1

57349 •

National Library
of CanadaBibliothèque nationale
du CanadaCANADIAN THESES
ON MICROFICHETHÈSES CANADIENNES
SUR MICROFICHE

NAME OF AUTHOR/NOM DE L'AUTEUR REGINALD J. KULPERGER

TITLE OF THESIS/TITRE DE LA THÈSE STATISTICAL INFERENCE FOR A
SIMPLE BRANCHING DIFFUSION
PROCESS

UNIVERSITY/UNIVERSITÉ CARLETON UNIVERSITY

DEGREE FOR WHICH THESIS WAS PRESENTED/
GRADE POUR LEQUEL CETTE THÈSE FUT PRÉSENTÉE DOCTOR OF PHILOSOPHY

YEAR THIS DEGREE CONFERRED/ANNÉE D'OBTENTION DE CE DEGRÉ 1978

NAME OF SUPERVISOR/NOM DU DIRECTEUR DE THÈSE DR. D. DAWSON

Permission is hereby granted to the NATIONAL LIBRARY OF
CANADA to microfilm this thesis and to lend or sell copies
of the film.

The author reserves other publication rights, and neither the
thesis, nor extensive extracts from it may be printed or other-
wise reproduced without the author's written permission.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈ-
QUE NATIONALE DU CANADA de microfilmer cette thèse et
de prêter ou de vendre des exemplaires du film.

L'auteur se réserve les autres droits de publication; ni la
thèse ni de longs extraits de celle-ci ne doivent être imprimés
ou autrement reproduits sans l'autorisation écrite de l'auteur.

DATED/DATE 14 June 1978 SIGNED/SIGNÉ Reginald J. Kulperger

PERMANENT ADDRESS/RÉSIDENCE FIXE 47 ADMIRAL ROAD, BRANTFORD, ONTARIO N3R 1L5

National Library of Canada

Cataloguing Branch
Canadian Theses Division

Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada

Direction du catalogage
Division des thèses canadiennes

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

**THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED**

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

**LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE**

STATISTICAL INFERENCE FOR A SIMPLE
BRANCHING DIFFUSION PROCESS

by

Reginald John Kulperger

A thesis submitted to the Faculty of Graduate
Studies in partial fulfillment of the requirements
for the degree of Doctor of Philosophy.

Carleton University

Ottawa, Ontario

May, 1978

The undersigned recommend to the Faculty
of Graduate Studies acceptance of the thesis
"Statistical Inference for a Simple
Branching Diffusion Process"
submitted by Reginald John Kulperger, M.Math.,
in partial fulfilment of the requirements for
the degree of Doctor of Philosophy.

L. Lawson

.....
Thesis Supervisor

D.R. Billinger

.....
External Examiner

Lawel

.....
Chairman, Department of Mathematics

Carleton University

Acknowledgement

I wish to express my thanks to my supervisor, Professor D.A. Dawson, and also to Dr. Gail Ivanoff, for many interesting discussions on branching diffusion processes and some related topics. I also wish to thank Mrs. Debbie Iscoe for an excellent typing job on this thesis and for generally being a co-operative and pleasant typist.

I would like to dedicate this thesis to my Mother and Father, and the rest of my family.

Abstract

In a simple biological population, for example of plankton, or of mutant genes introduced into a population of "normals", individuals migrate, give birth and die in some domain $D \subseteq \mathbb{R}^d$. A branching diffusion process is a mathematical model which describes this stochastic evolutionary behaviour. A particular model known as a simple branching diffusion process is considered here. The model is a simple branching process, with Brownian movement or migration added, in which individuals either die or split into two individuals. In addition, immigration is assumed to take place. Let θ be the parameter of the process, consisting of the birth and death rates, the diffusion rate, and the immigration rate.

The main purpose of this study is to consider the problem of estimating the parameter θ of the above mentioned process.

Let $\{X_t(x): x \in \mathbb{R}^d, t \geq 0\}$ be the stochastic evolutionary process. If $0 \leq \tau_1 < \dots < \tau_L$ are taken to be fixed times, then $\{X(x) = (X_{\tau_1}(x), \dots, X_{\tau_L}(x)): x \in \mathbb{R}^d\}$ is an L -dimensional point process on $\mathbb{R}^d$. Such an induced point process is called an L time slice version of the stochastic evolutionary process. For the simple branching diffusion immigration process, the parameter of an L time slice version is $(\theta, \tau_1, \dots, \tau_L)$. Estimating this parameter yields an estimate of θ .

Now consider a particular L time slice version of the simple branching diffusion immigration process, with times $0 < \tau_1 < \dots < \tau_L$. Let $N_\tau(A)$ be the number of individuals in A at time τ . For the two sets A and $B \subseteq R^d$, the random variables $N_{\tau_i}(A)$ and $N_{\tau_i}(B)$ are not independent. It is shown, however, that this process satisfies some Brillinger-type mixing conditions. Various central limit theorems (CLT's) are given for Brillinger mixing point processes. As a corollary, a CLT is immediately obtained for $(N_{\tau_1}(A), \dots, N_{\tau_L}(A))$ as $A \uparrow R^d$.

Using CLT's for Fourier transforms of Brillinger mixing point processes, an asymptotic likelihood function is defined. This leads to consistent estimates of the parameters of the L time slice version of the simple branching diffusion process. Under some conditions $A \uparrow R^d$, limiting normal distributions are obtained for these estimates.

The parameters of an L time slice version of the simple branching diffusion process are time scale dependent. If $L = 2$, with times τ and $\tau + t$, where $t > 0$ is an experimental parameter, then even the age, τ , of the process can be estimated. Since the process is observed on a large set $A \subseteq R^d$ at the fixed, unknown times, this means that with no "tagging" or following of the individuals, such parameters as the rate of movement (diffusion rate) and the age, τ , can still be estimated.

TABLE OF CONTENTS

	Page
Acknowledgement	i
Abstract	ii
§1 Introduction and Summary	1
§2 Preliminaries and Mixing	5
§3 Some Results in Estimation for Stationary Point Processes	54
§4 Parametric Inference for a Simple Branching Diffusion Process	91
§4.1 Introduction and Summary	91
§4.2 Approximate Likelihood Inference for a 1-Time Slice Immigration Process	94
§4.3 Asymptotic Likelihood Estimation for an L-Dimensional B-Mixing Process	135
§4.4 Asymptotic Minimum Squares Estimation	140
§5 Concluding Remarks	146
References	150
Appendix I	152
Appendix II	155

§1 Introduction and Summary

Many populations behave as stochastic evolutionary point processes, for example, animals in a forest, mutant genes or plankton. We consider a simple process giving rise to such phenomena, namely a simple branching diffusion process described below. It is a simple model, but is useful in genetical studies (Sawyer (1976)) or for simple biological processes such as plankton. This model is studied in some detail in Ivanoff (1976) and is also described in Dawson and Ivanoff (1977). The main interest of this work is to statistically estimate the parameters of this process.

The simple branching diffusion process is a Markov particle system in which the particles independently undergo a random motion in R^d and branching. The process is described as follows:

(a) Brownian spatial process. Each particle is assumed to perform an independent Brownian motion in R^d with probability transition density function, until the time it branches, given by

$$p(t, x, y) = (2\pi pt)^{-d/2} \exp\left\{ \frac{-|x-y|^2}{2pt} \right\}.$$

The parameter $p > 0$ is the diffusion constant or diffusion rate, and $t \geq 0$ is time.

(b) Branching rate. The probability that a given particle branches in the interval $[t, t + \Delta t)$ is $V\Delta t + o(\Delta t)$. The time to the first branch is an exponential random variable with mean V^{-1} . $V > 0$ is called the branching rate. Particles branch independently.

(c) Binary branching mechanism. When a particle branches, it dies with probability $1 - \alpha$ and is replaced in the same location by two descendants or offspring with probability α , $0 < \alpha < 1$. The two descendants then act independently.

The simple branching diffusion immigration process, called the immigration process for short, is described by (a) - (c) above and

(d) Immigration. Suppose at time zero there are no individuals but that immigrants arrive according to a Poisson process on $R^d \times R^+$ with immigration rate $r > 0$.

For this process, let $m = V(1 - 2\alpha)$, the death rate minus the birth rate. The process is sub-critical if $m > 0$, critical if $m = 0$ and supercritical if $m < 0$.

The parameters of the immigration process are p , V , m and r .

In establishing Brillinger type mixing conditions for the immigration process, three simple branching diffusion processes are considered:

- (i) initially one particle, at $a \in R^d$,
- (ii) an initial Poisson field of particles, no immigration,
- (iii) the immigration process.

Among other things, Ivanoff (1976) shows for each case

- (i) for each time $t \geq 0$, there are no multiple points, that is

$$P\{\text{there exists } a \in R^d \text{ such that } N_t(da) \geq 2\} = 0$$

where $N_t(A)$ is the number of individuals in A at time t ,

(ii) if $A \subseteq R^d$ is bounded, measurable and t is a fixed time, then almost surely there is no branch in A at time t ,

(iii) intensity functions of all orders exist.

In the problem of estimating the parameters of the immigration process, if one can observe the immigration place, branch place and corresponding times, of a large number of immigrants, then one can write down the likelihood function. This gives maximum likelihood estimates (mle) and is an obvious extension of similar results of Keiding (1975) for a birth-death process. This "complete tagging" is not a reasonable set of data to obtain in some biological populations.

In an evolutionary process, considering the process at one time gives rise to a one dimensional point process. Considering it at say L distinct times gives an L dimensional point process. We refer to this as an L time slice version of the evolutionary process. In this work, we generally work with an L time slice version of the immigration process, and usually with $L = 1$ or 2 . The data is assumed to be an L time slice version observed on some set $A \subseteq R^d$. No individual is followed through time. This type of data might be more reasonable for some biological population processes than that of the preceding paragraph.

Chapter 2 derives recursive formulae for the intensity functions of the L time slice version of the immigration process. Using these, we are able to prove that the L time slice version satisfies some Brillinger type mixing conditions.

Chapter 3 considers some CLT's of a Brillinger mixing point process in R^d , and of Fourier transforms of such processes. Also a strong law of large numbers (SLLN) is obtained. As immediate corollaries,

by the results of chapter 2, these results hold for an L time slice version of the immigration process. Periodograms are also introduced.

Chapter 4 is the main chapter. It introduces the estimation problem in more detail. The periodograms of Chapter 3 are asymptotically independent. Based on these, an approximate or asymptotic likelihood is defined. This allows estimates of the parameters of the immigration process to be obtained. These estimates are asymptotically normal. An optimality condition is defined, and "optimal normal" results are obtained. If $d \geq 2$, these results are also asymptotically fully Fisher efficient. The estimates are also almost surely consistent. To avoid certain technicalities, an asymptotic least squares procedure is given. The estimate is again almost surely consistent and asymptotically normal. "Optimal normal" results are obtained. In case $L \geq 2$, it turns out that even the age of the simple branching diffusion immigration process can be estimated, along with all other parameters.

Chapter 5 gives some concluding remarks. Among them, a partial test of fit statistic is proposed.

The important mathematical results and computations are the proof of the mixing conditions in Chapter 2, Theorem 3.3 and Theorem 4.2.5. Using the mixing condition proven in Chapter 2, bounds on the rate of growth of cumulants of Fourier transforms of the process are obtained in Theorem 3.3. Chapter 4 considers a quadratic functional of the Fourier transforms using a cumulant calculus of Leonov and Shiryaev bounds on the growth rate of the cumulants of the quadratic function are obtained in Theorem 4.2.5.

§2 Preliminaries and Mixing

Summary and Outline of §2

We develop recurrence formulae for multiple time intensity functions for the simple branching diffusion process in the case of one initial particle, the initial Poisson field case and the immigration process case. These allow us to prove the initial Poisson field and immigration process are stationary. Also we show that these processes have no multiple points.

Using generating functionals and the notion of cluster processes, we then show that the factorial cumulants of the t -time slice initial Poisson field are non-negative since they are integrals of the one initial particle intensity functions. This integral representation allows us to show Brillinger-mixing for the initial Poisson field (Theorem 2.12) and an analogue of Brillinger's ℓ -summable property (Assumption 2.6.3, Brillinger (1975), p. 27) (Proposition 2.16).

Defining a delayed immigration process allows us to write $(N_{\tau_1}^I(B), \dots, N_{\tau_L}^I(B))$ as a limiting sum of independent initial Poisson field simple branching diffusion processes for each bounded Borel set B , where $N_{\tau}^I(B)$ is the number of individuals of the immigration process in B at time τ . This gives the factorial cumulants of the immigration process as an integral of the factorial cumulants of the initial Poisson field case, and hence as an integral of the intensities of the one initial particle case. This allows a proof of B -mixing (Theorem 2.13) and an ℓ -summable property (Theorem 2.16) for the immigration process.

Consider the simple branching diffusion process described in the introduction. It is discussed in greater mathematical detail in Ivanoff (1976). The branching time for an individual is exponentially distributed with rate V . When a branch occurs, the binary mechanism results in a birth with probability α and death with probability $1 - \alpha$. Let $m = V(1 - 2\alpha)$. The process is supercritical when $m > 0$, critical when $m = 0$ and subcritical when $m < 0$.

Let $\{X(t), t \geq 0, x \in \mathbb{R}^d\}$ be the simple branching diffusion process. $X(0)$ then denotes the initial condition. The notation for factorial densities is given here. In the case that there is initially one particle located at $a \in \mathbb{R}^d$ at time $t = 0$,

$$P_1(t, a, b) h^d \sim \Pr\{\text{a particle is in } S_h(b) \text{ at time } t \mid X(0) = a\},$$

where $S_h(b) = \{x \in \mathbb{R}^d: \|x - b\|_\infty \leq \frac{h}{2}\}$ is a cube of side length h , and centred at b and where " $X(0) = a$ " means, one particle at time 0 and it is located at $a \in \mathbb{R}^d$. Similar for higher order densities

$$P_n(t, a, b_1, \dots, b_n) h_1^d \dots h_n^d \sim \Pr\{\text{particles in } S_{h_1}(b_1), \dots, S_{h_n}(b_n) \text{ at } t \mid X(0) = a\}.$$

For more than one time, we define

$$P_{(n,m)}(t_1, t_2, a, a_1, \dots, a_n, b_1, \dots, b_n) h_1^d \dots h_n^d s_1^d \dots s_m^d.$$

$\sim \Pr \{ \text{particles in } S_{n_i}(a_i) \text{ at } t_1, i=1, \dots, n \text{ and}$

$S_{s_j}(b_j) \text{ at } t_2, j=1, \dots, m \mid X(0) = a \}$,

for two times, and in general for times $t_1, \dots, t_k$

$P_{(n_1, \dots, n_k)}(t_1, \dots, t_k, a, a_{1,1}, \dots, a_{1,n_1}, a_{2,1}, \dots, a_{2,n_2}, \dots, a_{k,n_k})$

$\prod_{i=1}^k \prod_{j=1}^{n_i} h_{i,j}^d \sim \Pr \{ \text{particle in } S_{h_{i,j}}(a_{i,j}) \text{ at time}$

$t_i, j=1, \dots, n_i, i=1, \dots, k \mid X(0) = a \}$.

In case one of the $n_i = 0$, we set, by convention, this term to the corresponding one with the appropriate t_i deleted; for example

$$P_{(n,0)}(t_1, t_2, a, a_1, \dots, a_n) = P_n(t_1, a, a_1, \dots, a_n).$$

If there is an initial Poisson field, we similarly define factorial intensity functions

$$P_{(n_1, \dots, n_k)}(t_1, \dots, t_k, a_1, 1, \dots, a_{1,n_1}, a_{2,1}, \dots, a_{k,1}, \dots, a_{k,n_k}).$$

For the immigration process, we define

$$P_{(n_1, \dots, n_k)}^I(t_1, \dots, t_k, a_1, 1, \dots, a_{1,n_1}, a_{2,1}, \dots, a_{k,1}, \dots, a_{k,n_k}).$$

That these densities exist is shown below (Theorem 2.1). It is also clear the densities are measurable functions in all their arguments. Note that we sometimes write domain D in place of R^d . The recursion relations below are just extension of Ivanoff's to multiple times and are valid for motions other than just Brownian motion. Let $p(t,a,b)$ be the motion density function. For Brownian motion with diffusion rate p it is (suppressing the dependence on p)

$$p(t,a,b) = (2\pi pt)^{d/2} \exp\{-|a-b|^2/(2pt)\}.$$

Remark: To obtain the recursion formulae, we implicitly assume there are no multiple points. This is proven in Theorem 2.6, although it could be proven first. The order we prove the results in seems to be notationally more convenient.

Theorem 2.1

(a) (Ivanoff (1976))

$$p_1(t,a,b) = e^{-mt} p(t,a,b),$$

$$p_n(t,a,b_1,\dots,b_n) = 2V\alpha \int_0^t \int_D e^{-ms} p(s,a,y) f_n(t-s,y,b_1,\dots,b_n) dy ds,$$

where

$$\begin{aligned}
2f_n(t, a, b_1, \dots, b_n) &= p_1(t, a, b_1) p_{n-1}(t, a, b_2, \dots, b_n) \\
&+ p_1(t, a, b_2) p_{n-1}(t, a, b_1, b_3, \dots, b_n) \\
&+ \dots + p_1(t, a, b_n) p_{n-1}(t, a, b_1, \dots, b_{n-1}) \\
&+ p_2(t, a, b_1, b_2) p_{n-2}(t, a, b_3, \dots, b_n) \\
&+ \dots + p_2(t, a, b_{n-1}, b_n) p_{n-2}(t, a, b_1, \dots, b_{n-2}) + \dots \\
&+ \dots + p_{n-1}(t, a, b_1, \dots, b_{n-1}) p_1(t, a, b_n) \\
&= \sum_{k=1}^{n-1} \left(\sum_{\rho \in S_k} p_k(t, a, b_{\rho(1)}, \dots, b_{\rho(k)}) \right. \\
&\quad \left. p_{n-k}(t, a, b_{\rho(k+1)}, \dots, b_{\rho(n)}) \right)
\end{aligned}$$

where $\sum_{\rho \in S_k}$ is the sum over the $\binom{n}{k}$ combinations of dividing

$b_1, \dots, b_n$ into two subsets of size k and $n-k$.

(b) Let $0 < \tau_1 < \tau_2$.

$$P_{(1,n)}(\tau_1, \tau_2, a, a_1, b_1, \dots, b_n) = e^{-m\tau_1} p(\tau_1, a, a_1) p_n(\tau_2 - \tau_1, a_1, b_1, \dots, b_n)$$

$$+ 2V\alpha \int_0^{\tau_1} \int_D e^{-ms} p(s, a, y) f_{(1,n)}(\tau_1 - s_1, \tau_2 - s, y, a_1, b_1, \dots, b_n) dy ds,$$

and if $n \geq 2, m \neq 0$

$$\begin{aligned} P_{(n,m)}(\tau_1, \tau_2, a, a_1, \dots, a_n, b_1, \dots, b_m) \\ = 2V\alpha \int_0^{\tau_1} \int_D e^{-ms} p(s, a, y) f_{(n,m)}(\tau_1 - s_1, \tau_2 - s, y, a_1, \dots, a_n, \\ b_1, \dots, b_m) dy ds, \end{aligned}$$

where

$$\begin{aligned} f_{(n,m)}(\tau_1, \tau_2, a, a_1, \dots, a_n, b_1, \dots, b_m) \\ = \frac{1}{2} \sum_{j=0}^n \sum_{k=0}^m \sum_{\rho \in S_j} \sum_{\sigma \in S_k} p_{(j,k)}(\tau_1, \tau_2, a, a_{\rho(1)}, \dots, a_{\rho(j)}, \\ b_{\sigma(1)}, \dots, b_{\sigma(k)}) p_{(n-j, m-k)}(\tau_1, \tau_2, a, a_{\rho(j+1)}, \dots, a_{\rho(n)}, \\ b_{\sigma(k+1)}, \dots, b_{\sigma(m)}). \end{aligned}$$

(c) In general, where $0 < \tau_1 < \tau_2 < \dots < \tau_L$

$$\begin{aligned} P_{(1, n_2, \dots, n_L)}(\tau_1, \dots, \tau_L, a, a_{1,1}, a_{2,1}, \dots, a_{2, n_2}, a_{3,1}, \dots, a_{L, n_L}) \\ = e^{-m\tau_1} p(\tau_1, a, a_{1,1}) P_{(n_2, \dots, n_L)}(\tau_2 - \tau_1, \dots, \tau_L - \tau_1, \\ a_{1,1}, a_{2,1}, \dots, a_{2, n_2}, a_{3,1}, \dots, a_{L, n_L}) \end{aligned}$$

$$+ 2V\alpha \int_0^{\tau_1} \int_D e^{-ms} p(s, a, y) f_{(1, n_2, \dots, n_L)}(\tau_1 - s, \dots, \tau_L - s, y,$$

$$a_{1,1}, a_{2,1}, \dots, a_{2, n_2}, \dots, a_{L, n_L}) dy ds ,$$

and if $n_1 \geq 2$,

$$P_{(n_1, \dots, n_L)}(\tau_1, \dots, \tau_L, a, a_{1,1}, \dots, a_{1, n_1}, a_{2,1}, \dots, a_{L, n_L})$$

$$= 2V \int_0^{\tau_1} \int_D e^{-ms} p(s, a, y) f_{(n_1, \dots, n_L)}(\tau_1 - s, \dots, \tau_L - s, y, \\ a_{1,1}, \dots, a_{L, n_L}) dy ds ,$$

where

$$f_{(n_1, \dots, n_L)}(\tau_1, \dots, \tau_L, a, a_{1,1}, \dots, a_{1, n_1}, a_{2,1}, \dots, a_{L, n_L}) \\ = \frac{1}{2} \sum_{k_1=0}^{n_1} \dots \sum_{k_L=0}^{n_L} \sum_{p_1 \in s_{k_1}} \dots \sum_{p_L \in s_{k_L}} \{P_{(k_1, \dots, k_L)}(\tau_1, \dots, \tau_L, a, a \\ 0 \neq k_1 + \dots + k_L \neq n_1 + \dots + n_L$$

$$a_{1, p_1(k_1)}, a_{2, p_2(1)}, \dots, a_{L, p_L(k_L)}) P_{(n_1 - k_1, \dots, n_L - k_L)}$$

$$(\tau_1, \dots, \tau_L, a, a_{1, p_1(k_1+1)}, \dots, a_{1, p_1(n_1)}, a_{2, p_2(k_2+1)}, \dots,$$

$$a_{L, p_L(n_L)}) \} .$$

Proof:

(a) This is Ivanoff's Theorem 1, p. 20.

(b) Let the two times be $0 < \tau_1 < \tau_1 + \tau_2$.

$$P_{(1,1)}(\tau_1, \tau_1 + \tau_2, a, a_1, a_2)$$

= Pr{no branch occurs before τ_1 , the particle is at a_1 at τ_1 , and then at τ_2 later a descendant of this particle is a_2 } + Pr{a branch occur before τ_1 gives birth and one of the two or a combination of the two, particles results in (τ_1, a_1) and $(\tau_1 + \tau_2, a_2)$ }

(where (τ_1, a_1) refers to a particle at a_1 at time τ_1)

$$\begin{aligned} &= e^{-V\tau_1} p(\tau_1, a, a_1) P_1(\tau_2, a_1, a_2) \\ &+ 2V\alpha \int_0^{\tau_1} \int_D e^{-Vs} p(s, a, y) \{P_{(1,1)}(\tau_1 - s, \tau_1 + \tau_2 - s, y, a_1, a_2) \\ &+ f_{(1,1)}(\tau_1 - s, \tau_1 + \tau_2 - s, y, a_1, a_2)\} dy ds. \end{aligned} \quad (1)$$

The 2 in (1) comes about since the two offspring born at s behave independently and there are no multiple points. Let

$$h_{(1,1)}(\tau_1, \tau_1 + \tau_2, a, a_1, a_2) = P_{(1,1)}(\tau_1, \tau_1 + \tau_2, a, a_1, a_2)$$

$$-e^{-V\tau_1} p(\tau_1, a, a_1) P_1(\tau_1, a_1, a_2).$$

Then (1) gives

$$h_{(1,1)}(\tau_1, \tau_1 + \tau_2, a, a_1, a_2)$$

$$= 2V\alpha \int_0^{\tau_1} \int_D e^{-Vs} p(s, a, y) \{ h_{(1,1)}(\tau_1 - s, \tau_2 + \tau_1 - s, y, a_1, a_2) \\ + f_{(1,1)}(\tau_1 - s, \tau_1 + \tau_2 - s, y, a_1, a_2) + e^{-V(\tau_1 - s)} p(\tau_1 - s, y, a_1) \} \quad (2)$$

$$\cdot P_1(\tau_2, a_1, a_2) \} dy ds.$$

For each τ_2 , the Laplace transform of both sides exists. Equation (2) is the same as Ivanoff's equation (1), p. 23 (see also Dawson and Ivanoff (1978)). By the Laplace transform and semigroup technique she uses (see also Sawyer (1976) for similar computations) we get

$$P_{(1,1)}(\tau_1, \tau_1 + \tau_2, a, a_1, a_2) = h_{(1,1)}(\tau_1, \tau_1 + \tau_2, a, a_1, a_2) + e^{-V\tau_1}$$

$$p(\tau_1, a, a_1) P_1(\tau_2, a_1, a_2)$$

$$= e^{-V\tau_1} p(\tau_1, a, a_1) P_1(\tau_2, a_1, a_2) + 2V\alpha \int_0^{\tau_1} \int_D e^{-ms} p(s, a, y)$$

$$\{ f_{(1,1)}(\tau_1 - s, \tau_2 + \tau_1 - s, y, a_1, a_2) + e^{-V(\tau_1 - s)} p(\tau_1 - s, y, a_1) \}$$

$$\cdot P_1(\tau_2, a_1, a_2) \} dy ds$$

$$= e^{-m\tau_1} p(\tau_1, a, a_1) P_1(\tau_2, a_1, a_2) + 2V\alpha \int_0^{\tau_1} \int_D e^{-ms} p(s, a, y)$$

$$f_{(1,1)}(\tau_1-s, \tau_2+\tau_1-s, y, a_1, a_2) dy ds$$

by integration.

As in (1) we obtain

$$\begin{aligned} P_{(1,n)}(\tau_1, \tau_1+\tau_2, a, a_1, b_1, \dots, b_n) &= e^{-V\tau_1} p(\tau_1, a, a_1) P_n(\tau_2, a_1, b_1, \dots, b_n) \\ &+ 2V\alpha \int_0^{\tau_1} \int_D e^{-Vs} p(s, a, y) \{P_{(1,n)}(\tau_1-s, \tau_1+\tau_2-s, y, a_1, b_1, \dots, b_n) \\ &+ f_{(1,n)}(\tau_1-s, \tau_1+\tau_2-s, y, a_1, b_1, \dots, b_n)\} dy ds. \end{aligned}$$

Solving, as in (2),

$$\begin{aligned} P_{(1,n)}(\tau_1, \tau_1+\tau_2, a, a_1, b_1, \dots, b_n) &= e^{-V\tau_1} p(\tau_1, a, a_1) P_n(\tau_2, a_1, b_1, \dots, b_n) \\ &+ 2V\alpha \int_0^{\tau_1} \int_D e^{-ms} p(s, a, y) \{f_{(1,n)}(\tau_1-s, \tau_1+\tau_2-s, y, a_1, b_1, \dots, b_n) \\ &+ e^{-V(\tau_1-s)} p(\tau_1-s, y, a_1) P_n(\tau_2, a_1, b_1, \dots, b_n)\} dy ds \\ &= e^{-m\tau_1} p(\tau_1, a, a_1) P_n(\tau_2, a_1, b_1, \dots, b_n) \\ &+ 2V\alpha \int_0^{\tau_1} \int_D e^{-ms} p(s, a, y) f_{(1,n)}(\tau_1-s, \tau_1+\tau_2-s, y, a_1, b_1, \dots, b_n) dy ds. \end{aligned}$$

In general, for $P_{(n,m)}$, $n \geq 2$, there must be a birth before τ_1 , so via a renewal type equation as in (1) we obtain

$$\begin{aligned} & P_{(n,m)}(\tau_1, \tau_1 + \tau_2, a, a_1, \dots, a_n, b_1, \dots, b_m) \\ &= 2V\alpha \int_0^{\tau_1} \int_D e^{-Vs} p(s, a, y) \{P_{(n,m)}(\tau_1 - s, \tau_1 + \tau_2 - s, y, a_1, \dots, a_n, \\ & \quad b_1, \dots, b_m) + f_{(n,m)}(\tau_1 - s, \tau_1 + \tau_2 - s, y, a_1, \dots, a_n, b_1, \dots, b_m)\} dy ds. \end{aligned}$$

Solving via the semi-group technique gives the solution in (ii) of the Theorem.

(c) Again we treat $n_1 = 1$ as a special case, since in this case there need not be a birth before τ_1 . Let the time be $0 < \tau_1 < \tau_1 + \tau_2 < \dots < \tau_1 + \dots + \tau_L$.

$$\begin{aligned} & P_{(1, n_2, \dots, n_L)}(\tau_1, \dots, \tau_1 + \dots + \tau_L, a, a_{1,1}, \dots, a_{2, n_2}, a_{3,1}, \dots, a_{L, n_L}) \\ &= e^{-V\tau_1} p(\tau_1, a, a_{1,1}) P_{(n_2, \dots, n_L)}(\tau_2, \dots, \tau_2 + \dots + \tau_L, a_{1,1}, a_{2,1}, \dots, \\ & \quad a_{2, n_2}, \dots, a_{L, n_L}) + 2V\alpha \int_0^{\tau_1} \int_D e^{-Vs} p(s, a, y) \{P_{(1, n_2, \dots, n_L)} \\ & \quad (\tau_1 - s, \tau_1 + \tau_2 - s, \dots, \tau_1 + \dots + \tau_L - s, y, a_{1,1}, \dots, a_{L, n_L}) \\ & \quad + f_{(1, n_2, \dots, n_L)}(\tau_1 - s, \dots, \tau_1 + \dots + \tau_L - s, y, a_{1,1}, a_{2,1}, \dots, a_{2, n_2}, \\ & \quad a_{3,1}, \dots, a_{L, n_L})\} dy ds. \end{aligned} \quad (3)$$

For each $\tau_2, \dots, \tau_L$, (3) is again Ivanoff's functional equation.

Applying the semi-group technique gives the solution in the Theorem.

For $n_1 \geq 2$, we obtain the functional equation

$$\begin{aligned} P_{(n_1, \dots, n_L)}(\tau_1, \dots, \tau_1 + \dots + \tau_L, a, a_{1,1}, \dots, a_{1,n}, a_{2,1}, \dots, a_{L,n_L}) \\ = 2V\alpha \int_0^{\tau_1} \int_D e^{-Vs} p(s, a, y) \{ P_{(n_1, \dots, n_L)}(\tau_1 - s, \dots, \tau_1 + \dots + \tau_L - s, \\ y, a_{1,1}, \dots, a_{L,n_L}) + f_{(n_1, \dots, n_L)}(\tau_1 - s, \dots, \tau_1 + \dots + \tau_L - s, \\ y, a_{1,1}, \dots, a_{L,n_L}) \} dy ds. \end{aligned}$$

By Ivanoff's method, again we obtain the solution given in the Theorem (iii). //QED.

Theorem 2.2 Consider the simple branching diffusion process, with initial Poisson field intensity λ . Then

(a) (Ivanoff (1976), p. 48)

$$P_n^P(\tau, b_1, \dots, b_n) = \sum_{j=1}^n \lambda^j \int_{D^j} \dots \int P_{n,j}(\tau, b_1, \dots, b_n, a_1, \dots, a_j) da_1 \dots da_j,$$

where

$$\begin{aligned} P_{n,j}(\tau, b_1, \dots, b_n, a_1, \dots, a_j) \\ = \sum_{o(k_1, \dots, k_j)} P_{K_1}(\tau, a_1, b_{p(1)}, \dots, b_{p(k_1)}) P_{K_2}(\tau, a_2, b_{p(k_1+1)}, \dots, b_{p(k_1+k_2)}) \end{aligned}$$

$$P_{K_j}(\tau, a_j, b_{p(n-k_j+1)}, \dots, b_{p(n)}) ,$$

where $\sum_{p(k_1, \dots, k_j)}$ indicates the sum over all S_n^j (a Stirling number

of the second kind, see e.g. Ivanoff (1976), Appendix A, or Jordan (1960), §4) possible groupings of n indistinguishable objects into j non-empty subsets.

(b) For $n, m \neq 0$, $0 < \tau_1 < \tau_2$

$$\begin{aligned} & P_{(n,m)}(\tau_1, \tau_2, a_1, \dots, a_n, b_1, \dots, b_m) \\ &= \sum_{j_1=1}^n \sum_{j_2=1}^m \sum_{j_{1,2}=0}^{j_1 j_2} \lambda^{j_1+j_2-j_{1,2}} \int \dots \int_0^{j_1+j_2-j_{1,2}} H_{(j_1, j_2, j_{1,2})}^{(n,m)} \\ & \quad (\tau_1, \tau_2, a_1, \dots, a_n, b_1, \dots, b_m, x_1, \dots, x_{j_1+j_2-j_{1,2}}) dx_1 \dots dx_{j_1+j_2-j_{1,2}} \end{aligned}$$

where, if $0 \neq j_{1,2} < j_1 j_2 = \min \{j_1, j_2\}$

$$\begin{aligned} & H_{(j_1, j_2, j_{1,2})}^{(n,m)} (\tau_1, \tau_2, a_1, \dots, a_n, b_1, \dots, b_m, x_1, \dots, x_{j_1+j_2-j_{1,2}}) \\ &= \sum_{p(k_1, \dots, k_{j_1})} \sum_{\sigma(\ell_1, \dots, \ell_{j_2})} P_{K_1}(\tau_1, x_1, a_{p(1)}, \dots, a_{p(k_1)}) \dots \end{aligned}$$

$$P_{K_{j_1-j_{1,2}}}(\tau_1, x_{j_1-j_{1,2}}, a_{p(k_1+\dots+k_{j_1-j_{1,2}-1}+1)}, \dots, a_{p(k_1+\dots+k_{j_1-j_{1,2}})})$$

$$P_{\ell_1}(\tau_2, x_{j_1-j_{1,2}+1}, b_{\sigma(1)}, \dots, b_{\sigma(\ell_1)}) \dots$$

$$\begin{aligned}
& P_{j_2-j_{1,2}}^{(\tau_2, x_{j_1+j_2-2j_{1,2}}, b_{\sigma(\ell_1+\dots+\ell_{j_2-j_{1,2}-1+1)}, \dots, \\
& \quad b_{\sigma(\ell_1+\dots+\ell_{j_2-j_{1,2}})})} \\
& \cdot P_{j_1-j_2+1, j_2-j_{1,2}+1}^{(\tau_1, \tau_2, x_{j_1+j_2-2j_{1,2}+1}, a_{\rho(k_1+\dots+k_{j_1-j_{1,2}+1})}, \dots, \\
& \quad a_{\rho(k_1+\dots+k_{j_1-j_{1,2}+1})}, b_{\sigma(\ell_1+\dots+\ell_{j_2-j_{1,2}+1})}, \dots, b_{\sigma(\ell_1+\dots+\ell_{j_2-j_{1,2}+1})})} \dots \\
& \cdot P_{j_1, j_2}^{(\tau_1, \tau_2, x_{j_1+j_2-j_{1,2}}, a_{\rho(n-k_{j_1}+1)}, \dots, a_{\rho(n)}, b_{\sigma(m-\ell_{j_2}+1)}, \dots, \\
& \quad b_{\sigma(m)})} .
\end{aligned}$$

The above formula holds for $j_{1,2} = 0$ if we replace the $P_{(\cdot, \cdot)}(\cdot)$ terms by 1's and it holds for $j_{1,2} = j_1$ or j_2 if we replace the appropriate P_{j_1} or P_{j_2} terms by 1's.

The formula for H is just the probability of the desired configuration at times τ_1 and τ_2 given that the points at time τ_1 and j_1 initial ancestors and those at time τ_2 have j_2 initial ancestors, but they have $j_{1,2}$ in common, and the τ_1 ancestors are at $x_1, \dots, x_{j_1-j_{1,2}}, x_{j_1+j_2-2j_{1,2}+1}, \dots, x_{j_1+j_2-j_{1,2}}$, the τ_2 ancestors are at $x_{j_1-j_{1,2}+1}, \dots, x_{j_1+j_2-2j_{1,2}}$ and the common ancestors are at $x_{j_1+j_2-2j_{1,2}+1}, \dots, x_{j_1+j_2-2j_{1,2}}$.

Remark: A similar formula holds for other multiple times, but fortunately all we need later is that the formula has the form (b) and not the actual formula.

Proof: (b) $H_{(j_1, j_2, j_{1,2})}^{(n,m)}$ is the conditional probability as indicated above. The probability element that an individual is at x_j in the initial Poisson field is λx_j , and so $\lambda^{j_1+j_2-j_{1,2}} dx_1 \dots dx_{j_1+j_2-j_{1,2}}$ is the probability element of the ancestors being at $x_1, \dots, x_{j_1+j_2-j_{1,2}}$. Therefore the probability of observing $(\tau_1, a_1), \dots, (\tau_1, a_n), (\tau_2, b_1), \dots, (\tau_2, b_m)$ is (b) above. //QED.

Theorem 2.3

For the immigration process, immigration rate r
 (a) (Ivanoff (1976), p. 84)

$$P_n^I(\tau, b_1, \dots, b_n) = \int_0^\tau \int_D r g_n(s, y, b_1, \dots, b_n) dy ds,$$

where

$$\begin{aligned} g_n(s, y, b_1, \dots, b_n) \\ = \sum_{k=1}^{n-1} \sum_{\rho \in S_k} P_k(s, y, b_{\rho(1)}, \dots, b_{\rho(k)}) P_{n-k}^I(s, b_{\rho(k+1)}, \dots, b_{\rho(n)}) \\ + P_n(s, y, b_1, \dots, b_n), \end{aligned}$$

where $\sum_{\rho \in S_k}$ means the sum over the $\binom{n}{k}$ divisions of $b_1, \dots, b_n$ into two subsets of size k and $n-k$.

$$(b) P_{(n_1, \dots, n_L)}^I(\tau_1, \dots, \tau_L, a_{1,1}, \dots, a_{1,n_1}, a_{2,1}, \dots, a_{L,1}, \dots, a_{L,n_L})$$

$$= \int_0^{\tau_1} \int_D r g(n_1, \dots, n_L)(s, y, \tau_1, \dots, \tau_L, a_{1,1}, \dots, a_{1,n_1}, a_{2,1}, \dots, a_{L,1}, \dots, a_{L,n_L}) dy ds,$$

where $0 < \tau_1 < \tau_2 < \dots < \tau_L$ and

$$\begin{aligned} & g(n_1, \dots, n_L)(s, y, \tau_1, \dots, \tau_2, a_{1,1}, \dots, a_{1,n_1}, a_{2,1}, \dots, a_{L,n_L}) \\ &= \sum_{k_1=0}^{n_1} \dots \sum_{k_L=0}^{n_L} \sum_{\rho_1 \in S_{k_1}} \dots \sum_{\rho_L \in S_{k_L}} \{P_{(k_1, \dots, k_L)}(\tau_1-s, \dots, \tau_L-s, y, \\ & \quad a_{1,\rho_1(1)}, \dots, a_{1,\rho_1(k_1)}, a_{2,\rho_2(1)}, \dots, a_{L,\rho_L(1)}, \dots, a_{L,\rho_L(k_L)} \\ & \quad \cdot P_{(n_1-k_1, \dots, n_L-k_L)}(\tau_1-s, \dots, \tau_L-s, a_{1,\rho_1(k_1+1)}, \dots, a_{1,\rho_1(n_1)}, \\ & \quad a_{2,\rho_2(k_2+1)}, \dots, a_{L,\rho_L(k_L+1)}, \dots, a_{L,\rho_L(n_L)})\} \\ & \quad + P_{(n_1, \dots, n_L)}(\tau_1-s, \dots, \tau_L-s, y, a_{1,1}, \dots, a_{1,n_1}, \dots, a_{L,1}, \dots, a_{L,n_L}). \end{aligned}$$

Proof: (a) Ivanoff, p. 84

(b) We use an analogous argument to that of Ivanoff, arriving at an obvious analogue of her functional equation. Formally it is the same as her functional equation and so we can immediately write down the solution (b).

First consider $D \subseteq \mathbb{R}^d$, $|D| < \infty$. Suppose the first immigrant arrives at $(s, y) \in \mathbb{R}^+ \times D$. It happens with probability

$$r|D|dy \cdot e^{-r|D|} \frac{ds}{|D|} = re^{-r|D|} s dy ds$$

It may be the ancestor of k_i of those at $(\tau_i, a_{i,j})$, $j = 1, \dots, n_i$, $i = 1, \dots, L$ and $0 \leq k_i \leq n_i$. We write this as the common ancestor of $(k_1, \dots, k_L)$ of those at times $\tau_1, \dots, \tau_L$. The first immigrant is the common ancestor of

$$\begin{aligned} (0, \dots, 0) & \quad \text{with probability } P_{(n_1, \dots, n_L)}^I(\tau_1 - s, \dots, \tau_L - s, a_{1,1}, \dots, a_{L,n_L}), \\ & \quad \vdots \\ (k_1, \dots, k_L) & \quad \text{with probability} \end{aligned}$$

$$\begin{aligned} & \sum_{\rho_1 \in S_{k_1}} \dots \sum_{\rho_L \in S_{k_L}} \{P_{(k_1, \dots, k_L)}(\tau_1 - s, \dots, \tau_L - s, y, a_{1,\rho_1(1)}, \dots, a_{1,\rho_1(k_1)} \\ & \quad a_{2,\rho_2(1)}, \dots, a_{L,\rho_L(1)}, \dots, a_{L,\rho_L(k_L)}) \cdot P_{(n_1 - k_1, \dots, n_L - k_L)}^I \\ & \quad (\tau_1 - s, \dots, \tau_L - s, a_{1,\rho_1(k_1+1)}, \dots, a_{1,\rho_1(n_1)}, a_{2,\rho_2(k_2+1)}, \dots, \\ & \quad a_{L,\rho_L(k_L+1)}, \dots, a_{L,\rho_L(n_L)})\} \end{aligned}$$

since the $(n_1 - k_1, \dots, n_L - k_L)$ who are not ancestors of the first immigrant result from an immigration process beginning at time $0 < s < \tau_1$.

This is just a renewal argument. Therefore if $|D| < \infty$,

$$P_{(n_1, \dots, n_L)}^I(\tau_1, \dots, \tau_L, a_{1,1}, \dots, a_{1,n_1}, a_{2,1}, \dots, a_{L,1}, \dots, a_{L,n_L})$$

$$\begin{aligned}
&= \int_0^{\tau_1} \int_D e^{-r|D|s} r \{P_{(n_1, \dots, n_L)}(\tau_1-s, \dots, \tau_L-s, a_{1,1}, \dots, a_{1,n_1}, \dots, \\
&\quad a_{L,1}, \dots, a_{L,n_L}) + g_{(n_1, \dots, n_L)}(s, y, a_{1,1}, \dots, a_{1,n_1}, \dots, a_{L,1}, \dots, \\
&\quad a_{L,n_L})\} dy ds.
\end{aligned}$$

This is Ivanoff's equation on p. 86 and so may be solved by the Laplace transform and semi-group technique. Then letting $D \uparrow \mathbb{R}^d$ gives the Theorem. //QED.

We now show that the L-time slice immigration process satisfies a Brillinger mixing condition. Some side results of our proof is that the L-time slice initial Poisson field case satisfies the same mixing condition and that both these processes satisfy a cumulant summability condition (§3 of this thesis, and Assumption 2.6.3, p. 27 Brillinger (1975) with factorial cumulants).

Let $\{X(t): t \in \mathbb{R}^d\}$ be a stationary L dimensional point process on $\mathbb{R}^d$ with $X(t) = (X_1(t), \dots, X_L(t))$. Let $c_{a_1 \dots a_k}(b_1, \dots, b_k) = \{X_{a_1}(b_1), \dots, X_{a_k}(b_k)\}$ be a joint k^{th} factorial cumulant of X (see §3).

Definition

The stationary point process X is said to be Brillinger mixing, or B-mixing, if

$$(a) \quad \forall \quad k \geq 2, \quad \int \dots \int_{\mathbb{R}^{d(k-1)}} |c_{a_1 \dots a_k}(b_1, \dots, b_{k-1})| db_1 \dots db_{k-1} < \infty$$

and

$$(b) \int_{R^d} |u| |c_{a_1 a_2}(u)| du < \infty$$

For stationary processes, we use both notations $c_{a_1 \dots a_k}(b_1, \dots, b_k)$ and $c_{a_1 \dots a_k}(b_1 - b_k, \dots, b_{k-1} - b_k)$ interchangeably.

Remark: This definition of B-mixing is not to be confused with anyone else's possible use of the term.

Theorem 2.4

The L-time slice initial Poisson field and immigration process versions are strictly stationary.

Lemma 2.5

$$P_{(n_1, \dots, n_L)}(\tau_1, \dots, \tau_L, x, a_{1,1}+a, \dots, a_{1,n_1}+a, a_{2,1}+a, \dots, a_{L,1}+a, \dots, a_{L,n_L}+a) \\ = P_{(n_1, \dots, n_L)}(\tau_1, \dots, \tau_L, x-a, a_{1,1}, \dots, a_{1,n_1}, a_{2,1}, \dots, a_{L,1}, \dots, a_{L,n_L})$$

Proof: This can be proven by induction on L and $n_1, \dots, n_L$. It is also clear by noting that the left and right hand sides are $(n_1, \dots, n_L)^{th}$ order intensity functions for simple branching diffusion processes with initially one initial particle at x and $x-a$ respectively. //QED.

Proof of Theorem 2.4

Since the initial Poisson field and the immigration cases are proven in a similar way only the immigration case is proven here. First consider $L = 1$.

$$\text{If } n = 1, P_1^I(\tau, a+b_1) = r \int_0^\tau \int_D P_1(s, y, b_1+a) dy ds = P_1^I(\tau, b_1)$$

Suppose the theorem holds for $1, \dots, n$ when $L = 1$.

$$g_{n+1}(s, y, b_1 + a, \dots, b_{n+1} + a)$$

$$= \sum_{k=1}^n \sum_{\rho \in S_k} P_k(s, y, b_{\rho(1)} + a, \dots, b_{\rho(k)} + a) P_{n-k}^I(s, b_{\rho(k+1)} + a, \dots, b_{\rho(n+1)} + a) \\ + P_{n+1}(s, y, b_1 + a, \dots, b_{n+1})$$

$= g_{n+1}(s, y - a, b_1, \dots, b_{n+1})$ by the induction hypothesis on n and Lemma 2.5.

$$P_{n+1}^I(\tau, b_1 + a, \dots, b_{n+1} + a) = \int_0^\tau \int_D r g_{n+1}(s, y - a, b_1, \dots, b_{n+1}) dy ds \\ = P_{n+1}^I(\tau, b_1, \dots, b_{n+1}) :$$

Therefore the theorem holds for $L = 1$ by induction.

Suppose the theorem holds for $1, \dots, L$.

By the induction hypothesis on L , Lemma 2.5 and the definition of g

$$g(1, \dots, 1)(s, y, \tau_1, \dots, \tau_{L+1}, a_1 + a, \dots, a_{L+1} + a) \\ = g(1, \dots, 1)(s, y - a, \tau_1, \dots, \tau_{L+1}, a_1, \dots, a_{L+1}).$$

$$\begin{aligned}
& P_{(1, \dots, 1)}^I(\tau_1, \dots, \tau_{L+1}, a_1+a, \dots, a_{L+1}+a) \\
&= \int_0^{\tau_1} \int_D r g_{(1, \dots, 1)}(s, y-a, \tau_1, \dots, \tau_{L+1}, a_1, \dots, a_{L+1}) dy ds \\
&= P_{(1, \dots, 1)}^I(\tau_1, \dots, \tau_{L+1}, a_1, \dots, a_{L+1}) .
\end{aligned}$$

Suppose the theorem holds for $n_1 = \dots = n_L = 1$ and $1, \dots, n_{L+1}$ for $L+1$. Then

$$\begin{aligned}
& g_{(1, \dots, 1, n_{L+1}+1)}(s, y, \tau_1, \dots, \tau_{L+1}, a_1+a, \dots, a_L+a, b_1+a, \dots, b_{n_{L+1}+1}+a) \\
&= g_{(1, \dots, 1, n_{L+1}+1)}(s, y-a, \tau_1, \dots, \tau_{L+1}, a_1, \dots, a_L, b_1, \dots, b_{n_{L+1}+1})
\end{aligned}$$

$$\begin{aligned}
& \text{Therefore } P_{(1, \dots, 1, n_{L+1}+1)}^I(\tau_1, \dots, \tau_{L+1}, a_1+a, \dots, a_L+a, b_1+a, \dots, b_{n_{L+1}+1}+a) \\
&= P_{(1, \dots, 1, n_{L+1}+1)}^I(\tau_1, \dots, \tau_{L+1}, a_1, \dots, a_L, b_1, \dots, b_{n_{L+1}+1})
\end{aligned}$$

Then by induction on n_{L+1} , the Theorem holds for $n_1 = \dots = n_L = 1$, for all n_{L+1} . Similarly, it holds for all n_L, n_{L+1} , when $n_1 = \dots = n_{L-1} = 1$. Repeating $L-1$ more times, we see the theorem holds for $L+1$. Therefore by induction on L , the Theorem holds for all L . //QED.

Theorem 2.6

The L -time slice versions of each of the three cases of the simple branching diffusion process we consider have no multiple points.

Proof:

Let $N_\tau(A|a)$, $N_\tau^P(A)$ and $N_\tau^I(A)$ be the numbers of individuals in A at time τ in the initially one particle, initial Poisson field and the immigration cases respectively.

$$N(A)^{(2)} = N_\tau(A)[N_\tau(A) - 1] .$$

Then if $|A| < \infty$

$$E\{N_\tau(A)^{(2)}\} = \sum_{k=2}^{\infty} k(k-1) P\{N_\tau(A) = k\} \geq 2 \Pr\{N_\tau(A) \geq 2\} , \quad (1)$$

and $\Pr\{(N_{\tau_1}(A), \dots, N_{\tau_L}(A)) \notin \{0,1\}^L\}$

$$= \Pr\{N_{\tau_i}(A) \notin \{0,1\} \text{ for some } i\} \quad (2)$$

$$\leq \sum_{i=1}^L \Pr\{N_{\tau_i}(A) \geq 2\} .$$

(i) For the initially one point at a case, Sawyer (1976, Theorem 2.2) gives

$$E\{N_\tau(A|a)^{(2)}\} = o(|A|) . \quad (3)$$

By Jagers (1973, Proposition 2.4) $\{N_\tau(x|a)\}$ for each τ , has no multiple points.

For times $\tau_1, \dots, \tau_L$, (1), (2) and (3) give

$$P\{(N_{\tau_1}(A|a), \dots, N_{\tau_L}(A|a)) \notin \{0,1\}^L\} = o(|A|).$$

Thus for given $\tau_1, \dots, \tau_L$, by Jagers, there are no multiple points.

(ii) For the initial Poisson field and the immigration process by (i), P_2^P and P_2^I can be computed by Theorem 2.2 and 2.3. The rest of the Theorem now follows from Jagers (1973) as in (i). //QED.

Theorem 2.7

$$\int_D \int_{n_1 + \dots + n_L} P_{(n_1, \dots, n_L)}(\tau_1, \dots, \tau_L, x, a_{1,1}, \dots, a_{L,1}, \dots, a_{L,n_L}) da_{1,1} \dots da_{L,n_L}$$

is finite and independent of x , and bounded in $\tau_1, \dots, \tau_L$ if $\max(\tau_i) \leq T_0 < \infty$, where the bound depends on T_0 .

Proof: $L=1$.

$$\int_D P_1(\tau, x, a) da = e^{-m\tau} \int_D p(\tau, x, a) da = e^{-m\tau}.$$

Suppose for $L = 1$, the Theorem holds for $1, \dots, n$. By Theorem 2.1

$$\int \dots \int_{D^n} P_n(\tau, x, a_1, \dots, a_n) da_1 \dots da_n$$

$$\begin{aligned}
&= \frac{1}{2} \sum_{k=1}^{n-1} \binom{n}{k} \int \dots \int_{D^n} P_k(\tau, x, a_1, \dots, a_k) P_{n-k}(\tau, x, a_{k+1}, \dots, a_n) da_1 \dots da_n \\
&= \frac{1}{2} \sum_{k=1}^{n-1} \binom{n}{k} \left\{ \int \dots \int_{D^k} P_k(\tau, x, a_1, \dots, a_k) da_1 \dots da_k \right\} \left\{ \int \dots \int_{D^{n-k}} P_{n-k}(\tau, x, a_{k+1}, \dots, a_n) \right. \\
&\quad \left. da_{k+1} \dots da_n \right\},
\end{aligned}$$

which is finite and independent of x by the induction hypothesis.

Thus the Theorem holds for $L = 1$ by induction. Suppose the theorem holds for $1, \dots, L$. Consider the case $L+1$. By Theorem 2.1 (c),

$$\begin{aligned}
&\int \dots \int_{D^{L+1}} P(1, \dots, 1)(\tau_1, \dots, \tau_{L+1}, x, a_1, \dots, a_{L+1}) da_1 \dots da_{L+1} \\
&= e^{-m\tau_1} \int_D \int \dots \int_{D^L} P(1, \dots, 1)(\tau_2 - \tau_1, \dots, \tau_{L+1} - \tau_1, a_1, a_2, \dots, a_{L+1}) \\
&\quad da_2 \dots da_{L+1} p(\tau_1, x, a_1) da_1 \\
&= \frac{1}{2} \sum_{k_1=0}^1 \dots \sum_{k_{L+1}=0}^1 \sum_{\rho_1 \in S_{k_1+1}} \dots \sum_{\rho_{L+1} \in S_{k_{L+1}+1}} 2V\alpha \int_0^{\tau_1} \int_D e^{-ms} p(s, a, y) \int \dots \int_{D^{\sum k_i}} \\
&\quad 0 \neq \sum k_i \neq k+1
\end{aligned}$$

$$P(k_1, \dots, k_{L+1})(\tau_1 - s, \dots, \tau_{L+1} - s, y, a_{\rho_1(k_1)}, \dots, a_{\rho_{L+1}(k_{L+1})})$$

$$\int \dots \int_{D^{L+1-\sum k_i}} P(1-k_1, \dots, 1-k_{L+1})(\tau_1 - s, \dots, \tau_{L+1} - s, a_{\rho_1(k+1)}, \dots,$$

$$a_{\rho_{L+1}(k_{L+1}+1)} da_1 \dots da_{L+1} dy ds,$$

which is finite and independent of x by applying the induction

hypothesis to the inner part of the iterated integrals.

As in previous proofs, it can be shown by induction on n_{L+1} that the Theorem holds for all n_{L+1} when $n_1 \dots = n_L = 1$ for $L+1$. Repeating the argument L times, as before, we then obtain the Theorem holds for all $n_1, \dots, n_{L+1}$ for $L+1$, that is the Theorem holds for $L+1$. Thus by induction on L , the Theorem holds for all L . //QED.

We now proceed to show that an L -time slice version of the initial Poisson field simple branching diffusion process is B-mixing.

Let $0 < \tau_1 < \dots < \tau_L$. The L time slice version can be thought of as a cluster process where each individual of the initial Poisson field is replaced by an L -vector cluster, with values $\{0,1\}^L$.

That is, each point in the initial field is independently replaced by a cluster centred at the initial point, and the cluster is a point process with no multiple points. The probability generating functional (defined in §3) for a Poisson field, rate λ is

$$G[\xi] = E\left\{ \prod_{x_i} (1 + (\xi(x_i) - 1)) \right\}, \text{ where } x_i \text{ is an occurrence point}$$

$$= \exp\left\{ \lambda \int_{R^d} (\xi(x) - 1) dx \right\}.$$

For an individual at a , if the cluster process centred at a has p.g.fcnal $G[\xi_1, \dots, \xi_L | a]$ say, then the cluster process for the initial Poisson field has p.g.fcnal

$$H[\xi_1, \dots, \xi_L] = G[G[\xi_1, \dots, \xi_L | \cdot]]$$

$$= \exp\left\{\lambda \int_{R^d} (G[\xi_1, \dots, \xi_L | a] - 1) da\right\}.$$

The factorial cumulant generating functional (53) for the cluster process is

$$\begin{aligned} \log H[\xi_1, \dots, \xi_L] &= \lambda \int_{R^d} (G[\xi_1, \dots, \xi_L | a] - 1) da \\ &= \lambda \int_{R^d} \left\{ \sum_{v_1, \dots, v_L} \frac{1}{v_1! \dots v_L!} \int_{v_1 + \dots + v_L} \dots \int \prod_{j=1}^L \prod_{k=1}^{v_j} (\xi(x_{v_0 + \dots + v_{j-1} + k}) - 1) \right. \\ &\quad \left. \cdot P_{(v_1, \dots, v_L)}(\tau_1, \dots, \tau_L, a, x_1, \dots, x_{v_1 + \dots + v_L}) dx_1 \dots dx_{v_1 + \dots + v_L} \right\} da, \end{aligned}$$

where $v_0 = 0$ and $P_{(v_1, \dots, v_L)}$ is given in Theorem 2.1.

For the simple branching diffusion process, $G[\xi_1, \dots, \xi_L | a]$ exists as a corollary of the work of P.S. Puri (1968).

Remark: For $\xi_i(x) = \begin{cases} z_i & \text{on } A_i \\ 1 & \text{on } A_i^c \end{cases}$

where A_i is compact, and $1 < z_i < 2$.

$$\log H[\xi_1, \dots, \xi_L] = \lambda \int_{R^d} \left\{ \sum_{v_1, \dots, v_L} \frac{1}{v_1! \dots v_L!} \int_{\prod_{i=1}^L A_i} \dots \int \prod_{j=1}^L \prod_{k=1}^{v_j} (\xi_j(x_{v_0 + \dots + v_{j-1} + k}) - 1) \right.$$

$$\left. (\xi_j(x_{v_0 + \dots + v_{j-1} + k}) - 1) \cdot P_{(v_1, \dots, v_L)}(\tau_1, \dots, \tau_L, a, x_1, \dots, x_{v_1 + \dots + v_L}) \right.$$

$$\left. dx_1 \dots dx_{v_1 + \dots + v_L} da \right\}$$

$$= \lambda \sum_{v_1, \dots, v_L} \frac{1}{1! \dots L!} \prod_{j=1}^L (z_j - 1)^{v_j} \int_{\prod A_j^{v_j}} \int_{R^d} P(v_1, \dots, v_L) \\ (\tau_1, \dots, \tau_L, x_1, \dots, x_{v_1 + \dots + v_L}) da dx_1 \dots dx_{v_1 + \dots + v_L},$$

by Fubini's Theorem, since each term is positive.

Factorial cumulant densities can be defined in terms of factorial moment densities (by , Appendix I). By Theorem 2.2, factorial moment densities (intensity functions) exist for the initial Poisson field case, and hence the factorial cumulants exist. Therefore by the above generating functional, the factorial cumulants for the initial Poisson field case are

$$q_{(n_1, \dots, n_L)}^P(\tau_1, \dots, \tau_L, a_{1,1}, \dots, a_{1,n_1}, a_{2,1}, \dots, a_{L,1}, \dots, a_{L,n_L}) \\ = \lambda \int_{R^d} P(n_1, \dots, n_L)(\tau_1, \dots, \tau_L, a, a_{1,1}, \dots, a_{1,n_1}, a_{2,1}, \dots, a_{L,1}, \dots, \\ a_{L,n_L}) da.$$

Notice that the factorial cumulant densities are all non-negative.

In view of stationarity we can as before, abuse notation and interchangeably write both

$$q_{(n_1, \dots, n_L)}(\tau_1, \dots, \tau_L, x_{1,1}, \dots, x_{1,n_1}, \dots, a_{L,1}, \dots, a_{L,n_L}), \text{ and}$$

$$q_{(n_1, \dots, n_L)}(\tau_1, \dots, \tau_L, x_{1,1}^{-x_{L,n_L}}, \dots, x_{1,n_1}^{-x_{L,n_L}}, \dots, x_{L,1}^{-x_{L,n_L}}, \dots, \\ x_{L,n_L}^{-x_{L,n_L}}).$$

Also, whenever more convenient, we permute the n_i 's and the τ_i 's, but always in the same way, but generally require $\tau_1 < \tau_i$, $i \neq 1$.

Lemma 2.8

$$(i) \int \dots \int_{D^n} P_n(\tau, a, a_1, \dots, a_n) da_1 \dots da_n = n! (V \alpha M_{(\tau, m)})^{n-1} e^{-m\tau},$$

$$(ii) \int \dots \int_{D^n} P_n(\tau, a, a_1, \dots, a_n) da_1 \dots da_{n-1} da = n! (V \alpha M_{(\tau, m)})^{n-1} e^{-m\tau},$$

$$\text{where } M_{(\tau, m)} = \begin{cases} \frac{1 - e^{-m\tau}}{m} & \text{if } m \neq 0 \\ \tau & \text{if } m = 0 \end{cases}$$

Proof: (i) and (ii) are easily verified for $n = 1$.

Assuming (i) and (ii) hold for $1, \dots, n$, it can be shown (i) holds for $n+1$ using Theorem 2.1. It can then be shown that (ii) holds for $n+1$ using Theorem 2.1 and Lemma 2.5. //QED.

Lemma 2.9

$$(a) \int \dots \int_{D^n} q_{n+1}^p(\tau, a_1, \dots, a_{n+1}) da_1 \dots da_n = \lambda(n+1)! (V \alpha M_{(\tau, m)})^n e^{-m\tau},$$

$$(b) \int \dots \int_{D^{n_1 + \dots + n_L - 1}} q_{(n_1, \dots, n_L)}^p(\tau_1, \dots, \tau_L, a_{1,1}, \dots, a_{1,n_1}, \dots, a_{L,1}, \dots, a_{L,n_L}) da_{1,1} \dots da_{L,n_L - 1},$$

exists, is finite, is independent of a_{L,n_L} and is bounded in $\tau_1, \dots, \tau_L$

if $\max_i \tau_j \leq T_0 < \infty$ (the bound depends on T_0), and is linear in λ .

Proof:

$$\begin{aligned}
 (a) \quad & \int_{D^n} \dots \int q_{n+1}(\tau, a_1, \dots, a_{n+1}) da_1 \dots da_n \\
 &= \lambda \int_{D^{n+1}} \dots \int P_{n+1}(\tau, a, a_1, \dots, a_{n+1}) da da_1 \dots da_n \\
 &= \lambda(n+1)! (V \alpha M_{(\tau, m)})^n e^{-m\tau}, \quad \text{by Theorem 2:8 (b)}
 \end{aligned}$$

(b) By part (a), part (b) holds for $L = 1$.

Suppose it holds for $1, \dots, L$.

$$\begin{aligned}
 & \int_{D^{L+1-1}} \dots \int q_{(1, \dots, 1)}^P(\tau_1, \dots, \tau_{L+1}, a_1, \dots, a_{L+1}) da_1 \dots da_L \\
 &= \lambda \int_{D^{L+1}} \dots \int P_{(1, \dots, 1)}(\tau_1, \dots, \tau_{L+1}, a, a_1, \dots, a_{L+1}) da_1 \dots da_L da \\
 &+ \lambda \int_0^{\tau_1} \int_{D^{L+2}} \dots \int e^{-ms} p(s, a, y) f_{(1, \dots, 1)}(\tau_1 - s, \dots, \tau_{L+1} - s, y, a_1, \dots, a_{L+1}) \\
 &\quad dy da_1 \dots da_L da ds.
 \end{aligned}$$

By the induction hypothesis, the first term is clearly finite, using the formula in Theorem 2.1, is bounded in τ_i and is linear in λ . For the

second term, if $1 \leq k \leq j$, where $\rho(k) = L+1$, and where j of the k_i 's are 1's in the term of the formula for $f_{(1, \dots, 1)}$ (Theorem 2.1),

$$\begin{aligned} & \int_0^{\tau_1} \int_{D^{j+1}} \dots \int e^{-ms} P_{(k_1, \dots, k_{L+1})}(\tau_1 - s, \dots, \tau_{L+1} - s, y, a_{\rho(1)}, \dots, a_{\rho(j)}) \\ & \{ \int_{D^{L-j}} \dots \int P_{(1-k_1, \dots, 1-k_{L+1})}(\tau_1 - s, \dots, \tau_{L+1} - s, y, a_{\rho(j+1)}, \dots, \\ & a_{\rho(L+1)}) da_{\rho(j+1)} \dots da_{\rho(L+1)} \} \prod_{\substack{i \neq k \\ i=1}}^j da_{\rho(i)} dy ds \\ & = \int_0^{\tau_1} e^{-ms} h_{(k_1, \dots, k_j)}(\tau_1, \dots, \tau_{L+1}, s) h_{(k_{j+1}, \dots, k_{L+1})}^* \\ & (\tau_1, \dots, \tau_{L+1}, s) ds, \end{aligned}$$

where h and h^* are positive bounded and measurable, and hence $*$ is finite and bounded. Thus we see

$$\int_{D^L} \dots \int^P q_{(1, \dots, 1)}(\tau_1, \dots, \tau_{L+1}, a_1, \dots, a_{L+1}) da_1 \dots da_L$$

satisfies the Theorem. Thus the theorem holds for $n_1 = \dots = n_{L+1} = 1$ for $L+1$. As before, we can prove the Theorem holds for $L+1$ for all $n_1, \dots, n_{L+1}$ by successive induction on $n_{L+1}, n_L, \dots, n_1$ by considering terms like $*$.

The Lemma is proven by induction on L . //QED.

Remark: Lemma 2.9 proves condition (a) of the B-mixing definition for the L-time slice initial Poisson field case.

Lemma 2.10

For $p(\tau, y, x)$, the $N_d(y, \rho\tau I_d)$ density, $x = (x_1, \dots, x_d)$

$$\begin{aligned} \int_{\mathbb{R}^d} |x_1^2 + \dots + x_d^2|^{\ell/2} p(\tau, y, x) dx \\ = \frac{(\rho\tau)^{\frac{\ell}{2}} \cdot 2^{\frac{d+1}{2}} \Gamma(\frac{d+\ell}{2})}{\Gamma(\frac{d-1}{2}) \Gamma(\frac{1}{2})} \sum_{r=0}^{\infty} \frac{e^{-\frac{||y||^2}{\rho\tau}}}{(\rho\tau)^r} \frac{||y||^{2r} r^r}{(2r)!} B\left\{\frac{d-1}{2}, \frac{1}{2} + r\right\}. \end{aligned}$$

Proof:

$$\int_{\mathbb{R}^d} |x_1^2 + \dots + x_d^2|^{\ell/2} p(\tau, y, x) dx = E\{|x_1^2 + \dots + x_d^2|^{\ell/2}\},$$

where $x_1, \dots, x_d$ are independent $N(y_i, \rho\tau)$ r.v.'s. Then

$$Z = \sum_{i=1}^d [(\rho\tau)^{-1/2} x_i]^2 \sim \chi^2_{\left(d, \left(\frac{||y||}{\sqrt{\rho\tau}}\right)^2\right)}$$

that is, non-central χ^2 on d degrees of freedom and non-centrality parameter $||y||^2/(\rho\tau)$. (See Kendall and Stuart, Vol. I, p. 271, Ex. 11.12)

$$E\{|x_1^2 + \dots + x_d^2|^{\ell/2}\} = E\{Z^{\ell/2}\} (\rho\tau)^{\ell/2}$$

$$= (\rho\tau)^{d/2} \int_0^\infty \frac{z^{d/2} e^{-z/2} z^{d/2-1}}{2^{d/2} \Gamma(\frac{d-1}{2}) \Gamma(\frac{1}{2})} dz \sum_{r=0}^\infty e^{-\frac{\|y\|^2}{\rho\tau}} \left(\frac{\|y\|^2}{\rho\tau} \right)^r \frac{r^r}{(2r)!}$$

$$B\left\{\frac{1}{2}(d-1), \frac{1}{2} + r\right\}$$

$$= (\rho\tau)^{d/2} \frac{2^{\frac{d+1}{2}} \Gamma(\frac{d+1}{2})}{\Gamma(\frac{d-1}{2}) \Gamma(\frac{1}{2})} \sum_{r=0}^\infty \frac{e^{-\frac{\|y\|^2}{\rho\tau}}}{(\rho\tau)^r} \frac{\|y\|^{2r} r^r}{(2r)!}$$

$$B\left\{\frac{d-1}{2}, \frac{1}{2} + r\right\}, \quad //QED.$$

Lemma 2.11

Suppose $Y \sim \chi^2_{(d)}$. Then

$$\sum_{n=0}^\infty E\{Y^n\} \frac{n^n}{(2n)!} B\left\{\frac{1}{2}(d-1), \frac{1}{2} + n\right\} < \infty.$$

Proof:

$$\text{Recall } B\left\{\frac{d-1}{2}, \frac{1}{2} + n\right\} = \frac{\Gamma(\frac{d-1}{2}) \Gamma(\frac{1}{2} + n)}{\Gamma(\frac{d-1}{2} + \frac{1}{2} + n + 1)}$$

$$= \frac{\Gamma(\frac{d-1}{2}) \Gamma(\frac{1}{2} + n)}{\Gamma(\frac{d}{2} + n + 1)}$$

Using the m.g.f. of a $\chi^2_{(d)}$ r.v., we see

$$E\{Y^n\} = \frac{\frac{d}{2} (\frac{d}{2} + 1) \dots (\frac{d}{2} + n - 1)}{n!} 2^n = \frac{\Gamma(\frac{d}{2} + n)}{\Gamma(n + 1)} \cdot \frac{2^n}{\Gamma(\frac{d}{2})}$$

$$\begin{aligned} \sum_{n=0}^{\infty} E\{Y^n\} \frac{n^n}{(2n)!} B\{\frac{d-1}{2}, \frac{1}{2} + n\} \\ = \sum \frac{\Gamma(\frac{d}{2} + n)}{\Gamma(n + 1)} \cdot \frac{2^n}{\Gamma(\frac{d}{2})} \cdot \frac{n^n}{(2n)!} \cdot \frac{\Gamma(\frac{d-1}{2})}{\Gamma(\frac{d}{2} + n + 1)} \Gamma(n + \frac{1}{2}) \\ = \frac{\Gamma(\frac{d-1}{2})}{\Gamma(\frac{d}{2})} \sum \frac{(2n)^n}{(2n)!} \cdot \frac{1}{(\frac{d}{2} + n)} \cdot \frac{\Gamma(n + \frac{1}{2})}{\Gamma(n + 1)} \end{aligned}$$

From Parzen (1960), p. 162, formula (2.19)

$$\Gamma(n + \frac{1}{2}) = \frac{1 \cdot 3 \dots (2n-1)}{2^n} \sqrt{\pi} \leq (n-1)! \sqrt{\pi} = \Gamma(n) \sqrt{\pi}$$

$$\sum_{n=0}^{\infty} E\{Y^n\} \frac{n^n}{(2n)!} B\{\frac{d-1}{2}, \frac{1}{2} + n\} \leq \frac{\Gamma(\frac{d-1}{2})}{\Gamma(\frac{d}{2})} \sum \frac{(2n)^n}{(2n)!} \cdot \frac{\sqrt{\pi}}{(\frac{d}{2} + n)n}$$

$$\text{But } n^2 \cdot \frac{(2n)^n}{(2n)!} \cdot \frac{1}{n(\frac{d}{2} + n)} \leq \frac{2^n}{n!} \cdot \frac{n^n}{(2n)(2n-1) \dots (n+1)}$$

$$\leq \frac{1}{2} \cdot \frac{2^n}{n!} \rightarrow 0 \text{ as } n \rightarrow \infty$$

Therefore by the ratio test,

$$\sum_{n=0}^{\infty} E\{Y^n\} \frac{n^n}{(2n)!} B\left\{\frac{d-1}{2}, \frac{1}{2} + n\right\} < \infty.$$

//QED. Theorem 2.12

An L-time slice version of an initial Poisson field simple branching diffusion process is Brillinger mixing.

Proof:

By Theorem 2.9, it only remains to show

$$\int |x| q_2^P(\tau, a, a+x) dx \text{ and}$$

$$\int |x| |q_{(1,1)}^P(\tau_1, \tau_2, a, a+x)| dx$$

are finite. Without loss of generality $a = 0$.

$$\int_{\mathbb{R}^d} |x| q_2^p(\tau, 0, x) dx$$

$$= 2\lambda V_{\alpha} e^{-m\tau} \int_0^{\tau} \int_{\mathbb{D}^2} |x| e^{-m(\tau-s)} p(\tau-s, y, 0) p(\tau-s, y, x) dx dy ds$$

$$= 2\lambda V_{\alpha} e^{-m\tau} \frac{2^{d/2} \Gamma(\frac{d+1}{2})}{\Gamma(\frac{d-1}{2}) \Gamma(\frac{1}{2})} \int_0^{\tau} e^{-m(\tau-s)} \int_{\mathbb{D}} p(s, 0, y) \left\{ \sum_{n=0}^{\infty} \frac{e^{-||y||^2 / (\rho(\tau-s))}}{(\rho(\tau-s))^n} \right.$$

$$\left. \frac{||y||^{2n}}{(\rho(\tau-s))^n} \cdot \frac{n^n}{(2n)!} B\left\{\frac{d-1}{2}, \frac{1}{2} + n\right\} \right\} \quad (\text{by Lemma 2.10})$$

$$= 2\lambda V_{\alpha} e^{-m\tau} \frac{2^{d/2} \Gamma(\frac{d+1}{2})}{\Gamma(\frac{d-1}{2}) \Gamma(\frac{1}{2})} \int_0^{\tau} e^{-m(\tau-s)} \left\{ \sum_{n=0}^{\infty} \int_{\mathbb{D}} \left(\frac{||y||^2}{\rho(\tau-s)} \right)^n \cdot \frac{n^n}{(2n)!} \right.$$

$$\left. B\left\{\frac{d-1}{2}, \frac{1}{2} + n\right\} p(\tau-s, y, 0) dy \right.$$

(by Fubini's Theorem, and since $e^{-||y||^2 / (\rho(\tau-s))} \leq 1$)

$$= 2\lambda V_{\alpha} e^{-m\tau} \frac{2^{d/2} \Gamma(\frac{d+1}{2})}{\Gamma(\frac{d-1}{2}) \Gamma(\frac{1}{2})} \int_0^{\tau} e^{-m(\tau-s)} ds \sum_{n=0}^{\infty} E\{Y^n\} \frac{n^n}{(2n)!}$$

$$B\left\{\frac{d-1}{2}, \frac{1}{2} + n\right\}, \quad (1)$$

$$\text{since } \int_{\mathbb{D}} \left(\frac{||y||^2}{\rho(\tau-s)} \right)^n p(\tau-s, y, 0) dy = E\{Y^n\}, \quad \text{where } Y \sim \chi_{(d)}^2.$$

(1) is finite by Lemma 2.11.

$$\int_{\mathbb{R}^d} |x| q_{(1,1)}^p(\tau_1, \tau_2, 0, x) dx \quad (0 \leq \tau_1 < \tau_2)$$

$$= \lambda \int_0^{\tau_2} \int_0^{\tau_1} |x| e^{-m\tau_2} p(\tau_1, a, 0) p(\tau_2 - \tau_1, 0, x) dx da$$

$$+ 2\lambda V\alpha \int_0^{\tau_1} \iiint_{D^3} |x| e^{-ms} p(s, a, y) e^{-m(\tau_1 - s)} p(\tau_1 - s, y, 0) e^{-m(\tau_2 - s)}$$

$$p(\tau_2 - s, x) dy dx da ds .$$

$$\text{1st term} = \frac{\lambda 2^{d/2} \Gamma(\frac{d+1}{2})}{\Gamma(\frac{d-1}{2}) \Gamma(\frac{1}{2})} B(\frac{d-1}{2}, \frac{1}{2}) \cdot e^{-m\tau_2} \cdot \rho(\tau_2 - \tau_1) < \infty$$

$$\text{2nd term} = 2\lambda V\alpha \int_0^{\tau_1} \iint_{D^2} e^{ms} e^{-m\tau_1} e^{-m\tau_2} p(\tau_1 - s, y, 0) |x| p(\tau_2 - s, y, x) dx dy ds$$

$$\leq 2\lambda V\alpha \int_0^{\tau_1} e^{+ms} e^{-m\tau_1} e^{-m\tau_2} \left\{ \sum_{n=0}^{\infty} \left(\frac{\|y\|^2}{\rho(\tau_2 - s)} \right)^n \frac{n^n}{(2n)!} \right.$$

$$\left. B(\frac{d-1}{2}, \frac{1}{2} + n) p(\tau_1 - s, y, 0) dy ds \right\} ,$$

by Lemma 2.10

$0 < s < \tau_1 < \tau_2$, thus $\tau_2 - s > \tau_1 - s > 0$ and therefore

$$(\rho(\tau_2 - s))^{-n} \leq (\rho(\tau_1 - s))^{-n} \quad \text{for all } n \geq 0$$

$$\text{2nd term} \leq 2\lambda V\alpha \int_0^{\tau_1} \sum_{n=0}^{\infty} \int_{D^2} \left(\frac{\|y\|^2}{\rho(\tau_1 - s)} \right)^n p(\tau_1 - s, y, 0) dy$$

$$\frac{n^n}{(2n)!} B\left(\frac{d-1}{2}, \frac{1}{2} + n\right) \cdot e^{+ms} e^{-m(\tau_1 + \tau_2)} ds$$

$$= 2\lambda V \alpha \int_0^{\tau_1} e^{ms} ds \cdot e^{-m(\tau_1 + \tau_2)} \sum_{n=0}^{\infty} E\{Y^n\} \frac{n^n}{(2n)!} B\left(\frac{d-1}{2}, \frac{1}{2} + n\right),$$

where $Y \sim \chi_{(d)}^2$. Thus the 2^{nd} term is finite by Lemma 2.11 //QED.

We now have B-mixing in the case of an initial Poisson field. To obtain the result for the immigration case we introduce a 'delayed process'. (The idea for trying such a device came out of a conversation with Dr. Gail Ivanoff). The basic idea of the delayed process is to hold the immigrants at their arrival points for a short while, then letting them proceed at pre-determined (non-random) times. These delayed immigrants form Poisson fields.

Consider an L-time slice version of the immigration process, with times $0 < \tau_1 < \dots < \tau_L$. We need only consider those immigrants arriving before τ_L . Consider the following partition P_n of $[0, \tau_L]$; $P_n = \{t_{i,j}\}$ the open-closed intervals with endpoints $t_{i,j}$, where $t_{i,j} = \tau_{i-1} + \frac{j(\tau_i - \tau_{i-1})}{n}$, $j = 1, \dots, n$, $i = 1, \dots, L$, $\tau_0 = 0$. Consider an immigrant (a, t) ; that is an immigrant arriving at $a \in R^d$ at time t . Almost surely, we can speak of 'the' immigrant at (a, t) . Anyway, t determines i such that $\tau_{i-1} < t \leq \tau_i$, and $1 \leq j \leq n$ $t_{i,j-1} < j \leq t_{i,j}$ or $\tau_{i-1} < t \leq t_{i,1}$. The delay mechanism acts on the family tree of (a, t) in the following way.

Let $X_{(a,t)}(u,w)$ be the w^{th} sample family tree at time u for the immigrant $(a(w), t(w))$. $X_{(a,t)}(u,w) \in R^{d,d(w)}$ where $d(w)$ is the number in the 'family' at time u , with $R^{d,d(w)} = \phi$ if $d(w) = 0$. Define the delay with respect to the partition Π_n

$$Y_{(a,t)}^{(n)}(u,w) = \begin{cases} \phi & \text{if } u < t \\ X_{(a,t)}(t,w) & \text{if } t_{i,j-1} < u \leq t_{i,j} \text{ or } \tau_{i-1} < t_{i,1} \\ X_{(a,t)}(u - (t_{i,j} - t), w) & \text{if } u > t_{i,j} \end{cases}$$

where $t_{i,j}$ is determined. $Y_{(a,t)}^{(n)}$ is just $X_{(a,t)}$ delayed in time an amount $t_{i,j} - t$. As $n \rightarrow \infty$, the delay time tends to zero. Note, if $X_{(a,t)}(u,w) = \phi$ for some $u > t$, then $Y_{(a,t)}^{(n)}(v,w) = \phi$ for all $v > u + (t_{i,j} - t)$.

Figure 1 is a picture of a delayed process for the simple branching diffusion immigration process in R^1 , with immigrant at (a_0, t_1) . The motion is represented as straight lines for convenience. In figure 1, the original family tree is delayed by an amount δ . Given times $0 < \tau_1 < \dots < \tau_L$ and a compact set $B \subset R^d$, there is no branch in B at time τ_L almost surely. Thus figure 1 is the most general 'picture' that need be considered. A notion of convergence of family trees is needed. For this a Skorohod type metric is quite reasonable (Billingsley (1968), first several pages of chapter 3).

Let $S = \bigcup_{n=0}^{\infty} (R^d)^n$, and

FIGURE 1

- - on line T , these represent the original process
- x - on line T , these represent the delayed process for this immigrant (a_0, t_0) .

$E = \{x | x: [0, \infty] \rightarrow S, x \text{ right continuous with left hand limits}\}.$

Family trees are in E . Let $n = (n_1, n_2, \dots)$ and $T = (t_1, t_2, \dots)$ be the sequence of population sizes and the jump times respectively.

In figure 1, for the original tree these are, up to time τ ,

$$n = (0, 1, 2, 3, 4, 3, 2, 3, 4, 3, 4, \dots)$$

$$T = (t_1, t_2, t_3, t_4, t_5, t_6, t_7, t_8, t_9, t_{10}, \dots)$$

For a given $\tau > 0$, let Λ_τ be the set of continuous 1-1 maps on $[0, \tau]$.

For given τ , let n_τ , and T_τ be n and T defined only up to time τ .

Let $x, y \in E$. If $\dim x(t) = n = \dim y(t)$,

$$|x(t) - y(t)| = \begin{cases} 0 & \text{if } n = 0 \\ ||x(t) - y(t)||_n & \text{if } 0 < n < \infty \end{cases}$$

where, for $n \neq 0$, $x = (a_1, \dots, a_n)$, $y = (b_1, \dots, b_n)$,

$$||x - y||_n = \min_{\rho} \sum_{j=1}^n ||a_j - b_{\rho(j)}||$$

and when $|| \cdot ||$ is the usual Euclidean distance, and the minimum is over all permutations on $\{1, \dots, n\}$.

If $\dim x(t) \neq \dim y(t)$, then define

$$|x(t) - y(t)| \equiv 1.$$

A Skorohod type metric can now be defined by

$$d_{\tau}(x, y) = \begin{cases} \inf_{\varepsilon > 0} \left\{ \sup_{0 \leq t \leq \tau} |\lambda t - t| \leq \varepsilon, \sup_{0 \leq t \leq \tau} |x(t) - y(\lambda t)| \leq \varepsilon, \lambda \in \Lambda_{\tau} \right\} & \text{if } n_{\tau}(x) = n_{\tau}(y) \\ 1 & \text{otherwise} \end{cases}$$

It is clear that if $Y_{(a,t)}^{(n)}$ is the delayed family tree of $X_{(a,t)}$, the tree of the immigrant (a,t) , and $X_{(a,t)}$ has no branch at τ_L , then

$$d_{\tau}(X_{(a,t)}, Y_{(a,t)}^{(n)}) \rightarrow 0 \text{ as } n \rightarrow \infty$$

by the delay definition. In fact, for delay $\delta < \delta_0$ where δ_0 is small enough so that $\dim_{(\delta)} X_{(a,t)} = \dim_{(\delta)} X_{(a,t)}$ for all $0 < \delta \leq \delta_0$, where $X_{(a,t)}$ is the tree $X_{(a,t)}$ delayed by an amount δ , the function λ given by

$$\lambda(s) = \begin{cases} s(\frac{t}{t+\delta}), & 0 \leq s \leq t+\delta \\ (s-t_{i+1})\left(\frac{t_i-t_{i+1}}{t_i+\delta-t_{i+1}}\right)+t_{i+1}, & \text{if } t_i \leq s \leq t_{i+1} \\ (s-\tau)\left(\frac{t_r-\tau}{t_r+\delta-\tau}\right)+\tau, & \text{if } t_r \leq s \leq \tau \end{cases}$$

where $T_{\tau}(X(a,t)) = (t_1, t_2, \dots, t_r)$, gives $d_{\tau}(X(a,t), X^{(\delta)}(a,t)) \rightarrow 0$ as $\delta \downarrow 0$.

For each bounded Borel set $B \subseteq R^d$, let $N_{\tau_i}^{(n)}(B)$ be the number of individual in B at time τ_i for the delayed process $\gamma^{(n)}$, where $\gamma^{(n)}$ is obtained from the immigration process by delaying each immigrant according to the partition P_n . It easily follows that

$$(N_{\tau_1}^{(n)}(B), \dots, N_{\tau_L}^{(n)}(B)) \xrightarrow{\text{a.s.}} (N_{\tau_1}^I(B), \dots, N_{\tau_L}^I(B)) \text{ as } n \rightarrow \infty$$

in the Skorohod type metric d_{τ_L} since for each sample w , there are only finitely many immigrants to consider. Notice

$$N_{\tau_i}^{(n)}(B) = \sum_{\ell=1}^i \sum_{j=1}^n N_{\tau_i, \ell, j}^{(n)}(B)$$

where $N_{\tau_i, \ell, j}^{(n)}(B)$ is the number of individuals in B of those delayed immigrants arriving in $(t_{\ell, j-1}, t_{\ell, j}]$ if $j \neq 1$ and $(t_{\ell-1}, t_{\ell, 1}]$ if $j = 1$, $\ell = 1, \dots, i-1$. The descendants of immigrants arriving in separate intervals behave independently. The $N_{\tau_i, \ell, j}^{(n)}$ are thus independent initial Poisson field simple branching diffusion processes, with Poisson intensity $n \frac{(\tau_{\ell+1} - \tau_{\ell})}{n}$.

Almost surely there are no branches at $\tau_1, \dots, \tau_L$ in B . Thus it is clear by definition, for each $B_1, \dots, B_L$

$$(N_{\tau_1}^{(n)}(B_1), \dots, N_{\tau_L}^{(n)}(B_L)) \rightarrow (N_{\tau_1}^I(B_1), \dots, N_{\tau_L}^I(B_L)) \text{ a.s.}$$

Now consider cumulants.

$$\begin{aligned}
 (N_{\tau_1}^{(n)}(B_1), \dots, N_{\tau_L}^{(n)}(B_L)) &= \left(\sum_{j=1}^n N_{\tau,1,j}^{(n)}(B_1), \dots, \sum_{j=1}^n N_{\tau_L,1,j}^{(n)}(B_L) \right) \\
 &+ (0, \sum_{j=1}^n N_{\tau,2,j}^{(n)}(B_2), \dots, \sum_{j=1}^n N_{\tau_L,2,j}^{(n)}(B_L)) \\
 &+ \dots \\
 &+ (0, \dots, 0, \sum_{j=1}^n N_{\tau_L,j}^{(n)}(B_L)) ,
 \end{aligned}$$

where the right hand side is the sum of independent processes. A joint ℓ^{th} factorial cumulant of the left hand side is then given by (with $n_{j_1} + \dots + n_{j_k} = \ell, 1 \leq k \leq L$)

$$\begin{aligned}
 \int \dots \int_{\prod B_t} \sum_{1 \leq j} \sum_{j=1}^n q_{(n_{j_1}, \dots, n_{j_k})}^p \left(\tau_{j_1} - \frac{j(\tau_{j_1} - \tau_{j_1-1})}{n}, \dots, \tau_{j_k} - \frac{j(\tau_{j_k} - \tau_{j_k-1})}{n} \right. \\
 \left. a_{1,1}, \dots, a_{1,n_{j_1}}, a_{2,1}, \dots, a_{k,1}, \dots, a_{k,n_{j_k}} \right) \\
 da_{1,1} \dots da_{k,n_{j_k}} \cdot r \frac{(\tau_{j_i} - \tau_{j_i-1})}{n} \quad (*)
 \end{aligned}$$

By Theorem 2.9, * converges to the finite expression

$$\begin{aligned}
 \int \dots \int_{\prod B_t} \int_0^{\tau_{j_1}} q_{(n_{j_1}, \dots, n_{j_k})}^p (\tau_{j_1} - s, \dots, \tau_{j_k} - s, a_{1,1}, \dots, a_{1,n_{j_1}}, \dots, a_{k,1}, \dots, \\
 a_{k,n_{j_k}}) ds da_{1,1} \dots da_{k,n_k} .
 \end{aligned}$$

Since the $N_{\tau_i}^{(n)}(B_i) \rightarrow N_{\tau_i}^I(B_i)$ a.s. and since $*$ is bounded, then by Chung (1974, Theorem 4.5.2, p. 95)

$$\begin{aligned} & \int \dots \int_{\prod_{j=1}^n B_j} q_{(n_1, \dots, n_L)}^I(\tau_1, \dots, \tau_L, a_{1,1}, \dots, a_{L,n_L}) da_{1,1} \dots da_{L,n_L} \\ &= \left(\int_{\prod_{j=1}^n B_j} \int_{n_j} \int_0^{\tau_{j_1}} q_{(n_1, \dots, n_L)}^P(\tau_1-s, \dots, \tau_L-s, a_{1,1}, \dots, a_{L,n_L}) \right. \\ & \quad \left. ds da_{1,1} \dots da_{L,n_L} \right) \end{aligned}$$

where $j_1 = \min_i \{n_i \neq 0\}$. Since this holds for each Borel set,

$$\begin{aligned} q_{(n_1, \dots, n_L)}^I(\tau_1, \dots, \tau_L, a_{1,1}, \dots, a_{L,n_L}) &= r \int_0^{\tau_{j_1}} q_{(n_1, \dots, n_L)}^P(\tau_1-s, \dots, \tau_L-s, a_{1,1}, \dots, a_{L,n_L}) ds. \end{aligned} \quad (**)$$

Theorem 2.13

The L-time slice version of the immigration process is B-mixing.

Proof:

By **

$$\begin{aligned} & \int \dots \int_{D^{L-1}} q_{(n_1, \dots, n_L)}^I(\tau_1, \dots, \tau_L, a_{1,1}, \dots, a_{L,n_L}) da_{1,1} \dots da_{L,n_L-1} \\ &= r \int_0^{\tau_{j_1}} \int_{D^{L-1}} q_{(n_1, \dots, n_L)}^P(\tau_1-s, \dots, \tau_L-s, a_{1,1}, \dots, a_{L,n_L}) da_{1,1} \dots da_{L,n_L-1} ds. \end{aligned}$$

The Theorem is now easily proven using Theorem 2.1 and the normal characteristic function, just as in Theorem 2.12. //QED.

The L-time slice immigration process has spectral representation and spectral densities of all orders. We calculate the second order spectral densities, $0 < \tau_1 < \tau_2$.

$$f_{\tau_1, \tau_1}(\lambda) = (2\pi)^{-d} \int_{R^d} e^{-i\langle \lambda, x \rangle} \text{Cov}\{dN_{\tau_1}(y), dN_{\tau_1}(x+y)\} dx$$

$$= (2\pi)^{-d} \{rM_{(\tau_1, m)} + \int_{R^d} e^{-i\langle \lambda, x \rangle} q_2^I(\tau_1, 0, x) dx\}$$

$$\int_{R^d} e^{-i\langle \lambda, x \rangle} q_2^I(\tau_1, 0, x) dx$$

$$2V\alpha r \int_0^{\tau_1} \int_0^s \iiint_{R^{3d}} e^{-i\langle \lambda, x \rangle} e^{-\mu u} p(u, a, y) e^{-2m(s-u)} p(s-u, y, 0)$$

$$p(s-u, y, x) da dy dx du ds$$

$$= 2\text{Var} \int_0^{\tau_1} \int_0^s \iint_{\mathbb{R}^{2d}} e^{-mu} e^{-2m(s-u)} p(s-u, y, 0) e^{-i\langle \lambda, x \rangle} p(s-u, y, x) dx dy du ds$$

Recall $p(s-u, y, x)$ is the $N_d(y, \rho(s-u)I)$ density with characteristic function $\phi(\lambda) = \exp\{i\langle y, \lambda \rangle - \frac{1}{2} \rho(s-u) |\lambda|^2\}$.

Therefore,

$$\begin{aligned} & \int_{\mathbb{R}^d} e^{-i\langle \lambda, x \rangle} q_2^I(\tau_1, 0, x) dx \\ &= 2\text{Var} \int_0^{\tau_1} \int_0^s \int_{\mathbb{R}^d} e^{-mu} e^{-2m(s-u)} e^{+i\langle \lambda, y \rangle} e^{-\frac{1}{2} \rho(s-u) |\lambda|^2} p(s-u, y, 0) dy du ds \\ &= 2\text{Var} \int_0^{\tau_1} \int_0^s e^{-ms} e^{-(m+\rho|\lambda|^2)u} du ds, \quad (u \rightarrow s-u) \end{aligned}$$

Therefore,

$$(2\pi)^d f_{\tau_1, \tau_1}(\lambda) = \begin{cases} rM_{(\tau_1, m)} + 2\text{Var} \int_0^{\tau_1} s e^{-ms} ds & \text{if } m + \rho|\lambda|^2 = 0 \\ rM_{(\tau, m)} + 2\text{Var} \frac{\{M_{(\tau_1, m)} - M_{(\tau_1, 2m+\rho|\lambda|^2)}\}}{m + \rho|\lambda|^2} & \text{if } m + \rho|\lambda|^2 \neq 0 \end{cases}$$

For $\tau_1 < \tau_2$

$$(2\pi)^d f_{\tau_1, \tau_2}(\lambda) = r \int_{R^d} e^{-i\langle \lambda, x \rangle} \int_0^{\tau_1} \int_{R^d} p_{(1,1)}(\tau_1-s, \tau_2-s, a, 0, x) da ds dx$$

$$= r \int_0^{\tau_1} \iint_{R^{2d}} e^{-i\langle \lambda, x \rangle} e^{-m(\tau_1-s)} p(\tau_1-s, a, 0) P_1(\tau_2-s, 0, x) da dx ds$$

$$+ 2Var \int_0^{\tau_1} \iint_{R^{2d}} \int_0^{\tau_1-s} \int_{R^d} e^{-mu} e^{-i\langle \lambda, x \rangle} p(u, a, y) P_1(\tau_1-s-u, y, a)$$

$$P_1(\tau_2-s-u, y, x) dy du da dx ds$$

$$= e^{-\frac{1}{2}(2m + \rho|\lambda|^2)(\tau_2-\tau_1)} (2\pi)^d f_{\tau_1, \tau_1}(\lambda)$$

by using the normal characteristic function.

Theorem 2.14

For the immigration process, $0 < \tau_1 < \tau_2$

$$f_{\tau_1, \tau_1}(\lambda) = \begin{cases} (2\pi)^{-d} \{ rM_{(\tau_1, m)} + 2Var \int_0^{\tau_1} s e^{-ms} ds \} & \text{if } m + \rho|\lambda|^2 = 0 \\ (2\pi)^{-d} \{ rM_{(\tau_1, m)} + 2Var \frac{[M_{(\tau_1, m)} - M_{(\tau_1, 2m+\rho|\lambda|^2)}]}{m + \rho|\lambda|^2} \} & \text{if } m + \rho|\lambda|^2 \neq 0 \end{cases}$$

and

$$f_{\tau_1, \tau_2}(\lambda) = e^{-\frac{1}{2}(2m + \rho|\lambda|^2)(\tau_2 - \tau_1)} f_{\tau_1, \tau_1}(\lambda).$$

We can obtain some additional Brillinger type mixing conditions for the time slice immigration process. Each condition also holds for the initial Poisson field case as well.

Theorem 2.15

For the one time slice immigration process.

$$c_n = r \int \dots \int_{R^{d(n-1)}} q_n^I(\tau, x_1, \dots, x_{n-1}, 0) dx_1 \dots dx_{n-1}$$

$$= (n-1)! r (V\alpha M_{(\tau, m)})^{n-1} M_{(\tau, m)}, \quad n = 1, 2, \dots;$$

and so

$$r \sum_{n=1}^{\infty} \frac{1}{n!} c_n z^n = r \sum_{n=1}^{\infty} \frac{(V\alpha M_{(\tau, m)} z)^{n-1}}{n} \cdot M_{(\tau, m)} z = - \frac{r \log(1 - V\alpha M_{(\tau, m)} z)}{V\alpha}$$

for all $|z| < (V\alpha M_{(\tau, m)})^{-1}$

Proof: This follows directly from Theorem 2.8 (ii). //QED.

Theorem 2.16

For each $\ell \in N_0$,

$$\int \dots \int_{R^{d(n-1)}} |b_{i,j}|^{\ell} q_{(n_1, \dots, n_L)}^I(\tau_1, \dots, \tau_L, b_{1,1}, \dots, b_{L,1}, \dots, b_{L,n_L-1})$$

$$db_{1,1} \dots db_{L,n_L-1} < \infty$$

if $n_L = 1$, $(i,j) \neq (L,n_L)$, and

$$\int \dots \int_{\mathbb{R}^{d(n-1)}} |b_{L,1}|^{\ell} q_{(n_1, \dots, n_L)}(\tau_1, \dots, \tau_L, b_{1,1}, \dots, b_{1,n_1}, \dots, b_{L-1,n_{L-1}-1},$$

$$0, b_{L,1}) db_{1,1} \dots db_{L-1,n_{L-1}-1} db_{L,1} < \infty$$

if $n_L = 1$,

where $n_1 + \dots + n_L = n$.

Proof:

The case $\ell = 0$ by B-mixing. The case $\ell = 1$ is a similar induction on L and $n_1, \dots, n_L$. The theorem then follows by induction on ℓ beginning at $\ell = 2$. //QED.

§3 Some Results in Estimation for Stationary Point Processes.

Introduction and Summary

In Section 2, it is shown that the simple branching diffusion process at L times $0 < \tau_1 < \dots < \tau_L$ (this is referred to as an L -time slice immigration process) is an L dimensional point process on R^d with no multiple points. It also satisfies some Brillinger type mixing conditions. This chapter considers some of Professor Brillinger's work which develops various central limit theorems for the Brillinger type mixing processes. The proofs are basically Brillinger's, but with the details completed. A strong law of large numbers (SLLN) for these processes is also obtained. As immediate corollaries, we then obtain a CLT and SLLN for the simple branching diffusion process. Under some stronger mixing conditions, Brillinger obtains an almost sure growth rate bounded by $T^d \log T$. This allows a Glivenko-Cantelli type result to be obtained for Fourier transforms of processes satisfying such mixing conditions. The notion of a periodogram is also introduced.

Let $\{X(t): t \in R^d\}$ be a spatial point process on R^d with no multiple points, $X(t) = (X_1(t), \dots, X_L(t))$, $L \geq 1$. We may define an $(n_1 + \dots + n_L)^{th}$ order intensity function by

$$P\{\text{particles of type } a \text{ in } S(b_{a,i}, h_{a,i}), i = 1, \dots, n_a, a = 1, \dots, L\} \\ \sim P_{(n_1, \dots, n_L)}(b_{1,1}, \dots, b_{1,n_1}, b_{2,1}, \dots, b_{L,1}, \dots, b_{L,n_L}) \prod_{a=1}^L \prod_{i=1}^{n_a} (h_{a,i})^d$$

where $S(b, h) = \{x: \|x - b\|_\infty < \frac{h}{2}\}$ is a cube centred at b , with sides of length h . Let $N_a(A)$ = number of type a particles in A .

$$\therefore N_a(A) = \int_A \chi_a(t) dt, \text{ and}$$

$$Y^{(n)} = \begin{cases} Y(Y-1) \dots (Y-n+1) & \text{if } n \geq 1 \\ 1 & \text{if } n = 0 \end{cases}$$

$Y^{(n)}$ is known as " Y to the n factors". Then

$$E\{N_1(A)^{(n_1)} \dots N_L(A)^{(n_L)}\} = \int \dots \int_{n_1 + \dots + n_L} p(n_1, \dots, n_L)(a_{1,1}, \dots, a_{1,n_1}, a_{L,1}, \dots, a_{L,n_L}) \\ \cdot da_{1,1}, \dots, da_{L,n_L}$$

A probability generating functional is defined for suitable

$\xi_j: R^d \rightarrow R$, by

$$\begin{aligned} G[\xi_1, \dots, \xi_L] &= E\{\exp\left[\sum_{a=1}^L \int_{R^d} \log \xi_a(x) dN_a(x)\right]\} \\ &= E\left\{\prod_{a=1}^L \prod_{\substack{\text{type } a \\ \text{occurrence} \\ \text{at } x_a}} [1 + (\xi_a(x_a) - 1)]\right\} \\ &= \sum_{n_1, \dots, n_L} \frac{1}{n_1! \dots n_L!} \int_{R^d} \dots \int_{n_1 + \dots + n_L} \prod_{a=1}^L \prod_{j=1}^{n_a} (\xi_a(x_{a,j}) - 1) \\ &\quad \cdot P(n_1, \dots, n_L)(x_{1,1}, \dots, x_{1,n_1}, x_{2,1}, \dots, x_{L,n_L}) \\ &\quad \cdot dx_{1,1}, \dots, dx_{L,n_L} \end{aligned}$$

Let A_a be measurable, $z_a > 0$, $a = 1, \dots, L$ and

$$\xi_a(x) = \begin{cases} z_a & \text{if } x \in A_a \\ 1 & \text{if } x \in A_a^c \end{cases}.$$

$$G[\xi_1, \dots, \xi_L] = E\{\exp[\sum_{a=1}^L \log z_a \int_{A_a} dN_a(x)]\} = E\{\prod_{a=1}^L z_a^{N_a(A_a)}\}.$$

We may define factorial cumulants $q_{(n_1, \dots, n_L)}(x_{1,1}, \dots, x_{L,n_L})$ through the factorial cumulant generating functional

$$\begin{aligned} \log G[\xi_1, \dots, \xi_L] = & \sum_{n_1, \dots, n_L} \frac{1}{n_1! \dots n_L!} \int_{\mathbb{R}^{d(n_1 + \dots + n_L)}} \prod_{a=1}^L \prod_{j=1}^{n_a} (\xi_a(x_{a,j}) - 1) \\ & \cdot q_{(n_1, \dots, n_L)}(x_{1,1}, \dots, x_{1,n_1}, x_{2,1}, \dots, x_{L,n_L}) \\ & \cdot dx_{1,1}, \dots, dx_{L,n_L}. \end{aligned}$$

Example: For a Poisson process on $\mathbb{R}^d$, mean intensity λ

$$G[\xi] = \exp[\lambda \int (\xi(x) - 1) dx].$$

Notice that the relationship between factorial cumulants and factorial moments is the same as that between ordinary cumulants and ordinary moments, since the formal relationship, via generating

functionals is the same.

For a stationary process with the obvious abuse of notation ,

$$p_a(x) = q_a(x) = \text{constant} .$$

$$p_{(n_1, \dots, n_L)}(x_{1,1}, \dots, x_{L,n_L}) = p_{(n_1, \dots, n_L)}(x_{1,1} - x_{L,n_L}, \dots, x_{L,n_L} - 1 - x_{L,n_L}) ,$$

$$q_{(n_1, \dots, n_L)}(x_{1,1}, \dots, x_{L,n_L}) = q_{(n_1, \dots, n_L)}(x_{1,1} - x_{L,n_L}, \dots, x_{L,n_L} - 1 - x_{L,n_L}) .$$

We also sometimes use the alternate notation $p_{a_1, \dots, a_\ell}(x_1, \dots, x_\ell)$ for an ℓ^{th} order intensity function of type a_j particles at x_j , this alternate notation is also used for the factorial cumulants.

Definition 3.1 (Same as chapter two)

The stationary point process X on R^d is said to be Brillinger mixing, or B-mixing if

$$(a) \quad \forall \quad n_1, \dots, n_L \in N_0, \quad \Rightarrow \quad n_1 + \dots + n_L = \ell$$

$$\int \dots \int_{R^{d(\ell-1)}} |q_{(n_1, \dots, n_L)}(x_{1,1}, \dots, x_{L,n_L-1}, 0)| dx_{1,1}, \dots, dx_{L,n_L-1} < \infty ,$$

and

$$(b) \quad \text{if } n_1 + \dots + n_L = 2$$

$$\int_{\mathbb{R}^d} |u| |q_{(n_1, \dots, n_L)}(u)| du < \infty.$$

Comment. We call this Brillinger mixing even though this may not be historically correct.

Definition 3.2 (Brillinger (1975), Assumption 2.6.3)

The B-mixing point process is said to satisfy the generating function condition if

$$\sum_{k=1}^{\infty} q_k \frac{z^k}{k!} < \infty \quad \text{for } |z| \leq |z_0| \text{ for some } z_0 \neq 0$$

where $q_k = \sup_{a_1, \dots, a_k} \int \dots \int_{\mathbb{R}^{d(k-1)}} |q_{a_1, \dots, a_k}(x_1, \dots, x_{k-1})| dx_1, \dots, dx_{k-1}$.

There is a stronger mixing condition which is of use to us.

Definition 3.3 The B-mixing point process X is said to be Brillinger ℓ -mixing for an $\ell \in \mathbb{N}_0$ if its factorial cumulants also satisfy

$$\int \dots \int_{\mathbb{R}^{d(n-1)}} |u_i|^\ell |q_{a_1, \dots, a_n}(u_1, \dots, u_{n-1})| du_1, \dots, du_{n-1} < \infty \quad (1)$$

where $\sum a_i = n$.

Theorem 2.16 shows that the immigration L-time slice process is Brillinger ℓ -mixing for each $\ell \in \mathbb{N}_0$.

Let $c_{a_1, \dots, a_n}$ be an ordinary n^{th} order cumulant density of X . It is related to factorial cumulant densities by the same linear relation as ordinary and factorial cumulants are. It then follows

$$\int \dots \int_{\mathbb{R}^{d(n-1)}} |u_i|^{\ell} |c_{a_1, \dots, a_n}(u_1, \dots, u_{n-1})| du_1 \dots du_{n-1} < \infty$$

if X is Brillinger ℓ -mixing.

A B-mixing point process X has a spectral representation and spectra of all orders. It has a second order spectral density

$$f_{ab}(\lambda) = (2\pi)^{-d} \int_{\mathbb{R}^d} \text{cov}\{X_a(t+u), X_b(t)\} \exp\{-i\langle \lambda, u \rangle\} du$$

where

$$\text{Cov}\{X_a(t+u), X_b(t)\} = p_a \cdot \delta_a(b) \cdot \delta(u) + \{p_{ab}(u) - p_a p_b\}$$

where $\delta_a(b) = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{if } a \neq b \end{cases}$ and $\delta(u)$ is the Dirac delta function.

Therefore

$$f_{ab}(\lambda) = \begin{cases} (2\pi)^{-d} [p_a + \int_{R^d} [p_{aa}(u) - p_a^2] \exp\{-i\langle \lambda, u \rangle\} du] & \text{if } a = b \\ (2\pi)^{-d} \int_{R^d} [p_{ab}(u) - p_a p_b] \exp\{-i\langle \lambda, u \rangle\} du & \text{if } a \neq b \end{cases}$$

Let $f_{XX}(\lambda) = (f_{ab}(\lambda))$.

For the immigration process, with indices $\tau_1, \dots, \tau_L$, the second order spectral densities are given in Proposition 2.14.

Let $N_a(A)$ be the number of type a particles in A , $1 \leq a \leq L$.

We always assume A to be Lebesgue measurable and of finite positive measure. $N_{\tau_a}^I(A)$ is the number of particles in A at time τ_a for the immigration process, $1 \leq a \leq L$, $0 < \tau_1 < \dots < \tau_L$.

Let $N(A) = (N_1(A), \dots, N_L(A))$. For the remainder of this chapter, we assume X (and equivalently N) are B-mixing.

Theorem 3.1

(i) (Brillinger (1974), Theorem 1)

If we only assume condition (a) of B-mixing, then if $A \uparrow R^d$ through convex sets

$$|A|^{-1/2}(N(A) - EN(A)) \xrightarrow{D} N(0, (2\pi)^d f_{XX}(0)).$$

(ii) (Brillinger (1970), Theorem 3.2)

If in addition we assume condition (b) of B-mixing, and consider sets $A_T = T A$, $T > 0$, and $|A| > 0$ then if A satisfies

$$\int |\chi_A(t+u) - \chi_A(t)| dt \leq K|u|, \quad \text{for some constant } K(A),$$

$$\text{then } T^{-d/2}|A|^{-1/2}(N(A_T) - EN(A_T)) \xrightarrow{D} N(0, (2\pi)^d f_{XX}(0)) \text{ as } T \rightarrow \infty.$$

Proof:

Consider a k^{th} factorial cumulant of $\{N_1(A), \dots, N_L(A)\}$,
 $n_1 + \dots + n_L = k$.

$$|\text{fac. cum}_k \{ \underbrace{N_1(A), \dots, N_1(A)}_{n_1 \text{ times}}, \dots, \underbrace{N_L(A), \dots, N_L(A)}_{n_L \text{ times}} \}|$$

$$\leq \int \dots \int_{A^k} |q_{(n_1, \dots, n_L)}(u_1, \dots, u_k)| du_1, \dots, du_k$$

$$\leq \int_A \left\{ \int \dots \int_{\mathbb{R}^{d(k-1)}} |q_{(n_1, \dots, n_k)}(u_1, \dots, u_{k-1}, 0)| du_1, \dots, du_{k-1} \right\} du_k$$

by translation invariance

$$= |A| \cdot \text{const.}$$

An ordinary k^{th} cumulant related to a factorial cumulant through a linear (in $q_{(n_1, \dots, n_L)}$) relationship (cf Kendall and Stuart (1969), p. 75) so that an ordinary k^{th} cumulant of $\{N_1(A), \dots, N_L(A)\}$ is $O(|A|)$.

Thus

$$\text{cum}_{(n_1, \dots, n_L)} \{ \underbrace{|A|^{-\frac{k}{2}} N_1(A), \dots, |A|^{-\frac{k}{2}} N_1(A)}_{n_1 \text{ times}}, \dots, \underbrace{|A|^{-\frac{k}{2}} N_L(A), \dots, |A|^{-\frac{k}{2}} N_L(A)}_{n_L \text{ times}} \}$$

$$= |A|^{-k/2} O(|A|)$$

$$= O(|A|^{1 - k/2}) \rightarrow 0 \quad \text{if } k \geq 3 \text{ as } |A| \rightarrow \infty.$$

Similarly the normalized k^{th} cumulants in (ii) tend to zero as $T \rightarrow \infty$ if $k \geq 3$.

For (i), we now evaluate the second moments.

$$\begin{aligned} & \text{Cov}\{|A|^{-\frac{k}{2}}(N_a(A) - EN_a(A)) |A|^{-\frac{k}{2}}(N_b(A) - EN_b(A))\} \\ &= |A|^{-1} \int \int_{A^2} \text{cov}\{X_a(t), X_b(s)\} ds dt \\ &= p_a \delta(a_1 - b_1) + |A|^{-1} \int_A \int_{A-s} [p_{a,b}(u) - p_a p_b] du ds, \quad u \neq 0 \end{aligned}$$

Let $\epsilon > 0$. Let $\alpha \in (0, 1)$ be a function of ϵ to be given below.

If $y \in \alpha A$, then $(1 - \alpha)A \subseteq A - y$, since A is convex.

$(1 - \alpha)A \uparrow \mathbb{R}^d$ as A does. Since $\int_{\mathbb{R}^d} |c_{ab}(u)| du < \infty$, then for A sufficiently large,

$$\left| \int_{A-y} c_{ab}(u) du - (2\pi)^d f_{ab}(0) \right| < \epsilon, \text{ where } c_{ab}(u) = p_{ab}(u) - p_a p_b.$$

Also

$$|A|^{-1} \left| \int_{A \setminus \alpha A} \int_{A-y} c_{ab}(u) du dy \right| \leq \text{const.} \frac{|A \setminus \alpha A|}{|A|} = \text{const.} (1 - \alpha^d) < \epsilon \quad (1)$$

for $\alpha(\epsilon)$ sufficiently close to 1. Thus for A sufficiently large

$$|A|^{-1} \text{Cov}\{N_a(A), N_b(A)\} \leq |A|^{-1} \int_{\alpha A} [(2\pi)^d f_{ab}(0) + \epsilon] dy + \epsilon$$

$$\leq \alpha^d (2\pi)^d f_{ab}(0) + (1 + \alpha^d) \epsilon$$

Similarly $|A|^{-1} \text{Cov}\{N_a(A), N_b(A)\} \geq \alpha^d (2\pi)^d f_{ab}(0) - (1 + \alpha^d) \epsilon$ for A sufficiently large. Since $\epsilon \rightarrow 0$ implies $\alpha \rightarrow 1$, and ϵ is arbitrary we see

$$|A|^{-1} \text{Cov}\{N_a(A), N_b(A)\} \rightarrow (2\pi)^d f_{ab}(0).$$

For (ii), the covariance is calculated in similar manner to that in Theorem 3.3. There it is seen that the translation condition in the theorem allows us to dispense with a condition such as convexity. //QED.

Letting $\hat{p}_a = \frac{N_a(A)}{|A|}$, Theorem 3.1 says $\hat{p}_a$ is an asymptotically normal unbiased estimate of the mean intensity.

In order to have the given rate of convergence $|A|^{-1/2}$ and the nice limiting covariance, it is clear that some conditions, such as those given, are needed in addition to just $|A| \rightarrow \infty$, as the following simple example shows. The example is on the real line, but it is

clear that it generalizes in an obvious way to $\mathbb{R}^d$.

Example: Suppose $|A_T| = T$, $T \in \mathbb{N}_0, T \geq 3$. A_T will be a slightly "fuzzy" first T^2 integers.

$$A_T = \bigcup_{n=1}^{T^2} \left[n - \frac{1}{2T}, n + \frac{1}{2T} \right]$$

$$|A_T| = T^2 \cdot \frac{1}{T} = T.$$

$$\text{If } s \in A_T, A_T - s \subseteq \bigcup_{n=-2T^2}^{2T^2} \left[\frac{1}{n} - \frac{1}{T}, \frac{1}{n} + \frac{1}{T} \right] = B_T, \quad |B_T| = 4T^2 \cdot \frac{1}{2T} = 2T.$$

c_{aa} is an integrable function, so

$$|A_T|^{-1} \left| \int \int_{A_T} c_{aa}(u-v) du dv \right| = |A_T|^{-1} \left| \int \int_{A_T \times A_T - v} c_{aa}(a) dx dv \right|$$

$$\leq |A_T|^{-1} \int \int_{A_T \times B_T} |c_{aa}(x)| dx dv = \int_{B_T} |c_{aa}(x)| dx.$$

$$\text{Let } \varepsilon > 0, \quad \exists N_0 > 0 \quad \text{s.t.} \quad \int_{\mathbb{R} \setminus [-N, N]} |c_{aa}(x)| dx < \frac{\varepsilon}{2} \quad \forall N \geq N_0.$$

$$|B_T \setminus [-N_0, N_0]| \leq \frac{N_0}{T} \rightarrow 0 \quad \text{as } T \rightarrow \infty.$$

Therefore by Proposition 13, p. 85, Royden (1968), $\exists T_0$ s.t.

$$\int_{B_T \setminus [-N_0, N_0]} |c_{aa}(x)| dx < \frac{\varepsilon}{2} \quad \forall T \geq T_0.$$

$$|A_T|^{-1} \left| \int \int_{A_T} c_{aa}(u-v) du dv \right| \leq \int_{B_T \cap [-N_0, N_0]} |c_{aa}(x)| dx + \int_{[-N_0, N_0]^c} |c_{aa}(x)| dx$$

$$< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \quad \forall T \geq T_0.$$

Thus $\lim_{T \rightarrow \infty} |A_T|^{-1} \int \int_{A_T} c_{aa}(u-v) du dv = 0 \neq (2\pi)^d f_{aa}(0).$

Corollary to Theorem 3.1 (Includes Ivanoff (1976), Theorem 1)

For the immigration process, if $0 < \tau_1 < \dots < \tau_L$, then under the conditions of Theorem 3.1, as $A \uparrow \mathbb{R}^d$

$$|A|^{-1} \left(\left(\frac{N_{\tau_1}(A)}{|A|}, \dots, \frac{N_{\tau_L}(A)}{|A|} \right) - r(M_{(\tau_1, m)}, \dots, M_{(\tau_L, m)}) \right)$$

$$\xrightarrow{D} N_L(0, (2\pi)^d f_{XX}(0)),$$

where

$$(2\pi)^d f_{\tau_1, \tau_1}(\lambda) = \begin{cases} rM_{(\tau_1, m)} + 2\text{Var} \int_0^{\tau_1} s e^{-ms} ds & \text{if } m + \rho|\lambda|^2 = 0 \\ rM_{(\tau_1, m)} + 2\text{Var} \frac{[M_{(\tau_1, m)} - M_{(\tau_1, 2m + \rho|\lambda|^2)}]}{m + \rho|\lambda|^2} & \text{if } m + \rho|\lambda|^2 \neq 0, \end{cases}$$

and $f_{\tau_i, \tau_j}(\lambda) = e^{-\frac{1}{2}(2m + \rho|\lambda|^2)(\tau_j - \tau_i)} f_{\tau_i, \tau_i}(\lambda)$ for $\tau_i < \tau_j$.

For more general sets A , we have

$$\frac{N_T(A) - rM_{(\tau,m)}(A)}{(\text{Var } N_T(A))^{1/2}} \xrightarrow{D} N(0,1) \quad \text{as } A \rightarrow R^d,$$

but the limiting variance is not so easily calculated as in Theorem 3.1.

Proof:

By Theorem 2.13, an L -time slice immigration process is B -mixing. For the last part, (1) of the proof of Theorem 3.1 shows $\text{Var}(N_T(A)) = O(|A|)$. Thus the convergence factor in the limit theorem is at most $|A|^{1/2}$, so the normality easily follows. // QED.

Theorem 3.2 (Strong Law of Large Numbers for B -mixing Processes)

Under condition (a) of B -mixing, if $|A| \neq 0$

$$\frac{N(A_T)}{|A_T|} \rightarrow p_1 \quad \text{a.s. as } T \rightarrow \infty, \text{ where } p_1 dx = E\{N(dx)\}$$

Proof: Without loss of generality X is one dimensional.

From the proof to Theorem 3.1,

$$K_{\ell,T} = \text{cum}_{\ell} \left\{ \frac{N(A_T)}{|A_T|} - p_1 \right\} = \begin{cases} O(T^{d(1-\ell)}) & \text{if } \ell \geq 2 \\ 0 & \text{if } \ell = 1 \end{cases}$$

Using the formula in Appendix I,

$$E\left\{\left(\frac{N(A_T)}{|A_T|} - p_1\right)^4\right\} = O(T^{-2d}),$$

that is (since $\frac{N(A_T)}{|A_T|}$ is a real valued process), $\exists K_1 > 0, T_0 > 0$

$$\text{so that } E\left\{\left|\frac{N(A_T)}{|A_T|} - p_1\right|^4\right\} \leq \frac{K_1}{T^{2d}} \quad \text{for all } T \geq T_0. \quad (1)$$

Then

$$P\left\{\left|\frac{N(A_T)}{|A_T|} - p_1\right| \geq \varepsilon\right\} \leq E\left\{\left|\frac{N(A_T)}{|A_T|} - p_1\right|^4\right\} \cdot \varepsilon^{-4}$$

(which is proven in the same way as Chebyshev's Inequality is in Chung (1974)).

Now consider only positive rational $T \geq T_0, T \in \mathbb{N}$. Let

$$S_{n,T} = \left\{\omega \in \Omega: \left|\frac{N(A_T)}{|A_T|} - p_1\right|(\omega) \geq \frac{1}{n}\right\},$$

where the process is defined on (Ω, S, P) :

— For each $d \geq 1, 2d > 1$, so that by (1)

$$\sum_{T \in \mathbb{N}} P\{S_{n,T}\} < \infty.$$

$$T \in \mathbb{N}$$

$$T \geq T_0$$

Then by one half of the Borel-Cantelli Lemma (Theorem 4.2.1, p. 73

Chung (1974))

$P\{S_{n,T} \text{ infinitely often, } T \geq T_0\} = 0$ for each n , T_0 given in (1).

That is, $\exists$ a null set N_n such that if $\omega \in \Omega \setminus N_n$, $\exists T_n(\omega)$ such that

$$\left| \frac{N(A_T)}{|A_T|}(\omega) - p_1 \right| < \frac{1}{n} \quad \forall T \geq T_n(\omega).$$

Let $N = \bigcup_{n=1}^{\infty} N_n$. $\therefore P(N) = 0$.

If $\omega \in \Omega \setminus N$, then for each n , $\exists T_n(\omega)$ such that

$$\left| \frac{N(A_T)}{|A_T|}(\omega) - p_1 \right| < \frac{1}{n} \quad \forall T \geq T_n(\omega)$$

i.e., $\frac{N(A_T)}{|A_T|}(\omega) \rightarrow p_1$ as $T \rightarrow \infty$ through rationals. (2)

Now we extend (2) to $T \rightarrow \infty$ through reals.

We may assume without loss of generality $X \geq 0$ since X is bounded.

For $T \in \mathbb{R}$, let $T_1(T) \leq T \leq T_2(T)$ where $T_1, T_2 \in \mathbb{N}$ and $T_2 - T_1 = 1$. Therefore

$(T_2(T) - T_1(T))/T \rightarrow 0$ as $T \rightarrow \infty$. Then

$$N(A_{T_1}) \leq N(A_T) \leq N(A_{T_2}) \quad (3)$$

$$\text{Also } \lim_{T \rightarrow \infty} \frac{|A_{T_1}|}{|A_T|} = \lim_{T \rightarrow \infty} \frac{|A_{T_2}|}{|A_T|} = 1. \quad (4)$$

Then if $\omega \in \Omega \setminus N$, by (3)

$$\frac{|A_{T_1}|}{|A_T|} \cdot \frac{N(A_{T_1})}{|A_{T_1}|}(\omega) \leq \frac{N(A_T)}{|A_T|}(\omega) \leq \frac{|A_{T_2}|}{|A_T|} \cdot \frac{N(A_{T_2})}{|A_{T_2}|}(\omega), \quad (5)$$

and the two extreme members of (5) converge to p_1 by (4) and by the first part since $T_1, T_2 \in \mathbb{N}$.

Therefore, $\forall \omega \in \Omega \setminus N$, $\frac{N(A_T)}{|A_T|}(\omega) \rightarrow p_1$ as $T \rightarrow \infty$. // QED.

Corollary to Theorem 3.2

For either the initial Poisson field and the immigration simple diffusion process, for each $\tau \geq 0$, if $|A| \neq 0$

$$\frac{N_T^P(A_T)}{|A_T|} \rightarrow \lambda e^{-m\tau} \quad \text{a.s. as } T \rightarrow \infty,$$

$$\frac{N_T^I(A_T)}{|A_T|} \rightarrow m(\tau_1 m) \quad \text{a.s. as } T \rightarrow \infty.$$

Let R_1 be a set satisfying the translation integral property in Theorem 3.1 (ii), and let

$$R_T = T \cdot R_1 = \{Tx : x \in R_1\}.$$

R_T is the set on which the point process is observed. Brillinger (1970) then introduces a "taper" with the following (very important for the approximate maximum likelihood technique of the next chapter) assumption.

Assumption on Tapers

Let $h: \mathbb{R}^d \rightarrow \mathbb{R}$ be a bounded measurable non-negative function

with compact support R_1 , that is, $h(R_1^c) = \{0\}$, and such that

$$\int_{R^d} |h(t+u) - h(t)| dt \leq K|u|$$

for some constant K .

For each $T > 0$, $h(\cdot/T)$ has support R_T and h is called a taper.

As in real time series analysis, we work with the tapered data

$$h(t/T)X(t), \quad t \in R^d.$$

$$\text{Let } H_K(\lambda) = \int_{R^d} h(t)^k e^{-i\langle \lambda, t \rangle} dt.$$

Example: Before proceeding any further, an example of a taper is given.

Take $R_1 = \prod_{j=1}^d [-b_j, b_j]$, $b_j > 0$, and $h = \chi_{R_1}$. Let $b = (b_1, \dots, b_d)$,

$$u = (u_1, \dots, u_d).$$

$$\begin{aligned} \int_{R^d} |h(t+u) - h(t)| dt &= \prod_{j=1}^d \int_{R} |\chi_{[-b_j, b_j]}(x + u_j) - \chi_{[-b_j, b_j]}(x)| dx \\ &= \prod_{j=1}^d a_j \end{aligned}$$

$$\text{where } a_j = \begin{cases} 4b_j & \text{if } |u_j| > 2b_j \\ 2|u_j| & \text{if } |u_j| \leq 2b_j \end{cases}.$$

Therefore, if all $|u_j| \geq 2b_j$,

$$\int_{R^d} |h(t+u) - h(t)| dt = \prod_{j=1}^d (4b_j) = K||u||$$

and if $|u_j| < 2b_j$ for some j ,

$$\int_{R^d} |h(t+u) - h(t)| dt \leq \prod_{k \neq j} (4b_k) 2|u_j| \leq K \|u\|$$

where

$$K = \max \left\{ \prod_1^d (4b_j), \max_{1 \leq i \leq d} 2 \prod_{k \neq i} (4b_k) \right\}.$$

$$H_K(\lambda) = \prod_1^d \int_{-b_j}^{b_j} e^{-ix\lambda_j} dx,$$

$$\text{where } \int_{-b_j}^{b_j} e^{-ix\lambda_j} dx = \begin{cases} 2b_j & \text{if } \lambda_j = 0 \\ \frac{2 \sin(\lambda_j b_j)}{\lambda_j} & \text{if } \lambda_j \neq 0 \end{cases}$$

Brillinger (1970) give some other examples of tapers.

A point process is a generalized random process, in the sense of Schwartz. By Theorem 3, p. 271 of Gel'fand and Vilenkin (1964), X has spectral representation $Z_X(d\lambda)$, that is,

$$X(t) - p_1 \stackrel{\text{a.s.}}{\sim} \int_{R^d} e^{i\langle \lambda, t \rangle} Z_X(d\lambda).$$

We may estimate p_1 by $\hat{p}_1 = \frac{1}{|R_T|} \int_{R_T} X(t) dt$ via Theorems 3.1 and 3.2.

Consider the finite Fourier transform

$$\hat{d}_a^{(T)}(\lambda) = \int h(t/T)[X_a(t) - \hat{p}_a]e^{-i\langle\lambda, t\rangle} dt, \quad a = 1, \dots, L$$

$$= \int g(s, T, \lambda) X_a(s) ds$$

$$\text{where } g(s, T, \lambda) = \begin{cases} h(s/T)e^{-i\langle\lambda, s\rangle} - \frac{T^d}{|R_T|} H_1(T\lambda) & \text{if } s \in R_T \\ 0 & \text{if } s \notin R_T. \end{cases}$$

By the spectral representation

$$\begin{aligned} \hat{d}_a^{(T)}(\lambda) &= \int h(t/T)e^{-i\langle\lambda, t\rangle} [X_a(t) - p_a] dt + (p_a - \hat{p}_a) \int h(t/T)e^{-i\langle\lambda, t\rangle} dt \\ &= T^d \int H_1(T(\lambda - \alpha)) Z_X(d\alpha) - T^d H_1(T\lambda) (\hat{p}_a - p_a). \end{aligned}$$

This indicates $T^{-d/2} \hat{d}_a^{(T)}(\lambda)$ and $T^{d/2}(\hat{p}_a - p_a)$ are asymptotically uncorrelated if $\lambda \neq 0$ (by Riemann-Lebesgue Theorem $H_1(T\lambda) \rightarrow 0$ as $T \rightarrow \infty$).

Theorem 3.3

Let $\lambda_1, \dots, \lambda_k$ be k distinct elements in $\mathbb{R}^d \setminus \{0\}$ such that $\lambda_i \pm \lambda_j \neq 0$ for all $i \neq j$. Then $T^{-d/2} \hat{d}_X^{(T)}(\lambda_1), \dots, T^{-d/2} \hat{d}_X^{(T)}(\lambda_k), T^{-d/2} \hat{d}^{(T)}(0)$ and $T^{d/2}(\hat{p}_1 - p_1)$ asymptotically complex normal and independent, as $T \rightarrow \infty$,

$$\frac{\hat{d}_X^{(T)}(\lambda)}{T^{d/2}} \xrightarrow{D} N_L^c(0, (2\pi)^d f_{XX}(\lambda_j) H_2(0)), \quad j = 1, \dots, k$$

$$\text{and } T^{d/2} \left(\frac{\hat{d}_X^{(T)}(0)}{T^d}, \hat{p}_1 - \hat{p}_1 \right) \xrightarrow{D} N_{2L}(0, (2\pi)^d \begin{bmatrix} f_{XX}(0) & 0 \\ 0 & f_{XX}(0)H_2(0) \end{bmatrix}).$$

Remark : In this work, complex normal means, unless otherwise stated, the special case of uncorrelated (independent) real and imaginary parts, as considered by Goodman (1963). This allows immediate application of his definition and results of the complex Wishart distribution. In the proof below we must consider joint cumulants of the random variables in the theorem and the complements of these random variables.

Proof of Theorem 3.3

$$\text{Let } \kappa_1 = \sup_{\lambda, t, T} |g(t, T, \lambda)| \leq \sup_{t, T} |h(t/T)| + \sup_{T, \lambda} |H_1(T\lambda)|$$

$$\leq \sup_{t, T} |h(t/T)| + \int |h(t)| dt < \infty.$$

$$\text{Notice } E\{T^{-d/2} \hat{d}_X^{(T)}(\lambda)\} = T^{-d/2} \int h(t/T) e^{-i\langle \lambda, t \rangle} [EX(t) - E\hat{p}_1] dt = 0, \forall T > 0.$$

Consider a joint ℓ^{th} factorial cumulant involving ℓ_1 of the $N_a(A_T)$'s and $\ell - \ell_1$ of the $\hat{d}_X^{(T)}(\lambda)$'s, $\lambda = \pm \lambda_j, j = 1, \dots, k, \ell \geq 2$

$$|\text{fac. cum}_{\ell} \{ \int_{R_T} \chi_{a_1}(t) dt, \dots, \int_{R_T} \chi_{a_{\ell_1}}(t) dt, \int_{R_T} g(t, T, \pm \lambda_{j_1}) \chi_{a_{j_1}}(t) dt, \dots, \\ \int_{R_T} g(t, T, \lambda_{j_{\ell - \ell_1}}) \chi_{a_{j_{\ell - \ell_1}}}(t) dt \}|$$

$$\leq \kappa_1^{\ell-\ell_1} \int_{R_T} \left\{ \int \dots \int_{R^{d(\ell-1)}} |q_{a_1, \dots, a_{\ell_1}, a_{j_1}, \dots, a_{j_{\ell-\ell_1}}}(s_1, \dots, s_{\ell-1})| \right. \\ \left. \cdot ds_1 \dots ds_{\ell-1} \right\} \chi_{R_T}(s_\ell) ds_\ell$$

= const. T^d , by B-mixing. Therefore as in Theorem 3.1, an ordinary joint ℓ^{th} cumulant is $O(T^d)$. Thus a normalized joint ℓ^{th} cumulant is

$$T^{-d\ell/2} O(T^d) = O(T^{d(1-\ell/2)}) \rightarrow 0 \text{ as } T \rightarrow \infty \text{ if } \ell \geq 3.$$

Notice the $O(T^{d(1-\ell/2)})$ is uniform in $\lambda_1, \dots, \lambda_k$ for each ℓ .

Now we need only check the second moments.

$$\text{Cov}\{T^{-d/2} \hat{d}_{a_1}^{(T)}(\gamma_1), T^{-d/2} \hat{d}_{a_2}^{(T)}(\gamma_2)\}$$

$$= T^{-d} \int \int_{R^{2d}} g(s_1, T, \gamma_1) g(s_2, T, \gamma_2) c_{a_1 a_2}(s_1 - s_2) ds_1 ds_2$$

where $c_{a_1 a_2}$ is the covariance density for X

$$= T^{-d} \int \int h(s_1/T) h(s_2/T) e^{-i\langle \gamma_1, s_1 \rangle + \langle \gamma_2, s_2 \rangle} c_{a_1 a_2}(s_1 - s_2) ds_1 ds_2$$

$$= T^{-d} \int \int h(s_1/T) e^{-i\langle \gamma_1, s_1 \rangle} \chi_{R_T}(s_2) \frac{H_1(-T\gamma_2)}{|R_T|} c_{a_1 a_2}(s_1 - s_2) ds_1 ds_2$$

$$= T^{-d} \int \int h(s_2/T) e^{i\langle \gamma_2, s_2 \rangle} \chi_{R_T}(s_1) \frac{H_1(T\gamma_1)}{|R_T|} c_{a_1 a_2}(s_1 - s_2) ds_1 ds_2 \quad (1)$$

$$+ T^{-d} \frac{H_1(T\gamma_1)H_1(-T\gamma_2)}{|R_T|^2} \iint X_{R_T}(s_1)X_{R_T}(s_2)c_{a_1a_2}(s_1-s_2)ds_1ds_2.$$

Consider the first term in (1). Note

$$T^{-d} \left| \iint h\left(\frac{s_1}{T}\right)h\left(\frac{s_2}{T}\right)e^{-i\langle\gamma_1, s_1\rangle}e^{-i\langle\gamma_2, s_2\rangle}c_{a_1a_2}(s_1-s_2)ds_1ds_2 \right.$$

$$\left. - \iint h^2\left(\frac{s_1}{T}\right)e^{-i\langle\gamma_1-\gamma_2, s_1\rangle}e^{-i\langle\gamma_2, u\rangle}c_{a_1a_2}(u)ds_1du \right|$$

$$\leq T^{-d} \iint h\left(\frac{s_1}{T}\right)\left|h\left(\frac{s_1+u}{T}\right) - h\left(\frac{s_1}{T}\right)\right| |c_{a_1a_2}(u)|ds_1du$$

$$\leq (\max h(s))K \int \frac{|u|}{T} c_{a_1a_2}(u)du = O(T^{-1}).$$

Therefore the first term in (1) is

$$T^{-d} \iint h^2\left(\frac{s_1}{T}\right)e^{-i\langle\gamma_1-\gamma_2, s_1\rangle}e^{-i\langle\gamma_2, u\rangle}c_{a_1a_2}(u)ds_1du + O(T^{-1})$$

$$= (2\pi)^d f_{a_1a_2}(\gamma_2)H_2(T(\gamma_1-\gamma_2)) + O(T^{-1}).$$

Similarly applying the translation property to the other terms in (1)

gives if $\gamma_1 \neq 0, \gamma_2 \neq 0$

$$\text{cov}\{T^{-d/2} \hat{d}_{a_1}^{(T)}(\gamma_1), T^{-d/2} \hat{d}_{a_2}^{(T)}(\gamma_2)\}.$$

$$= (2\pi)^d f_{a_1a_2}(\gamma_2)H_2(T(\gamma_1-\gamma_2)) + O(T^{-1})$$

$$\left\{ \begin{array}{ll} (2\pi)^d f_{a_1 a_2}(\gamma_2) H_2(0) & \text{if } \gamma_1 = \gamma_2 \\ 0 & \text{otherwise} \end{array} \right. \quad (2)$$

where the $O(T^{-1})$ term is uniform in γ_1, γ_2 . Similarly

$$\begin{aligned} \text{cov}\{T^{-d/2} \hat{d}_{a_1}^{(T)}(\gamma_1), T^{d/2} \hat{p}_{a_2}\} \\ = (2\pi)^d [f_{a_1 a_2}(\gamma_1) - f_{a_1 a_2}(0)] H_2(T\gamma_1) + O(T^{-1}) \end{aligned} \quad (3)$$

//QED.

Remark on Proof of Theorem 3.3

Suppose the B-mixing point process is also Brillinger ℓ -mixing for $\ell = 1$ (definition 3.3). Then if $\lambda_1, \dots, \lambda_k$ are non-zero,

$$\begin{aligned} \text{cum}_k\{T^{-d/2} \hat{d}_{a_1}^{(T)}(\lambda_1), \dots, T^{-d/2} \hat{d}_{a_k}^{(T)}(\lambda_k)\} \\ = (2\pi)^d f_{a_1, \dots, a_k}(\lambda_1, \dots, \lambda_{k-1}) T^{d(1-k/2)} H_k(T(\lambda_1 + \dots + \lambda_k)) + O(T^{d(1-k/2)-1}) \end{aligned}$$

where $f_{a_1, \dots, a_k}$ is an ordinary k^{th} cumulant spectral density and the $O(T^{d(1-k/2)-1})$ term is uniform in $\lambda_1, \dots, \lambda_k$.

Remark: The taper h satisfies the translation property

$$\int_{\mathbb{R}^d} |h(t+u) - h(t)| dt \leq \kappa |u|.$$

In the above Theorem, we cannot obtain asymptotic uncorrelatedness at a rate faster than $O(T^{-1})$ in this class of tapers. This observation

is useful when we consider parametric estimation via an approximate likelihood technique in the next chapter.

One can obtain several estimates near given frequencies by the following.

Theorem 3.4 (Brillinger (1970), Theorem 3.2)

Let $\lambda_j(T) \rightarrow \lambda \neq 0$, $j = 1, \dots, k$ such that $T|\lambda_{j_1}(T) - \lambda_{j_2}(T)| \rightarrow \infty$ as $T \rightarrow \infty$ for $j_1 \neq j_2$. Suppose f_{XX} continuous at λ , then $T^{-d/2} \hat{d}_X^{(T)}(\lambda_1(T)), \dots, T^{-d/2} \hat{d}_X^{(T)}(\lambda_k(T))$ are asymptotically i.i.d. $N_L^C(0, (2\pi)^d f_{XX}(\lambda) H_2(0))$ random variables.

Proof: The normality follows as in Theorem 3.3. For asymptotic independence, we obtain (cf (2) in the proof of Theorem 3.3) for $j_1 \neq j_2$

$$\begin{aligned} & \text{Cov}\{T^{-d/2} \hat{d}_a^{(T)}(\lambda_{j_1}(T)), T^{-d/2} \hat{d}_b^{(T)}(\lambda_{j_2}(T))\} \\ & \sim (2\pi)^d f_{XX}(\lambda) H_2(T(\lambda_{j_1}(T) - \lambda_{j_2}(T))) \rightarrow 0 \end{aligned}$$

by Riemann-Lebesgue Theorem. //QED.

Theorem 3.5 (Brillinger (1970))

If the point process X is B-mixing satisfying the generating function condition, then for each compact set K in R^d ,

$$\lim_{T \rightarrow \infty} \sup_{\lambda \in K} \{ |d_X^{(T)}(\lambda)| / (T^d \log T)^{1/2} \} \leq \{ 2(p^2 + p + 1) H_2(0) \sup_{\lambda \in K} \{ f_{XX}(\lambda) \} \}^{1/2}$$

almost surely, where $d_X^{(T)}(\lambda) = \int h(t/T) e^{-i\langle \lambda, t \rangle} [X(t) - p_1] dt$

Outline of Proof:

We only note the changes in the proof of Theorem 3.3 of Brillinger (1970). In the proof Lemma 9.1 (p. 68 of Brillinger (1970)), we can write $|\text{cum}_\ell \{ \text{Re } d_X^{(T)}(\lambda_1), \dots, \text{Re } d_X^{(T)}(\lambda_\ell) \}| \leq L^k T^d C_k$ where $L = \sup |h(t)|$. Then we obtain

$$\begin{aligned} & |\log E \{ \exp[\alpha \text{Re } d_X^{(T)}(\lambda) (f_{XX}(\lambda))^{-1/2}] \} - \alpha^2 \gamma/2 | \\ & \leq \sum_{k=2}^{\infty} (L^k |\alpha|^k C_k T^d / k!) < \infty \quad \text{for } \alpha \text{ sufficiently small.} \end{aligned}$$

The remainder of the proof now follows in an obvious way the rest of Brillinger's proof. //QED.

Corollary to Theorem 3.5

- (i) Theorem 3.5 holds for $\hat{d}_X^{(T)}$.
- (ii) $\sup_{\lambda \in R^d} \frac{|\hat{d}_X^{(T)}(\lambda)|}{T^d} \rightarrow 0 \quad \text{a.s. as } T \rightarrow \infty.$

Proof: (i) Obvious by Theorems 3.5 and 3.2.

(ii) For each $m \in \mathbb{N}$ $\exists$ a finite set $\Gamma_m = \{\gamma_{1,m}, \dots, \gamma_{j_m,m}\}$ $|\gamma| \leq (2\pi)^d$, such that for all $\lambda \in R^d$ $\exists \gamma \in \Gamma_m$ $\rightarrow$ $|e^{-i\langle \lambda, t \rangle} - e^{-i\langle \gamma, t \rangle}| < \frac{1}{m}$ (this follows since $e^{-i\langle \lambda, t \rangle}$ is periodic) for all $t \in K$, compact. Then since X is bounded, say by M , there exists

$\gamma \in \Gamma_m$ for each $\omega \in \Omega$

$$\begin{aligned}
 & |T^{-d} \int h(t/T) e^{-i\langle \lambda, t \rangle} [X(t) - \hat{p}_1] dt \\
 & - \int h(t) e^{-i\langle \gamma, t \rangle} [X(Tt) - \hat{p}_1] dt| \\
 & = |\int h(t) [e^{-i\langle T\lambda, t \rangle} - e^{-i\langle \gamma, t \rangle}] [X(Tt) - \hat{p}_1] dt| \\
 & \leq H_1(0)(M + |p_1(\omega)|) \cdot 1/m
 \end{aligned}$$

By the same technique as in the proof of Theorem 3.2, $\frac{\hat{d}^{(T)}(\lambda)}{T^d} \rightarrow 0$ almost surely for each λ , so that

$$\sup_{\lambda \in \Gamma_m} \frac{|\hat{d}^{(T)}(\lambda)|}{T^d} \rightarrow 0 \quad \text{a.s.} \quad \text{as } T \rightarrow \infty \quad **$$

Let N_m be the exceptional null set in **.

Let N_0 be a null set from Theorem 3.2 $\hat{p}_1(\omega) \neq p_1$ on N_0^c . Then

$N = \bigcup_{m=0}^{\infty} N_m$ is a null set. Then if $\omega \in \Omega \setminus N$ so $\hat{p}_1(\omega) \leq \kappa(\omega) < \infty$, then for all $m \in \mathbb{N}$, and if $T \geq 1$

$$\begin{aligned}
 \sup_{\lambda \in \mathbb{R}^d} \frac{|\hat{d}^{(T)}(\lambda)|}{T^d}(\omega) & \leq \frac{H_1(0)[M + \kappa(\omega)]}{m} + \sup_{\gamma \in \Gamma_m} \frac{|\hat{d}^{(T)}(\gamma/T)|}{T^d}(\omega) \\
 & \leq \frac{H_1(0)[M + \kappa(\omega)]}{m} + \sup_{|\gamma| \leq (2\pi)^d} \frac{|\hat{d}^{(T)}(\gamma)|}{T^d}(\omega)
 \end{aligned}$$

The set $K = \{\gamma: |\gamma| \leq (2\pi)^d\}$ is compact, so that by Theorem 3.4

there exists a null set N_K such that $\sup_{\gamma \in K} \frac{|\hat{d}^{(T)}(\gamma)|}{T^d} \rightarrow 0$ on N_K^c .

$N^* = N_K \cup N$ is a null set, so for $\omega \in \Omega \setminus N^*$,

$$\sup_{\lambda \in R^d} \frac{|\hat{d}^{(T)}(\lambda)|}{T^d}(\omega) \leq \frac{H_1(0)[M + K(\omega)]}{m} + \epsilon \quad \text{for } T \text{ sufficiently large.}$$

Since m and ϵ are arbitrary, Corollary (ii) follows. //QED.

The periodogram is given by

$$\hat{I}_{XX}^{(T)}(\lambda) = (2\pi)^{-d} T^{-d} H_2(0)^{-1} \hat{d}^{(T)}(\lambda) * \hat{d}^{(T)}(\lambda).$$

Corollary to Theorems 3.3 and 3.4

For $\lambda_1, \dots, \lambda_k$, distinct non-zero elements of R^d such that $\lambda_{j_1} \pm \lambda_{j_2} \neq 0$,

and for positive integers $\ell_1, \dots, \ell_k$, $\lambda_{n,j}(T) \rightarrow \lambda_j$, as $T \rightarrow \infty$,

$n = 1, \dots, \ell_j$, $j = 1, \dots, k$ such that $T|\lambda_{n_1,j}(T) - \lambda_{n_2,j}(T)| \rightarrow \infty$

as $T \rightarrow \infty$, then $\sum_{n=1}^{\ell_1} \hat{I}^{(T)}(\lambda_{n,1}(T)), \dots, \sum_{n=1}^{\ell_k} \hat{I}^{(T)}(\lambda_{n,k}(T))$ are

asymptotically independent complex Wishart (see for example Goodman (1963)).

$$\sum_{n=1}^{\ell_j} \hat{I}^{(T)}(\lambda_{n,j}(T)) \xrightarrow{D} W_L^c(\ell_j, f_{XX}(\lambda_j))$$

Proof: This follows directly from Goodman's (1963) paper. //QED.

Remarks: (1) In order to have asymptotically non-singular distributions,

it is necessary that $\ell_j \geq L$.

(2) In case $L = 1$, if $\lambda_1, \dots, \lambda_k$ are distinct, $\lambda_j - \lambda_i \neq 0$, then $\hat{I}^{(T)}(\lambda_1), \dots, \hat{I}^{(T)}(\lambda_k)$ are asymptotically independent and if $f(\lambda) > 0$,

$$\frac{\hat{I}^{(T)}(\lambda_j)}{\sqrt{f(\lambda_j)}} \xrightarrow{D} \chi_{(2)}^2, \quad \text{that is, exp. 1, if } \lambda_j \neq 0$$

$$\xrightarrow{D} \chi_{(1)}^2 \quad \text{if } \lambda_j = 0$$

Corollary to Theorem 3.3

(i) $E\{\hat{I}^{(T)}(\lambda)\} = f_{XX}(\lambda) + O(T^{-1})$ uniformly in λ

(ii) $E\{\hat{I}^{(T)}(\lambda)\} = \frac{\int |H_1(\alpha)|^2 f_{XX}(\lambda - \alpha/T) d\alpha}{\int |H_1(T\alpha)|^2 d\alpha} + \frac{\int |G_1(\alpha)|^2 f_{XX}(\alpha/T) d\alpha}{\int |H_1(T\alpha)|^2 d\alpha}$

$$\cdot \frac{|H_1(T\lambda)|^2}{(2\pi) d_{H_2}(0)} - 2 \operatorname{Re} \left\{ \frac{H_1(T)}{(2\pi) d_{H_2}(0)} \int H_1(\alpha) G_1(T\lambda - \alpha) \cdot f_{XX}(\lambda - \alpha/T) d\alpha \right\},$$

where $G_1(T\alpha) = \int e^{i\langle \alpha, s \rangle} \chi_{R_T}(s) ds$.

Proof:

(i) $E\{\hat{I}^{(T)}(\lambda)\} = (2\pi)^{-d_{H_2}(0)-1} \operatorname{Cov}\{\hat{d}_X^{(T)}(\lambda), \hat{d}_X^{(T)}(\lambda)\}$

$= f_{XX}(\lambda) + O(T^{-1})$ by (2) of the proof of Theorem 3.3.

(ii) This follows from the Cramér representation.

//QED.

part (ii) follows.

//QED.

As in real time series analysis, the periodogram is not a consistent estimate of the spectral density, and so we are led to consider weighted combinations of the periodogram.

Assumption on Weights

Let $W: \mathbb{R}^d \rightarrow \mathbb{R}$ be measurable, real valued, even, and vanishing at infinity (that is, $\forall \epsilon > 0, \exists$ compact $K \ni |W(x)| < \epsilon \forall x \in K^c$), such that

$$(i) \int_{\mathbb{R}^d} W(\alpha) d\alpha = 1,$$

(ii) if $\omega(u) = \int W(\alpha) e^{-i\langle u, \alpha \rangle} d\alpha$, then ω is absolutely integrable and $\int |u| |\omega(u)| du < \infty$.

Example: For each $\dim \mathbb{R}^d$, an example of such a W is given by

$$W(x) = \begin{cases} (1 - |x|)^\beta, & \text{if } |x| < 1 \\ 0 & \text{otherwise} \end{cases}$$

and where $\beta + \frac{d}{2} \geq \frac{3}{2}$ and $\beta > 0$.

Let $\{B_T: T > 0\}$ be a "sequence" of positive real numbers and $W^{(T)}(\alpha) = B_T^{-d} W(B_T^{-1}\alpha)$.

Also define

$$\hat{f}_{XX}^{(T)}(\lambda) = \int W^{(T)}(\lambda - \alpha) \hat{I}_{XX}^{(T)}(\alpha) d\alpha.$$

Notice, if W is non-negative, then $\hat{f}_{XX}^{(T)}(\cdot)$ is non-negative definite and also Hermitian.

$$\begin{aligned} \hat{f}_{XX}^{(T)}(\lambda) &= \int_{R^d} W^{(T)}(\lambda - \alpha) \hat{I}_{XX}^{(T)}(\alpha) d\alpha \\ &= \frac{(2\pi)^{-d}}{R^{2d}} \int \int \omega(B_T u) e^{i\langle u, \alpha - \lambda \rangle} \hat{I}_{XX}^{(T)}(\alpha) du d\alpha, \text{ (since } W \text{ is even)} \\ &= (2\pi)^{-2d} H_2(0)^{-1} T^{-d} \int \int \int \int \omega(B_T u) e^{i\langle u, \alpha - \lambda \rangle} h(t/T) h\left(\frac{t+s}{T}\right) e^{i\langle \alpha, s \rangle} \\ &\quad \int_{R^{4d}} \end{aligned}$$

$$\text{where } Y(t) = \int Y(t) * Y(t+s) dt ds du d\alpha,$$

where $Y(t) = X(t) - \hat{p}_1$.

If $\int (\cdot) d\alpha$ is $\oint (\cdot) d\alpha$, the Cauchy principle value

(that is, $\int_{R^d} f(\alpha) d\alpha = \lim_{n \rightarrow \infty} \int_{[-n, n]^d} f(x) dx$)

then

$$\hat{f}_{XX}^{(T)}(\lambda) = (2\pi)^{-d} H_2(0)^{-1} T^{-d} \iiint_{R^{3d}} \omega(B_T u) e^{-i\langle u, \lambda \rangle} h(t/T) h\left(\frac{t+s}{T}\right) \delta(u-s) Y(t) * \\ \cdot Y(t+s) dt ds du$$

$$= (2\pi)^{-d} H_2(0)^{-1} T^{-d} \iint \omega(B_T u) e^{-i\langle u, \lambda \rangle} h(t/T) h\left(\frac{t+s}{T}\right) Y^*(t) Y(t+u) dt du,$$

where $\delta(u)$ is the Dirac delta function.

Theorem 3.6 (Brillinger (1970), Theorem 4.4)

Suppose $W \geq 0$ satisfies the Weight assumption, $f_{ab}(\lambda)$ is bounded in λ for all $1 \leq a, b \leq L$, is continuous at λ and the fourth cumulants $c_{a_1 a_2 a_3 a_4}(u_1, u_2, u_3)$ of the B-mixing point process X satisfies $\iiint_{R^{3d}} |u_i| |c_{a_1 a_2 a_3 a_4}(u_1, u_2, u_3)| du_1 du_2 du_3 < \infty$, $i = 1, 2, 3$.

Suppose also $B_T \rightarrow 0$ and $T B_T \rightarrow \infty$ as $T \rightarrow \infty$. Then

$$(i) \ E\{\hat{f}_{XX}^{(T)}(\lambda)\} = \int W^{(T)}(\lambda - \alpha) f_{XX}(\alpha) d\alpha + O(T^{-1}) \\ + f_{XX}(\lambda) \quad \text{as } T \rightarrow \infty,$$

and where $O(T^{-1})$ is uniform in λ .

$$(ii) \ \text{Cov}\{\hat{f}_{a_1 b_1}^{(T)}(\lambda), \hat{f}_{a_2 b_2}^{(T)}(\mu)\}$$

$$\begin{aligned}
&= (2\pi)^d H_4(0) H_2(0)^{-2} \{ f_W^{(T)}(\lambda - \alpha) W^{(T)}(\mu - \alpha) f_{a_1 a_2}(-\alpha) f_{b_1 b_2}(\alpha) d\alpha \\
&\quad + f_W^{(T)}(\lambda - \alpha) W^{(T)}(\mu + \alpha) f_{a_1 b_2}(-\alpha) f_{a_2 b_1}(\alpha) d\alpha \} \\
&\quad + O((TB_T)^{-d-1}) + o((B_T T)^{-d})
\end{aligned}$$

where the error terms are uniform in λ, μ . If in addition, $B_T^{d+1} = O(T^{-1})$, the error term is uniformly $O((TB_T)^{-d-1})$. Also

$$\begin{aligned}
&\lim_{T \rightarrow \infty} (B_T T)^d \text{Cov}\{\hat{f}_{a_1 b_1}^{(T)}(\lambda), \hat{f}_{a_2 b_2}^{(T)}(\mu)\} \\
&= (2\pi)^d H_4(0) H_2(0)^{-2} \{ \delta(\lambda - \mu) f_{a_1 b_2}(-\lambda) f_{a_2 b_1}(\lambda) + \delta(\lambda + \mu) \\
&\quad \cdot f_{a_1 a_2}(-\lambda) f_{b_1 b_2}(\lambda) \}
\end{aligned}$$

(iii) [If $B_T = O(T^{-(1-\frac{1}{2})})$ and W is bounded] $(B_T T)^{d/2} \hat{f}_{XX}^{(T)}(\lambda_j)$, $j = 1, \dots, k$, with $\lambda_1, \dots, \lambda_k$ given, are asymptotically normal with the asymptotic covariance structure in (ii).

Proof:

(i) By Corollary to Theorem 3.3 (i),

$$\begin{aligned}
E\{\hat{f}_{ab}^{(T)}(\lambda)\} &= \int W^{(T)}(\lambda - \alpha) f_{ab}(\alpha) d\alpha + O(T^{-1}) \\
&= \int W^{(T)}(\alpha) f_{ab}(\lambda - \alpha) d\alpha + O(T^{-1})
\end{aligned}$$

Let $\epsilon > 0$. By continuity of f at λ , choose $a_0 > 0$

$\forall \alpha \in [-a_0, a_0]^d, |f_{ab}(\lambda - \alpha) - f(\lambda)| < \frac{\epsilon}{2}$.

Choose n_0 s.t. $\forall n > n_0, \int_{([-n, n]^d)^c} |W(\alpha)| d\alpha < \frac{\epsilon}{4B}$ where $B = \sup_{\gamma, a, b} |f_{a,b}(\gamma)|$

Choose $T_0(\epsilon)$ s.t. $\forall T > T_0, B_T^{-1} a_0 > n_0$, since $B_T \rightarrow 0$. Then

$$|\int W^{(T)}(\gamma) f_{ab}(\lambda - \alpha) d\alpha - f_{ab}(\lambda)|$$

$$\leq |B_T^{-d} \int W(\alpha) [f_{ab}(\lambda - \alpha) - f(\alpha)] d\alpha|$$

$$\leq B_T^{-d} \int_{([-a_0, a_0]^d)^c} W(B_T^{-1} \alpha) 2B d\alpha + B_T^{-d} \int_{[-a_0, a_0]^d} |W(B_T^{-1} \alpha)| \frac{\epsilon}{2} d\alpha$$

$$\leq \int_{([- \frac{a_0}{B_T}, \frac{a_0}{B_T}]^d)^c} W(\alpha) d\alpha 2B + \frac{\epsilon}{2} \int_{R^d} W(\alpha) d\alpha$$

$$\leq \frac{\epsilon}{4B} 2B + \frac{\epsilon}{2} = \epsilon.$$

(ii) and (iii).

We first show that the joint ℓ^{th} cumulants, $\ell \geq 3$, of the normalized $\hat{f}^{(T)}$'s tend to zero as $T \rightarrow \infty$. Consider a joint ℓ^{th} factorial moment of $\hat{f}^{(T)}$. Up to a constant factor it is bounded by

$$T^{-d\ell} \int_R \int_{\prod_{k=1}^{\ell} [-\frac{a_0}{B_T}, \frac{a_0}{B_T}]^d} |W(B_T u_k) h(t_k/T) h(\frac{t_k + u_k}{T})| \text{fac cum}_{\ell} \{Y_{a_1}(t_1) Y_{b_1}(t_1 + u_1), \dots, Y_{a_{\ell}}(t_{\ell}) Y_{b_{\ell}}(t_{\ell} + u_{\ell})\} | dt_1 du_1 \dots dt_{\ell} du_{\ell}. \quad (1)$$

We need to consider factorial cumulants of products of random variables. To do this we apply the theorem of Leonov and Shiryaev (1959) (see also Appendix II). Consider an indecomposable partition of the table

$$\begin{array}{cc} a_1 & b_1 \\ a_2 & b_2 \\ \vdots & \vdots \\ a_\ell & b_\ell \end{array}$$

Since $E\{Y_a(t)\} = 0$, we need only consider those indecomposable partitions which contain no sets of size 1. By Appendix I,

$$\text{fac-cum}_k\{Y_{a_1}(s_1), \dots, Y_{a_k}(s_k)\} = \sum_{j_1=0}^1 \sum_{j_k=0}^1 \text{fac-cum}_k\{Z_{1,j_1}, \dots, Z_{k,j_k}\} \quad (2)$$

$$\text{where } Z_{i,j_i} = \begin{cases} X_{a_i}(s_i) - p_{a_i} & \text{if } j_i = 0 \\ p_{a_i} - \hat{p}_{a_i} & \text{if } j_i = 1 \end{cases}$$

Considering the first terms only, that is, $Z_{ij} = X_{a_i}(\cdot) - p_{a_i}$, we have for an indecomposable partition

$$\begin{aligned} \int \dots \int_R \prod_{k=1}^{\ell} |w(B_T u_k) h(t_k/T) h(\frac{t_k + u_k}{T})| q_{v_1}(s_1, \dots, s_{N_1-1}) |q_{v_2}(s_{N_1+1}, \dots, s_{N_1+N_2-1})| \dots \\ |q_{v_i}(s_{N_1+\dots+N_{j-1}+1}, \dots, s_{N_1+\dots+N_j-1})| \\ dt_1 \dots dt_\ell du_1 \dots du_\ell \end{aligned} \quad (3)$$

where s_i is of the form $t_{j_1} - t_{j_2} + \varepsilon_i u_{j_1} - \gamma_i u_{j_2}$, $\varepsilon_i, \gamma_i = 0$ or 1 and the s_i 's are all distinct, and $v_1, \dots, v_j$ are collections of $a_1, \dots, a_\ell$ and $b_1, \dots, b_\ell$ (this is Brillinger's notation from the quoted Theorem). Since the partition is indecomposable we can transform in a one to one manner

$$(t_1, \dots, t_\ell, u_1, \dots, u_\ell) \rightarrow (t_1, s_1, \dots, s_{N_1-1}, s_{N_1+1}, \dots, s_{N_1+N_2-1}, \dots, s_{N_1+\dots+N_{j-1}+1}, \dots, s_{N_1+\dots+N_j-1}, u_{i_1}, \dots, u_{i_{j-1}})$$

Also bound $|w(B_T u_n)| \leq \int |W(\alpha)| d\alpha < \infty$ if $n \in \{1, \dots, \ell\} \setminus \{i_1, \dots, i_{j-1}\}$ and

$$b\left(\frac{t_i + u_{i_n}}{T}\right) \leq K, \text{ if } n \in \{i_1, \dots, i_{\ell-j}\}.$$

Then letting $J = \{i_1, \dots, i_j\}$ bound (3) by

$$\begin{aligned} & \text{const.} \int \dots \int_{\mathbb{R}} \prod_J |w(B_T u_j)| |h(t_1/T)| q_{v_1}(s_1, \dots, s_{N_1-1}) | \dots \\ & |q_{v_i}(s_{N_1+\dots+N_{j-1}+1}, \dots, s_{N_1+\dots+N_j-1})| dt_1 ds_N \dots ds_{N_1-1} \\ & \dots ds_{N_1+\dots+N_{j-1}+1} \dots ds_{N_1+\dots+N_j-1} du_{i_1} \dots du_{i_{j-1}} \\ & = \text{const.} T^{d_{B_T} - d(j-1)} \leq \text{const.} (TB_T^{-\ell+1})^d \text{ for } T \geq T_0 \end{aligned}$$

since $B_T < 1 \quad \forall T \geq T_0$, (since $1 \leq j \leq \ell$)

Other terms resulting from the expansion (2) are of smaller order.

Thus from (3) we obtain a joint ℓ^{th} factorial cumulant of $T_{f_{a_i b_i}}^{d\hat{f}(T)}$'s

is $O(T_{B_T}^{d_{B_T}-d(\ell-1)})$, and so a joint ℓ^{th} cumulant of $T_{f_{a_i b_i}}^{d\hat{f}(T)}$'s is

$$O(T_{B_T}^{d_{B_T}-d(\ell-1)} + \dots + O(T_{B_T}^{d_{B_T}-d(\ell-1)}) = O(T_{B_T}^{d_{B_T}-(\ell-1)d}) \text{ since } B_T \rightarrow 0.$$

Therefore a joint ℓ^{th} cumulant of $f_{a_i b_i}^{(T)}$'s is $O(T_{B_T}^{d(1-\ell)-d(1-\ell)})$.

Now considering the normalized $\hat{f}^{(T)}$'s, it follows that a joint ℓ^{th}

cumulant of $(B_T T)^{d/2} f_{a_i b_i}^{(T)}$'s is $(B_T T)^{d\ell/2} O(T_{B_T}^{d(1-\ell)-d(\ell-1)d}) =$

$O(T_{B_T})^{d(1-\ell/2)} \rightarrow 0$ as $T \rightarrow \infty$ if $\ell \geq 3$, since $T_{B_T} \rightarrow \infty$ as $T \rightarrow \infty$.

To calculate the standardized covariance limit is fairly straightforward and so is omitted. //QED.

For the special case of isotropic processes (that is, $f_{XX}(\lambda)$ is constant on the set $\{\lambda: |\lambda| = r\}$ for each $r \geq 0$) a better estimate is available (of course for $\dim R^d \geq 2$). Let $\Gamma(\beta)$ be Lebesgue measure on the unit sphere S_d . Consider the estimate, for $\lambda \neq 0$

$$\hat{g}_{XX}^{(T)}(\lambda) = \int_{S_d} \hat{f}_{XX}^{(T)}(|\lambda|\beta) d\Gamma(\beta)$$

OF/DE

$$= \int_{R^d} \hat{I}_{XX}^{(T)}(\alpha) \left\{ \int_{S_d} W^{(T)}(|\lambda|\beta - \alpha) d\Gamma(\beta) \right\} d\alpha$$

Theorem 3.7 (Brillinger (1970), Theorem 4.6)

Under the same conditions as Theorem 3.6, if $\mu, \lambda \neq 0$, and the process is isotropic

$$(i) E\{\hat{g}_{XX}^{(T)}(\lambda)\} = \iint W^{(T)}(\lambda - \gamma) f_{XX}(\gamma) \left\{ \int W^{(T)}(|\lambda|\beta - \alpha) d\beta \right\} d\alpha d\gamma + O(T^{-1})$$

$$\rightarrow f_{XX}(\lambda) \text{ as } T \rightarrow \infty$$

$$(ii) \lim_{T \rightarrow \infty} B_T^{d-1} T^d \text{Cov}\{\hat{g}_{a_1 b_1}^{(T)}(\lambda), \hat{g}_{a_2 b_2}^{(T)}(\mu)\}$$

$$= (2\pi)^d H_2(0) H_2(0)^{-2} \delta(|\lambda| - |\mu|) [f_{a_1 a_2}(\lambda) f_{b_1 b_2}(\lambda)$$

$$+ f_{a_1 b_2}(\lambda) f_{a_2 b_1}(\lambda)] \cdot 2^{-1} |\lambda|^{-1} \int |u|^{-1} \omega(u)^2 du$$

(iii) $\hat{g}_{XX}^{(T)}(\lambda_1), \dots, \hat{g}_{XX}^{(T)}(\lambda_k)$ are asymptotically normal with covariance given in (ii) and with rate $(B_T^{d-1} T^d)^{1/2}$.

Remark: (1) The proof is essentially the same as that of Theorem 3.6 and is omitted.

(2) The rate of convergence for the isotropic estimate is $B_T^{-1/2}$ times the rate for the estimate in Theorem 3.6, and $B_T^{-1/2} \rightarrow \infty$ as $T \rightarrow \infty$.

§4 Parametric Inference for a Simple Branching Diffusion Process

§4.1 Introduction and Summary of §4

The problem of estimating the parameters of a simple branching diffusion immigration process is considered. The observed data is an L -time slice version of the process on a set $A \subseteq \mathbb{R}^d$. This contrasts with methods for similar processes in which data is observed over many "times" and the asymptotics are as the "number of times" tends to infinity. The type of data considered here may be more realistic in some applications.

The observed data is assumed to be the locations of all the individuals in the set A . In this sense the data is continuous and individuals are indistinguishable. In particular, individuals are not followed through time.

To be specific, the L -time slice version considered here are for $L = 1$ and $L = 2$, say at times τ and $\tau + t$, $t > 0$ being an experimental parameter. The parameter for the time slice versions are

- ρ - diffusion rate
- r - immigration rate
- V - branching rate
- m - death rate minus birth rate
- τ - time of the (first) time slice.

The parameters $\theta = (m, V, \rho, r, \tau)$ are scale dependent. When $L = 1$, only a subset of the parameters may be estimated relative to some values of the others. In particular, it is assumed either τ is known, or that $m \neq 0$ is known. In case $L = 2$, $t > 0$ is used to determine the time

scale, so that even τ can be estimated.

It is also assumed that $\rho > 0$, $r > 0$, $V > 0$, $\tau > 0$ and $0 < \alpha < 1$ where $\alpha = (V-m)/(2V)$. These ensure there is immigration, branching and movement. If $\alpha = 0$, the process is a pure death process and the second order spectral density function becomes

$$f(\lambda|\theta) = (2\pi)^{-d} m(\tau, m) = \text{constant, independent of } \lambda$$

The method of estimation used exploits the variation in $f(\lambda|\theta)$.

The limit theorems obtained in this chapter are for some special large sets, namely $A_T = TA$, where $T > 0$ and $T \rightarrow \infty$. The method used to obtain consistent estimates is outlined is Brillinger's Lemma given at the end of this section.

In chapter 3, it is shown that periodograms are asymptotically independent. Using this, one can write down an asymptotic likelihood function by using the limiting distributions instead of the correct ones. This gives rise to a quadratic function in the random Fourier transforms (Theorem 3.4). By Brillinger's Lemma, a consistent estimate $\hat{\theta}_T$ of θ' , the true value of the parameter is obtained, where the process is observed on A_T . If A_1 is a rectangle, $\hat{\theta}_T \xrightarrow{\text{a.s.}} \theta'$. Also when A_1 is a rectangle, a limiting normal distribution of $\hat{\theta}_T$ is obtained. Section 4.2 considers in detail the case $L = 1$ and §4.3 considers $L \geq 2$. Some asymptotic Fisher information results are obtained. Another optimality criterion is also defined.

Taking the point of view of estimating equations, instead of using an asymptotic likelihood procedure, one could use an analogue of least squares. When $L \geq 2$, this has some advantages over asymptotic

likelihood. This is considered in §4.4.

Theorem 4.1.1 (Brillinger's Lemma, Brillinger (1974), page 94)

Suppose

- (i) θ is an open subset of R^k ,
- (ii) $(\Omega, \mathcal{A}, P)$ is a probability space, Ω complete, separable, metric,
- (iii) $Q^{(T)}(\theta, \omega)$ is real valued, Borel measurable for each $(\theta, \omega) \in \theta \times \Omega$ and $T > 0$,
- (iv) $Q(\theta)$ is real valued, Borel measurable, lower semi-continuous and $Q(\theta) > Q(\theta')$ for all $\theta \neq \theta'$ (that is Q has a unique minimum at θ'),
- (v) $Q^{(T)}(\theta', \omega) = Q(\theta') + o_p(1)$ as $T \rightarrow \infty$
 $Q^{(T)}(\theta, \omega) \geq Q(\theta) + o_p(1)$
- (vi) Given $\varepsilon > 0$, $\eta > 0$, $\theta_1 \neq \theta'$, there exists a neighbourhood U_1 of θ_1 and T_0 such that

$$P(\omega: Q^{(T)}(\theta', \omega) - \inf_{U_1} Q^{(T)}(\theta, \omega) \geq -\varepsilon) < \eta \text{ for all } T \geq T_0,$$
- (vii) There exists a compact set $K \ni \theta$, $\theta' \in K^0$, $\hat{\theta}_T \in K$, such that

$$Q^{(T)}(\hat{\theta}_T, \omega) \leq Q^{(T)}(\theta, \omega) \text{ for all } \theta \in K$$

Then there exists $\hat{\theta}_T$, a local minimum of $Q^{(T)}(\theta, \omega)$ such that

$$\hat{\theta}_T \xrightarrow{\text{Pr}} \theta' \text{ as } T \rightarrow \infty.$$

§4.2 Approximate Likelihood Inference for a 1-Time Slice Immigration Process

In this section we consider a generalization of Brillinger's (1974) remark on "approximate" likelihood. It is approximate in the sense that a family of asymptotic distributions is used instead of the "correct" family.

For the immigration process, the second order spectral density is

$$f_{\tau,\tau}(\lambda|\theta) = (2\pi)^{-d} \{rM_{(\tau,m)} + 2\text{Var}F(\lambda|\theta)\},$$

where

$$F(\lambda|\theta) = \begin{cases} \int_0^\tau se^{-ms}ds & \text{if } m + \rho|\lambda|^2 = 0 \\ \frac{(M_{(\tau,m)} - M_{(\tau,2m+\rho|\lambda|^2)})}{m+\rho|\lambda|^2} & \text{if } m + \rho|\lambda|^2 \neq 0 \end{cases}$$

and where $\theta = (m, V, \rho, r, \tau)$ are the parameters of the process.

It is seen, from chapter 2, that $f_{\tau,\tau}(\lambda|\theta) > 0$ for all λ, θ .

By Theorem 3.3, if $\pm\lambda_1, \dots, \pm\lambda_k$ are distinct, non-zero,

$$\frac{T^{-d/2} \hat{d}^{(T)}(\lambda_j)}{(2\pi)^{d/2} H_2(0)^{1/2} (f(\lambda_j|\theta))^{1/2}}, \quad j = 1, \dots, k$$

are asymptotically i.i.d. $N_1^c(0,1)$ as $T \rightarrow \infty$, when $\theta = \theta'$. Let

$$g(\lambda_j|\theta) = f(\lambda_j|\theta) / ((2\pi)^d r_{M(\tau,m)})$$

$$= 1 + 2V_\alpha G(\lambda|\theta)$$

where $G(\lambda|\theta) = \frac{F(\lambda|\theta)}{M(\tau,m)}$

Then by Corollary p. 93, Chung (1974),

$$\frac{T^{-d/2} \hat{d}^{(T)}(\lambda_j)}{(2\pi)^d \hat{p}_1 H_2(0)^{1/2} g(\lambda_j|\theta)}, \quad j = 1, \dots, k$$

are asymptotically i.i.d. $N_1^C(0, I)$ r.v.'s as $T \rightarrow \infty$ at $\theta = \theta'$, and

where $\hat{p}_1(T) = \frac{N_T^I(R_T)}{|R_T|}$ is an estimate of $p_1 = r_{M(\tau,m)}$ obtained in

Theorem 3.1. Thus, when $\theta = \theta'$,

$$\frac{\hat{I}^{(T)}(\lambda_j)}{(2\pi)^d \hat{p}_1}, \quad j = 1, \dots, k$$

are asymptotically independent $g(\lambda_j|\theta) \chi_{(2)}^2$ r.v.'s, or equivalently exponentially distributed r.v.'s with means $g(\lambda_j|\theta)$.

The "approximate" or asymptotic log likelihood function for given data $\frac{\hat{I}^{(T)}(\lambda_1)}{(2\pi)^d \hat{p}_1}, \dots, \frac{\hat{I}^{(T)}(\lambda_k)}{(2\pi)^d \hat{p}_1}$ is defined to be the log likelihood

obtained using the asymptotic distribution in place of the exact one;

that is

$$L(\theta) = - \sum_{j=1}^k \left\{ \log g(\lambda_j | \theta) + \frac{\hat{I}^{(T)}(\lambda_j)}{(2\pi)^d \hat{p}_1} \cdot \frac{1}{g(\lambda_j | \theta)} \right\}. \quad (4.2.1)$$

The r.v.'s are asymptotically independent, but not identically distributed. In this type of situation, maximum likelihood estimates are not always consistent (see §5, concluding remarks, for a simple example).

Aside: Instead of working with the normalized periodograms $\frac{\hat{I}^{(T)}(\lambda)}{(2\pi)^d \hat{p}_1}$

one could work with just the $\hat{I}^{(T)}(\lambda)$. This is briefly discussed at the end of this section.

At this point, the frequencies $\lambda_j(T)$ and $k = k(T)$ are parameters of the approximate likelihood function. Recall the process is observed on TR_1 . The next two lemmas about the function g motivate a choice for these frequencies and for $k(T)$.

Lemma 4.2.1

For the immigration process, for each θ' , $\theta \in \Theta$, the parameter space,

$$\int_{R^d} \left[\log g(\lambda | \theta) + \frac{g(\lambda | \theta')}{g(\lambda | \theta)} - 1 \right] d\lambda$$

exists as a Riemann integral as is finite.

Proof:

Since the process is isotropic, the Lemma holds if and only if

$$\int_0^\infty \left[\log g(\lambda_x | \theta) + \frac{g(\lambda_x | \theta')}{g(\lambda_x | \theta)} - 1 \right] dx$$

exists and is finite, where $\lambda_x = (x, 0, \dots, 0) \in \mathbb{R}^d$.

$$g(\lambda_x | \theta) = 1 + 2V\alpha G(\lambda_x | \theta)$$

where $2V\alpha G(\lambda_x | \theta)$ is positive, bounded, continuous in (x, θ) and for each θ is $O((1 + x^2)^{-1})$. The Lemma now easily follows. //QED.

$$\text{Let } \Lambda(\theta; \beta) = - \int_0^\beta \left[\log g(\lambda_x | \theta) + \frac{g(\lambda_x | \theta')}{g(\lambda_x | \theta)} - 1 \right] dx$$

Lemma 4.2.2

For the simple branching diffusion immigration process, and for each $0 < \beta \leq \infty$, $\Lambda(\theta; \beta)$ has a unique maximum at θ' , the true parameter value, in either Θ_1 or Θ_2 . In Θ_2 , if $m = 0$, this result holds if either V or α is also fixed.

Proof:

$$\begin{aligned} \Lambda(\theta; \beta) - \Lambda(\theta'; \beta) &= - \int_0^\beta \left[\log g(\lambda_x | \theta) + \frac{g(\lambda_x | \theta')}{g(\lambda_x | \theta)} - 1 - \log g(\lambda_x | \theta') \right] dx \\ &= \int_0^\beta \left[\log \frac{g(\lambda_x | \theta')}{g(\lambda_x | \theta)} - \frac{g(\lambda_x | \theta')}{g(\lambda_x | \theta)} + 1 \right] dx \end{aligned}$$

Notice, if $y \in \mathbb{R}^+$, $\log y = y + 1 = 0$ if $y = 1$
 < 0 if $y \neq 1$

Since $\frac{g(\lambda_x|\theta')}{g(\lambda_x|\theta)}$ is positive and continuous in x , $\Lambda(\theta;\beta) - \Lambda(\theta';\beta) \leq 0$,

and equals zero if and only if,

$$\frac{g(\lambda_x|\theta')}{g(\lambda_x|\theta)} = 1 \quad \forall 0 < x < \beta. \quad (1)$$

It can be shown that for the immigration process (1) implies $\theta = \theta'$.

//QED.

Adding function of T to $L(\theta)$ given by (4.2.1), and multiplying by functions of T do not change the extreme points of $L(\theta)$. Since $\Lambda(\theta;\beta)$ of Proposition 4.2.2 distinguishes an exceptional point θ' , the true value of θ , Lemmas 4.2.1 and 4.2.2 suggest considering

$$\Lambda^{(T)}(\theta;\beta,\phi) = -\frac{1}{T\phi} \sum_{s=1}^{S_T} \left\{ \log g\left(\frac{\lambda_s}{T\phi} \mid \theta\right) + \frac{\hat{I}^{(T)}\left(\frac{\lambda_s}{T\phi}\right)}{(2\pi)^d \hat{p}_1 g\left(\frac{\lambda_s}{T\phi} \mid \theta\right)} - 1 \right\} \quad (4.2.2)$$

where $\phi > 0$ is an experimental parameter, and S_T is a non-decreasing function of T , and $\lambda_s = (s, 0, \dots, 0) \in \mathbb{R}^d$. In case $\beta < \infty$ S_T is taken to be $S_T = [\beta T^\phi]$, where $[\]$ denotes the greatest integer function. When $\beta = \infty$, another form for S_T is given. The parameter ϕ determines the spacing of the frequencies $\frac{1}{T\phi}$. Larger ϕ means the separation of frequencies used in (4.2.2) is smaller. Also, later a definition of optimal separation is given.

$\Lambda^{(T)}(\theta;\beta,\phi)$ takes advantage of the isotropic property of the point process.

The next few theorems verify the conditions of Brillinger's Lemma where $Q^{(T)} = -\Lambda^{(T)}(\cdot, \beta, \phi)$ and $Q = -\Lambda(\cdot, \beta)$.

Condition (i) of Brillinger's Lemma is clear. Condition (ii) is satisfied by using a canonical representation (of Dawson and Ivanoff (1978), §2). Condition (iii) is satisfied, since for each θ , T , $\Lambda^{(T)}$ is a finite sum of random variables, and $1 \leq g(\lambda_x | \theta) < \infty$ and $g(\lambda_x | \cdot)$ is continuous in θ in either θ_1 or θ_2 .

$\Lambda(\theta; \beta)$ is continuous at each θ , and so together with Lemma 4.2.2 it follows that condition (iv) of Brillinger's Lemma is satisfied.

Proposition 4.2.3

Suppose the taper h satisfies $|H_2(T\lambda)| \leq \frac{\text{constant}}{T|\lambda_0|}$ for all

$|\lambda| \geq |\lambda_0| \neq 0$ (for example $h = \chi_{\Pi[-b_j, b_j]}$).

(i) $\Lambda^{(T)}(\theta; \beta, \phi) \xrightarrow{\text{Pr}} \Lambda(\theta; \beta)$ where

$$S_T = \begin{cases} [\beta T^\phi] & \text{if } \beta < \infty \\ O(T^{\phi+\delta}) & \text{if } \beta = \infty, \text{ with } 0 < \delta < \frac{1 \wedge \phi}{2} \end{cases}$$

(ii) For each β , ϕ , and for each $\theta_1 \in \theta_1$ (or θ_2), and for all $\epsilon, \eta > 0$, there exists $U_1(\theta_1, \epsilon)$, a neighbourhood of θ_1 , and $T_1(\epsilon, \eta, \theta_1)$ such that

$$P\left\{\sup_{\theta \in U_1} |\Lambda^{(T)}(\theta; \beta, \phi) - \Lambda(\theta; \beta)| < \epsilon\right\} \geq 1 - \eta \quad \text{for all } T \geq T_1.$$

(iii) Suppose $h = \chi_{\Pi}[-b_j, b_j]$, where $b_1 = \pi$,

$$\phi = \begin{cases} 1 & \text{if } \dim R^d \geq 2 \\ \frac{1}{2} & \text{if } \dim R^d = 1 \end{cases}$$

Then for each compact set K in Θ_1 (or Θ_2),

$$\sup_K |\Lambda^{(T)}(\theta; \beta, \phi) - \Lambda(\theta; \beta)| \xrightarrow{\text{a.s.}} 0 \text{ as } T \rightarrow \infty. \quad (1)$$

Using Brillinger 1-mixing (Definition 3.3), (1) also holds with

$\phi = 1$ if $\dim R^d = 1$.

Proof:

$$(i) E\{\Lambda^{(T)}(\theta; \beta, \phi)\}$$

$$= -\frac{1}{T^\phi} \sum_{s=1}^{S_T} \left\{ \log g\left(\frac{\lambda_s}{T^\phi} | \theta\right) + E \frac{\hat{I}^{(T)}\left(\frac{\lambda_s}{T^\phi}\right)}{(2\pi)^d \hat{p}_1} \frac{1}{g\left(\frac{\lambda_s}{T^\phi} | \theta\right)} - 1 \right\}$$

$$= -\frac{1}{T^\phi} \sum_{s=1}^{S_T} \left\{ \log g\left(\frac{\lambda_s}{T^\phi} | \theta\right) + \frac{g\left(\frac{\lambda_s}{T^\phi} | \theta'\right)}{g\left(\frac{\lambda_s}{T^\phi} | \theta\right)} - 1 \right\} + O(T^{-\phi} S_T \cdot T^{-1})$$

$$\rightarrow \Lambda(\theta; \beta) \text{ since } O(T^{-\phi} S_T \cdot T^{-1}) \rightarrow 0 \text{ as } T \rightarrow \infty.$$

It now suffices, by Chebyshev's Inequality, to show

$$\text{Var}\{\hat{\Lambda}^{(T)}(\theta; \beta, \phi)\} \approx \frac{1}{T^{2\phi}} \sum_{s=1}^{S_T} \sum_{t=1}^{S_T} \frac{\text{Cov}\{\hat{I}^{(T)}(\frac{\lambda_s}{T\phi}), \hat{I}^{(T)}(\frac{\lambda_t}{T\phi})\}}{(2\pi)^{2d} p_1^{2\phi} g(\frac{\lambda_s}{T\phi}|\theta) g(\frac{\lambda_t}{T\phi}|\theta)} \quad (2)$$

$\rightarrow 0$ as $T \rightarrow \infty$.

$$\begin{aligned} & \text{Cov}(\hat{I}^{(T)}(\lambda_1), \hat{I}^{(T)}(\lambda_2)) (2\pi)^{2d} H_2(\phi)^2 \\ &= \text{cum}_4(T^{-d/2} \hat{d}^{(T)}(\lambda_1), T^{-d/2} \hat{d}^{(T)}(-\lambda_1), T^{-d/2} \hat{d}^{(T)}(\lambda_2), T^{-d/2} \hat{d}^{(T)}(-\lambda_2)) \\ &+ \text{cum}_2(T^{-d/2} \hat{d}^{(T)}(\lambda_1), T^{-d/2} \hat{d}^{(T)}(\lambda_2)) \text{cum}_2(T^{-d/2} \hat{d}^{(T)}(-\lambda_1), T^{-d/2} \hat{d}^{(T)}(-\lambda_2)) \\ &+ \text{cum}_2(T^{-d/2} \hat{d}^{(T)}(\lambda_1), T^{-d/2} \hat{d}^{(T)}(-\lambda_2)) \text{cum}_2(T^{-d/2} \hat{d}^{(T)}(-\lambda_1), T^{-d/2} \hat{d}^{(T)}(\lambda_2)) \quad (3) \end{aligned}$$

By the proof of Theorem 3.3, $\text{cum}_4(\quad) = O(T^{-d})$ uniformly in the frequencies, that is there exists constants K and T_0 such that

$$|\text{cum}_4(T^{-d/2} \hat{d}^{(T)}(\lambda_1), \dots, T^{-d/2} \hat{d}^{(T)}(\lambda_4))| \leq K T^{-d} \text{ for all } T \geq T_0.$$

In particular this holds for $\lambda_1 = -\lambda_2 = \frac{\lambda_s}{T\phi}$, $\lambda_3 = -\lambda_4 = \frac{\lambda_t}{T\phi}$. Therefore

the fourth cumulant contribution of (3) to (2) is $T^{-2\phi} O(S_T^2 T^{-d})$.

For $\beta = \infty$, that is

$$O(T^{-2\phi} T^{2\phi+2\delta} T^{-d}) = O(T^{2\delta-d}) \rightarrow 0 \text{ since } d \geq 1.$$

Similarly it tends to zero if $\beta < \infty$.

Consider $\beta = \infty$. The second term of (3) contributes to (2)

$$\begin{aligned}
 & T^{-2\phi} \sum \left\{ \left[g\left(\frac{\lambda_s}{T^\phi} \middle| \theta'\right) H_2\left(T \cdot \frac{\lambda_{s-t}}{T^\phi} + o(T^{-1}) \right) \right] \left[g\left(\frac{\lambda_t}{T^\phi} \middle| \theta'\right) H_2\left(T \cdot \frac{\lambda_{t-s}}{T^\phi} + o(T^{-1}) \right) \right] \right. \\
 & \quad \left. + \left[g\left(\frac{\lambda_s}{T^\phi} \middle| \theta\right) g\left(\frac{\lambda_t}{T^\phi} \middle| \theta\right) \right] \right\} \\
 & = T^{-2\phi} \sum \sum \frac{g\left(\frac{\lambda_s}{T^\phi} \middle| \theta'\right) g\left(\frac{\lambda_t}{T^\phi} \middle| \theta'\right)}{g\left(\frac{\lambda_s}{T^\phi} \middle| \theta\right) g\left(\frac{\lambda_t}{T^\phi} \middle| \theta\right)} H_2\left(T \cdot \frac{\lambda_{s-t}}{T^\phi}\right) H_2\left(T \cdot \frac{\lambda_{t-s}}{T^\phi}\right) \quad (4) \\
 & \quad + o(T^{2\delta-1}) + o(T^{2\delta-2})
 \end{aligned}$$

For the tapers considered if $\frac{|s-t|}{T^\phi} \geq T^a$,

$$|H_2\left(T \cdot \frac{\lambda_{s-t}}{T^\phi}\right)| \leq K \cdot T^{-(1+a)}.$$

|1st term in (4)|

$$\leq T^{-2\phi} \sum_s \sum_t \left| H_2\left(T \cdot \frac{\lambda_{s-t}}{T^\phi}\right) \right| \left| H_2\left(T \cdot \frac{\lambda_{t-s}}{T^\phi}\right) \right| c_1(\theta', \theta)$$

$$\frac{|s-t|}{T} \leq T^a$$

$$+ T^{-2\phi} \sum \sum_{|s-t| > T^{a+\phi}} \left| H_2\left(T \cdot \frac{\lambda_{s-t}}{T^\phi}\right) \right| \left| H_2\left(T \cdot \frac{\lambda_{t-s}}{T^\phi}\right) \right| c_2(\theta', \theta)$$

$$= o(T^{-2\phi} \cdot S_T T^{a+\phi}) + o(T^{-2\phi} \cdot S_T^2 T^{-2(1+a)})$$

Thus (4) $\rightarrow 0$ as $T \rightarrow \infty$ if

$$-2\phi + \phi + \delta + a + \phi = \delta + a < 0$$

(5)

$$-2\phi + 2\phi + 2\phi - 2 - 2a = 2(\delta - 1 - a) < 0$$

This holds for $a = \frac{-1}{2}$.

The other term in (3) behaves similarly.

$\text{Var}\{\Lambda^{(T)}(\theta; \beta, \phi)\} \rightarrow 0$ as $T \rightarrow \infty$ if $\beta = \infty$.

The case $\beta < \infty$ is proven similarly. Thus (i) is proved.

$$(iii) |\Lambda^{(T)}(\theta; \beta, \phi) - \Lambda^{(T)}(\theta_1; \beta, \phi)|$$

$$\leq \frac{1}{T^\phi} \left| \sum_{s=1}^{S_T} \left\{ \log g\left(\frac{1}{T^\phi} \lambda_s | \theta\right) - \log g\left(\frac{1}{T^\phi} \lambda_s | \theta_1\right) \right\} \right| \quad (6)$$

$$+ \frac{1}{T^\phi} \sum_{s=1}^{S_T} \frac{\hat{I}^{(T)}\left(\frac{1}{T^\phi} \lambda_s\right)}{(2\pi)^d \hat{p}_1} \frac{2|V_1 \alpha_1 G\left(\frac{1}{T^\phi} \lambda_s | \theta_1\right) - V \alpha G\left(\frac{1}{T^\phi} \lambda_s | \theta\right)|}{g\left(\frac{1}{T^\phi} \lambda_s | \theta\right) g\left(\frac{1}{T^\phi} \lambda_s | \theta_1\right)}$$

Let $\epsilon > 0$. Suppose first $\beta < \infty$.

It can be shown that $\exists \delta_1(\epsilon) > 0$ such that if $\theta \in S(\theta_1, \delta_1)$ then

$$2|V_1 \alpha_1 G(\lambda_y | \theta_1) - V \alpha G(\lambda_y | \theta)| \leq \frac{\epsilon}{2 \int_0^\beta g(\lambda_x | \theta') dx} \quad \forall 0 \leq y \leq \beta.$$

The second term on the right-hand side of (5) is then bounded by

$$\frac{1}{T^\phi} \sum_{s=1}^{S_T} \frac{\hat{I}^{(T)}\left(\frac{1}{T^\phi} \lambda_s\right)}{(2\pi)^d \hat{p}_1} \cdot \frac{\varepsilon^n}{2 \int_0^\beta g(\lambda_x | \theta') dx} \xrightarrow{\text{Pr}} \frac{\varepsilon}{2} \quad (7)$$

by the method used in proving parts (i) and (ii) above.

$\int_0^\beta \log g(\lambda_x | \theta) dx$ is continuous in θ . Therefore, considering the first

term on the right-hand side of (6), it follows $\exists 0 < \delta_2(\theta_1, \varepsilon) \leq \delta_1$

$\cdot \exists \cdot$ if $\theta \in \bar{S}(\theta_1, \delta_2)$, $\exists T_1(\varepsilon, \theta_1) \cdot \exists \cdot$

$$\frac{1}{T^\phi} \left| \sum_{s=1}^{S_T} \left\{ \log g\left(\frac{1}{T^\phi} \lambda_s | \theta\right) - \log g\left(\frac{1}{T^\phi} \lambda_s | \theta_1\right) \right\} \right| \leq \varepsilon \quad \forall T \geq T_1. \quad (8)$$

By (7), $\forall n > 0$, $\exists T_2(\varepsilon, n) \cdot \exists \cdot$ for all $T \geq T_2$.

$$P\left\{ \sup_{\bar{S}(\theta_1, \delta_2)} \frac{1}{T^\phi} \sum_{s=1}^{S_T} \left| \frac{\hat{I}^{(T)}\left(\frac{1}{T^\phi} \lambda_s\right)}{(2\pi)^d \hat{p}_1} \left[\frac{1}{g\left(\frac{1}{T^\phi} \lambda_s | \theta_1\right)} - \frac{1}{g\left(\frac{1}{T^\phi} \lambda_s | \theta\right)} \right] \right| > \varepsilon \right\} < n. \quad (9)$$

From (6), if $\sup_{\bar{S}(\theta_1, \delta_2)} |\Lambda^{(T)}(\theta; \beta, \phi) - \Lambda^{(T)}(\theta_1; \beta, \phi)| > 2\varepsilon$, then

$$\sup_{\bar{S}(\theta_1, \delta_2)} \frac{1}{T^\phi} \sum_{s=1}^{S_T} \left| \frac{\hat{I}^{(T)}\left(\frac{1}{T^\phi} \lambda_s\right)}{(2\pi)^d \hat{p}_1} \left[\frac{1}{g\left(\frac{1}{T^\phi} \lambda_s | \theta_1\right)} - \frac{1}{g\left(\frac{1}{T^\phi} \lambda_s | \theta\right)} \right] \right| > \varepsilon.$$

Therefore if $T \geq \max\{T_1, T_2\}$,

$$P\left\{\sup_{\theta \in \bar{S}(\theta_1, \delta_2)} |\Lambda^{(T)}(\theta; \beta, \phi) - \Lambda^{(T)}(\theta_1; \beta, \phi)| > 2\epsilon\right\} < \eta \quad (10)$$

Case: $\beta = \infty$. Let $0 < \epsilon < 1/9$. Bound the second term on the right-hand side (6) by

$$\begin{aligned} & \frac{1}{T^\phi} \sum_{s=1}^{S_T} \frac{\hat{I}^{(T)}(\frac{1}{T^\phi} \lambda_s)}{(2\pi)^d \hat{p}_1} \frac{2|V_1 \alpha_1 G(\frac{1}{T^\phi} \lambda_s | \theta_1) - V \alpha G(\frac{1}{T^\phi} \lambda_s | \theta)|}{g(\frac{1}{T^\phi} \lambda_s | \theta) g(\frac{1}{T^\phi} \lambda_s | \theta_1)} \\ & \leq \frac{1}{T^\phi} \sum_{s=1}^{[x_2 T^\phi]} \frac{\hat{I}^{(T)}(\frac{1}{T^\phi} \lambda_s)}{(2\pi)^d \hat{p}_1} 2|V_1 \alpha_1 G(\frac{1}{T^\phi} \lambda_s | \theta_1) - V \alpha G(\frac{1}{T^\phi} \lambda_s | \theta)| \\ & \quad + \frac{1}{T^\phi} \sum_{s=[x_2 T^\phi]+1}^{S_T} \frac{\hat{I}^{(T)}(\frac{1}{T^\phi} \lambda_s)}{(2\pi)^d \hat{p}_1} V_1 \alpha_1 G(\frac{1}{T^\phi} \lambda_s | \theta_1) \cdot 18 \epsilon, \end{aligned} \quad (11)$$

whenever $\theta \in \bar{S}(\theta_1, \delta_3)$.

{ By the case $\beta < \infty$ (with $\beta = x_2$), choose $0 < \delta_4(\theta_1, \epsilon) \leq \delta_3$ so

$$\text{that } \sup_{\bar{S}(\theta_1, \delta_4)} \left| \frac{1}{T^\phi} \sum_{s=1}^{[x_2 T^\phi]} \frac{\hat{I}^{(T)}(\frac{1}{T^\phi} \lambda_s)}{(2\pi)^d \hat{p}_1} 2(V_1 \alpha_1 G(\frac{1}{T^\phi} \lambda_s | \theta_1) - V \alpha G(\frac{1}{T^\phi} \lambda_s | \theta)) \right|$$

converges in probability to something bounded by $\frac{\epsilon}{2}$ as $T \rightarrow \infty$.

For the other term on the right-hand side (11), by an argument analogous to that used in proving (i) and (ii), it follows that this term converges in probability as $T \rightarrow \infty$ to

$$18\epsilon \int_{x_2}^{\infty} g(\lambda_x | \theta') v_{1\alpha_1} G(\lambda_x | \theta_1) dx = c(\theta_1, \theta') \epsilon.$$

Now a statement like (10) follows. Part (iii) of the proposition follows by the continuity of $\Lambda(\theta; \delta)$.

(iv) Consider the second term of the right-hand side of (6). It is a sum of periodogram, and its cumulants can be calculated. In particular, consider the bound (11) of it.

$$\sup_{\bar{S}(\theta_1, \delta_3)} \frac{1}{T^\phi} \sum_{s=1}^{S_T} \frac{\hat{I}^{(T)}(\frac{1}{T^\phi \lambda_s})}{(2\pi)^d \hat{p}_1} 2 |v_{1\alpha_1} G(\frac{1}{T^\phi \lambda_s} | \theta_1) - v_{\alpha} G(\frac{1}{T^\phi \lambda_s} | \theta)|$$

$$\leq \epsilon \cdot c_1(\theta_1, \theta') \frac{1}{T^\phi} \sum_{s=1}^{x_2 T^\phi} \frac{\hat{I}^{(T)}(\frac{1}{T^\phi \lambda_s})}{(2\pi)^d \hat{p}_1}$$

(12)

$$+ \epsilon c_2(\theta_1, \theta') \frac{1}{T^\phi} \sum_{x_2 T^\phi + 1}^{S_T} \frac{\hat{I}^{(T)}(\frac{1}{T^\phi \lambda_s})}{(2\pi)^d \hat{p}_1} v_{1\alpha_1} G(\frac{1}{T^\phi \lambda_s} | \theta_1)$$

Rather than writing out details twice, it follows from the method of the proof of Theorem 4.2.5 (iii) that an ℓ^{th} cumulant κ_ℓ of each of the two sums in (12) is $O(T^{-\ell d/3})$. Therefore by Appendix I

$$E \left\{ \frac{1}{T^\phi} \sum_{s=1}^{x_2 T^\phi} \frac{\hat{I}^{(T)}(\frac{1}{T^\phi \lambda_s})}{(2\pi)^d \hat{p}_1} - E \frac{1}{T^\phi} \sum_{s=1}^{x_2 T^\phi} \frac{\hat{I}^{(T)}(\frac{1}{T^\phi \lambda_s})}{(2\pi)^d \hat{p}_1} \right\}^\ell = O(T^{-\ell d/3}), \text{ if } \ell = 2 \text{ or } 4.$$

$$E \left\{ \frac{1}{T^\phi} \sum_{s=1}^{x_2 T^\phi+1} \frac{\hat{I}^{(T)}(\frac{1}{T^\phi \lambda_s})}{(2\pi)^d \hat{p}_1} - V_1 \alpha_1 G(\frac{1}{T^\phi \lambda_s} | \theta_1) - E \left\{ \frac{1}{T^\phi} \sum_{s=1}^{x_2 T^\phi+1} \frac{\hat{I}^{(T)}(\frac{1}{T^\phi \lambda_s})}{(2\pi)^d \hat{p}_1} \right\} \right\}^\ell = O(T^{-\ell d/3}) \text{ if } \ell = 2 \text{ or } 4. \quad (13)$$

$$V_1 \alpha_1 G(\frac{1}{T^\phi \lambda_s} | \theta_1)^\ell = O(T^{-\ell d/3}) \text{ if } \ell = 2 \text{ or } 4.$$

If $\ell = 2$, $\frac{\ell d}{3} > 1$ if $d \geq 2$, and if $\ell = 4$, $\frac{\ell d}{3} > 1 \quad \forall d \geq 1$.

Therefore the argument below may be trivially modified to use fewer cumulant conditions if $d \geq 2$.

The following inequality is proven in exactly the same way as Chebyshev's Inequality is (in Chung (1974) for example):

$$P\{|Y - EY| > \epsilon\} \leq \epsilon^{-2k} E\{|Y - EY|^{2k}\} \quad k = 1, 2, \dots$$

Therefore for ϵ given, $\exists T_0$.

$$P\left\{ \frac{1}{T^\phi} \sum_{s=1}^{x_2 T^\phi} \frac{\hat{I}^{(T)}(\frac{1}{T^\phi \lambda_s})}{(2\pi)^d \hat{p}_1} - E \left\{ \frac{1}{T^\phi} \sum_{s=1}^{x_2 T^\phi} \frac{\hat{I}^{(T)}(\frac{1}{T^\phi \lambda_s})}{(2\pi)^d \hat{p}_1} \right\} > \epsilon \right\} \leq \frac{\epsilon^{-4k_1}}{T^{4d/3}}.$$

for $T \geq T_0(\epsilon)$. Let

$$S_{T,\epsilon} = \left\{ \omega: \frac{1}{T^\phi} \sum_{s=1}^{x_2 T^\phi} \frac{\hat{I}^{(T)}(\frac{1}{T^\phi \lambda_s})}{(2\pi)^d \hat{p}_1}(\omega) - E \sum_{s=1}^{x_2 T^\phi} \frac{\hat{I}^{(T)}(\frac{1}{T^\phi \lambda_s})}{(2\pi)^d \hat{p}_1} > \epsilon \right\}.$$

Therefore $\sum P\{S_{T,\epsilon}\} < \infty$ where the sum is over an equi-spaced subset of Q beginning at $T_1 \geq T_0$. Thus by Borel-Cantelli $P\{S_{T,\epsilon} \text{ i.o.}\} = 0$ when the indices T are in the sequence. If $N_{T_1,\epsilon}$ is the exceptional set for the sequence starting with T_1 , then $N_\epsilon = \bigcup_{T \geq T_0} N_{T,\epsilon}$ is a null set.

Thus $P\{S_{T,\epsilon} \text{ i.o.}, T \geq T_0\} = 0$. The result now will follow by a continuity argument on the Fourier transforms and by compactness of $K \subseteq \Theta_1$. //QED.

Proposition 4.2.3 (i) and (ii) show conditions (v) and (vi) of Brillinger's Lemma are satisfied for the immigration process.

Theorem 4.2.4

Let $\theta' \in U_1 \subset U_2$ where U_1 is open and U_2 is compact (in Θ_1 or Θ_2). Let $\hat{\theta}_T(U_2) \in U_2 \Rightarrow \Lambda^{(T)}(\hat{\theta}_T; \beta, \phi) \geq \Lambda^{(T)}(\theta; \beta, \phi) \quad \forall \theta \in U_2$.

Then under the conditions of Theorem 4.2.3, $P\{\hat{\theta}_T \in U_2 \setminus U_1\} \rightarrow 0$ as $T \rightarrow \infty$.

If in addition a taper $\chi_{\Pi}[-b_j, b_j]$ is used with $b_1 = \pi$ (cf (iv) of Proposition 4.2.3), then $\hat{\theta} \in U_1$ a.s. as $T \rightarrow \infty$, if $\phi = \frac{1}{2}$ or 1.

Proof: $\Lambda^{(T)}$ is continuous in θ , so $\hat{\theta}_T$ is a random variable.

Let $\epsilon = \Lambda(\theta'; \beta) - \sup_{U_2 \setminus U_1} \Lambda(\theta; \beta)$. $\epsilon > 0$ since Λ has a unique maximum

at θ' and since $U_2 \setminus U_1$ is compact. For each $\theta_1 \in U_2$, choose $\delta_1(\theta_1, \epsilon) \Rightarrow$

if $S_{T, \theta_1, \epsilon} = \{ \sup_{\bar{S}(\theta_1, \delta_1)} |\Lambda^{(T)}(\theta; \beta, \phi) - \Lambda(\theta_1, \beta)| > \frac{\epsilon}{2} \}$, then

$P\{S_{T, \theta_1, \epsilon}\} \rightarrow 0$ as $T \rightarrow \infty$, and such that

$$\sup_{S(\theta_1, \delta_1)} |\Lambda(\theta; \beta) - \Lambda(\theta_1; \beta)| < \frac{\varepsilon}{2}.$$

$$S(\theta_1, \delta_1)$$

$U_2 \setminus U_1$ is compact, so choose a finite subcover $S(\theta_1, \delta_1), \dots, S(\theta_\ell, \delta_\ell)$.

Also let $N_{T, \theta', \varepsilon} = \{|\Lambda^{(T)}(\theta'; \beta, \phi) - \Lambda(\theta'; \beta)| > \frac{\varepsilon}{2}\}$. Let $\eta > 0$.

Choose $T_i(\theta_i, \varepsilon, \eta) \rightarrow \infty$ $P\{S(\theta_i, \delta_i)\} < \frac{\eta}{\ell+1} \quad \forall T \geq T_i, 1 \leq i \leq \ell$.

Choose $T_{\ell+1}(\theta', \varepsilon, \eta) \rightarrow \infty$ $P\{N_{T, \theta', \varepsilon}\} < \frac{\eta}{\ell+1} \quad \forall T \geq T_{\ell+1}$.

$$\text{Let } T_0 = \max_{1 \leq i \leq \ell+1} T_i, \quad N_\varepsilon = N_{T, \theta', \varepsilon} \cup \bigcup_{i=1}^{\ell} S_{T, \theta_i, \varepsilon} \quad (1)$$

Then $P\{N_\varepsilon\} < \eta$.

Let $\theta \in U_2 \setminus U_1$. Choose $\theta_i \rightarrow \theta \in S(\theta_i, \delta_i)$. If $\omega \in \Omega \setminus N$, $T \geq T_0$,

$$\Lambda^{(T)}(\theta'; \beta, \phi)(\omega) \geq \Lambda(\theta'; \beta) - \frac{\varepsilon}{4} \geq \Lambda(\theta_i; \beta) + \varepsilon - \frac{\varepsilon}{4} \geq \Lambda^{(T)}(\theta; \beta, \phi)(\omega) + \frac{\varepsilon}{4}$$

that is, on N^c , $\hat{\theta}_T \in U_1$, so $P\{\hat{\theta}_T \in U_1\} \geq 1 - \eta \quad \forall T \geq T_0$.

For the almost sure statement, the proof is analogous, but we now work with null sets as in (iv) of Theorem 4.2.3. //QED.

Corollary to Theorem 4.2.4

There exists $\hat{\theta}_T \rightarrow \theta$ $\frac{\partial \Lambda^{(T)}(\theta; \beta, \phi)}{\partial \theta} \big|_{\theta = \hat{\theta}_T} = 0$, and such that

$\hat{\theta}_T \xrightarrow{\text{Pr}} \theta'$. For the special taper $\chi_{\Pi}[-b_j, b_j]$, $b_1 = \pi$ of Proposition 4.2.3

and Theorem 4.2.4, if $\phi = \frac{1}{2}$ or 1, then $\hat{\theta}_T \rightarrow \theta'$ a.s.

The same thing holds if $\hat{\theta}_T \in U_2$ as in Theorem 4.2.4, except now $\hat{\theta}_T$

is a root of $\frac{\partial \Lambda^{(T)}(\theta; \beta, \phi)}{\partial \theta}$ with probability tending to 1 as $T \rightarrow \infty$,

or almost surely as $T \rightarrow \infty$.

Proof: By continuity $\Lambda(\theta'; \beta) - \sup_{U_2 \setminus S(\theta', \delta)} \Lambda(\theta; \beta) \rightarrow 0$ as $\delta \rightarrow 0$.

Thus $\epsilon_m \rightarrow 0$ where,

$$\epsilon_m = \Lambda(\theta'; \beta) - \sup_{U_2 \setminus S(\theta', \frac{1}{m})} \Lambda(\theta; \beta).$$

For the almost sure parts, for example, we work with $N_m = \{\hat{\theta}_T \in U_2 \mid S(\theta'; \delta_m)\}$.

N_m is a null set by Theorem 4.2.4, and so $N = \bigcup_{m=1}^{\infty} N_m$ is a null set.

If $\omega \in \Omega \setminus N$, then $\forall m \exists T_m \cdot |\hat{\theta}_{T_m}(\omega) - \theta'| < \frac{1}{m}$. The other parts are similar. //QED.

Remark: In Theorem 4.2.4, the compactness of U_2 is used to conclude

$$\epsilon = \Lambda(\theta'; \beta) - \sup_{U_2 \setminus U_1} \Lambda(\theta; \beta) > 0.$$

If we could show $\Lambda(\theta'; \beta) - \sup_{\theta \in U_1} \Lambda(\theta; \beta) > 0$ and

$$\sup_{\theta} |\Lambda^{(T)}(\theta; \beta, \phi) - \Lambda(\theta; \beta)| \rightarrow 0 \text{ in pr. or a.s.,}$$

then we could conclude "the global maximum point of $\Lambda^{(T)}(\cdot, \beta, \phi)$ is consistent".

For the remainder of this section we use the taper χ_A where A is the rectangle $\Pi[-b_j, b_j]$. Without loss of generality $b_1 = \pi$.

Let $\Lambda_j^{(T)}(\theta; \beta, \phi) = \frac{\partial}{\partial \theta} \Lambda^{(T)}(\theta; \beta, \phi)$, θ_j a component of θ .

This derivative clearly exists since

$$\Lambda_j^{(T)}(\theta; \beta, \phi) = -\frac{1}{T\phi} \sum \frac{\partial \log g(\frac{\lambda_s}{T\phi} | \theta)}{\partial \theta_j} \left[1 - \frac{\hat{I}^{(T)}(\frac{\lambda_s}{T\phi})}{g(\frac{\lambda_s}{T\phi} | \theta)} \right]$$

Theorem 4.2.5

Let $h = \chi_A$, $A = \Pi[-b_j, b_j]$, $b_1 = \pi$. Consider the immigration process.

(i) if $\beta = \infty$, $E\left\{ \frac{\partial \Lambda^{(T)}(\theta'; \beta, \phi)}{\partial \theta} \right\} = O(T^{-1+\delta})$, $0 < \delta < \frac{1}{2}\phi$,

(ii) if $\beta < \infty$, $E\left\{ \frac{\partial \Lambda^{(T)}(\theta'; \beta, \phi)}{\partial \theta} \right\} = O(T^{-1})$,

(iii) $T^{\phi/2} \left\{ \frac{\partial \Lambda^{(T)}(\theta'; \beta, \phi)}{\partial \theta} - E\left\{ \frac{\partial \Lambda^{(T)}(\theta'; \beta, \phi)}{\partial \theta} \right\} \right\}$,

$$\xrightarrow{D} N(0, V), \quad \text{where } \phi = \begin{cases} \frac{1}{2} & \text{if } d = 1 \\ 1 & \text{if } d \geq 2 \end{cases}$$

$$\text{where } V_{jk} = \int_0^\beta \left(\frac{\partial \log g(\lambda_x | \theta')}{\partial \theta_j} \frac{\partial \log g(\lambda_x | \theta')}{\partial \theta_k} \right) dx,$$

(iv) if the process is also Brillinger 1-mixing, then with $\phi = 1$ and $\dim R^d = 1$,

$$T^{\frac{1}{2}}(\hat{\theta}_T - \theta') \xrightarrow{D} N(0, V)$$

$$\text{where } V_{jk}(\theta'; \beta) = \int_0^\beta \left(\frac{\partial \log g(\lambda_x | \theta')}{\partial \theta_j} \frac{\partial \log g(\lambda_x | \theta')}{\partial \theta_k} \right) dx$$

$$+ \frac{(2\pi)H_4(0)}{H_2(0)^2} \int_0^\beta \int_0^\beta \frac{\partial \log g(\lambda_x | \theta')}{\partial \theta_j} \frac{\partial \log g(\lambda_x | \theta')}{\partial \theta_k} \cdot \frac{f^{(4)}(\lambda_x, -\lambda_x, \lambda_y | \theta')}{f(\lambda_x | \theta') f(\lambda_y | \theta')} dx dy,$$

where $f^{(4)}$ is a fourth order cumulant spectral density.

Proof:

$$(i) \text{ and } (ii) \quad E\{\Lambda_j^{(T)}(\theta'; \beta, \phi)\}$$

$$= - \frac{1}{T^\phi} \sum_{s=1}^{S_T} \frac{\partial \log g(\frac{1}{T^\phi} \lambda_s | \theta')}{\partial \theta_j} \left[1 - E \left\{ \frac{\hat{I}^{(T)}(\frac{1}{T^\phi} \lambda_s)}{(2\pi)^d \hat{p}_1} \cdot \frac{1}{g(\frac{1}{T^\phi} \lambda_s | \theta')} \right\} \right]$$

$$= - \frac{1}{T^\phi} \sum_{s=1}^{S_T} \frac{\partial \log g(\frac{1}{T^\phi} \lambda_s | \theta')}{\partial \theta_j} O(T^{-1}), \text{ where } O(T^{-1}) \text{ is uniform}$$

(by Corollary to Theorem 3.3)

$$= O(T^{-\delta} S_T T^{-1})$$

$$= \begin{cases} O(T^{\delta-1}) & \text{if } \beta = \infty \\ O(T^{-1}) & \text{if } \beta < \infty \end{cases}$$

(iii) Consider a joint ℓ^{th} cumulant of $T^{\phi/2} \frac{\partial \Lambda^{(T)}(\theta'; \beta, \phi)}{\partial \theta_j}$, $\ell \geq 3$

$$\text{cum}_\ell \left\{ T^{\phi/2} \frac{\partial}{\partial \theta_{j_1}} \Lambda^{(T)}(\theta'; \beta, \phi), \dots, T^{\phi/2} \frac{\partial}{\partial \theta_{j_\ell}} \Lambda^{(T)}(\theta'; \beta, \phi) \right\}$$

$$= T^{\ell\phi/2} \left(\frac{1}{T^\phi} \right)^\ell \sum_{s_1=1}^{S_T} \sum_{s_\ell=1}^{S_T} \prod_{k=1}^{\ell} \frac{\partial \log g(\frac{1}{T^\phi} \lambda_{s_k} | \theta')}{\partial \theta_{j_k}} \cdot \frac{1}{f(\frac{1}{T^\phi} \lambda_{s_k} | \theta')} \quad (1)$$

$$\cdot \text{cum}_\ell \left\{ \hat{I}^{(T)}\left(\frac{1}{T^\phi} \lambda_{s_1}\right), \dots, \hat{I}^{(T)}\left(\frac{1}{T^\phi} \lambda_{s_\ell}\right) \right\}$$

Again we use the calculations of Leonov and Shiryaev (1959). Recall from the proof of Theorem 3.3

$$\text{cum}_1\{T^{-d/2}\hat{d}^{(T)}(\lambda)\} \equiv 0,$$

$$\text{cum}_2\{T^{-d/2}\hat{d}^{(T)}(\lambda_1), T^{-d/2}\hat{d}^{(T)}(\lambda_2)\} = (2\rho)^d f(\lambda_1|\theta') H_2(T \cdot (\lambda_1 - \lambda_2)) + O(T^{-1}), \quad (2)$$

$$\text{cum}_\ell\{T^{-d/2}\hat{d}^{(T)}(\lambda_1), \dots, T^{-d/2}\hat{d}^{(T)}(\lambda_\ell)\} = O(T^{d(1-\ell/2)}), \quad \ell \geq 3,$$

where the "0" terms are uniform in the λ 's. In particular, since $\sin(\pi n) = 0$ for $n \in \mathbb{Z}$, if $\phi = 1$, and $s, t \in \mathbb{Z}$,

$$\text{cum}_2\{T^{-d/2}\hat{d}^{(T)}(\frac{1}{T^\phi\lambda_s}), T^{-d/2}\hat{d}^{(T)}(\frac{1}{T^\phi\lambda_t})\} = \begin{cases} (2\pi)^d f(\frac{1}{T^\phi\lambda_s}|\theta') H_2(0) + O(T^{-1}) & \text{if } s = t \\ O(T^{-1}) & \text{if } s \neq t \end{cases},$$

$$\text{cum}_\ell\{\hat{I}^{(T)}(\frac{1}{T^\phi\lambda_{s_1}}), \dots, \hat{I}^{(T)}(\frac{1}{T^\phi\lambda_{s_\ell}})\} (2\pi)^{d\ell} H_2(0)^\ell$$

$$= \text{cum}_\ell\{T^{-d/2}\hat{d}^{(T)}(\frac{1}{T^\phi\lambda_{s_1}}), T^{-d/2}\hat{d}^{(T)}(-\frac{1}{T^\phi\lambda_{s_1}}), \dots, T^{-d/2}\hat{d}^{(T)}(\frac{1}{T^\phi\lambda_{s_\ell}}),$$

$$T^{-d/2}\hat{d}^{(T)}(-\frac{1}{T^\phi\lambda_{s_\ell}})\}.$$

The table we must consider in order to form indecomposable partitions is of the form of TABLE 1 for each cum_ℓ term in (1). Thus

TABLE 1

(1,1)	(1,2)
(2,1)	(2,2)
$\vdots$	$\vdots$
(ℓ ,1)	(ℓ ,2)

$$\begin{aligned}
 & \text{cum}_{\ell} \left\{ T^{\phi/2} \frac{\partial \Lambda^{(T)}(\theta'; \beta, \phi)}{\partial \theta_{j_1}}, \dots, T^{\phi/2} \frac{\partial \Lambda^{(T)}(\theta'; \beta, \phi)}{\partial \theta_{j_{\ell}}} \right\} \\
 &= \sum_{\mathbf{P} = \{P_1, \dots, P_t\}}^* T^{\ell\phi/2} \left(\frac{1}{T^{\phi}} \right)^{\ell} \sum_{s_1=1}^{S_T} \dots \sum_{s_{\ell}=1}^{S_T} \prod_{k=1}^{\ell} \frac{\partial \log g\left(\frac{1}{T^{\phi}} s_k \mid \theta'\right)}{\partial \theta_{j_k}} \cdot \frac{1}{f\left(\frac{1}{T^{\phi}} s_k \mid \theta'\right)} \\
 & \quad \cdot C(P_1) \dots C(P_k),
 \end{aligned}$$

where $\sum^*$ is the sum over all indecomposable partitions $\mathbf{P} = \{P_1, \dots, P_t\}$ of TABLE 1 and $C(P_i)$ is the joint cumulant of the r.v.'s whose TABLE 1 indices are in P_i . Thus, since there are only finitely many indecomposable partitions, we may consider each indecomposable partition separately. Also we need only consider indecomposable partitions with no sets of size 1, because if there is a P_i of size 1, $C(P_i) = \text{cum}_1 \{ T^{-d/2} \hat{\Lambda}^{(T)}(\lambda) \} = 0$, and so $C(P_1)C(P_2) \dots C(P_t) = 0$.

Consider an indecomposable partition $\mathbf{P}$. Let k be the number of sets of size 2, and i_0 the number of sets of size bigger than 2, say of size $n_1, \dots, n_{i_0}$. If $k = \ell$, then $i_0 = 0$. If $0 \leq k < \ell$,

then $1 \leq i_0 \leq \frac{2}{3}(\ell-k)$. When $k = 0$ and $\ell = 0 \pmod{3}$, there exists an indecomposable partition with $i_0 = \frac{2}{3}\ell$. For example, in TABLE 1, takes sets consisting of the first three elements in column 1, then the next three, etc. This partitions the first column, since $\ell = 0 \pmod{3}$. Do the same thing with the second column, starting at the second element, putting the first element in the last partition set.

Let the indecomposable partition being considered be $\mathcal{P} = \{P_{1,1}, \dots, P_{1,k}, P_1, \dots, P_{i_0}\}$ where $P_{1,1}, \dots, P_{1,k}$ are sets of size $n_1, \dots, n_{i_0}$, $n_j \geq 3$. The term we are considering is

$$T^{\ell\phi/2} \left(\frac{1}{T^\phi}\right)^\ell \sum_{s_1} \dots \sum_{s_\ell} \prod_{i=1}^{\ell} \frac{\partial \log g\left(\frac{1}{T^\phi} \lambda_{s_i} \mid \theta'\right)}{\partial \theta_j} \cdot \frac{1}{f\left(\frac{1}{T^\phi} \lambda_{s_i} \mid \theta'\right)} \cdot C(P_{1,1}) \dots C(P_{1,k}) C(P_1) \dots C(P_{i_0}) \quad (3)$$

If $k = 0$

$$T^{\ell\phi/2} \left(\frac{1}{T^\phi}\right)^\ell \sum_{s_1} \dots \sum_{s_\ell} \prod_{i=1}^{\ell} \frac{\partial}{\partial \theta_{j_i}} \log g\left(\frac{1}{T^\phi} \lambda_{s_i} \mid \theta'\right) \frac{1}{f\left(\frac{1}{T^\phi} \lambda_{s_i} \mid \theta'\right)} \cdot C(P_1) \dots C(P_{i_0})$$

$$<< T^{\ell\phi/2} \left(\frac{1}{T^\phi}\right)^\ell \sum_{s_1} \dots \sum_{s_\ell} \prod_{i=1}^{\ell} \left| \frac{\partial}{\partial \theta_{j_i}} \log g\left(\frac{1}{T^\phi} \lambda_{s_i} \mid \theta'\right) \frac{1}{f\left(\frac{1}{T^\phi} \lambda_{s_i} \mid \theta'\right)} \right|$$

$$\prod_{j=1}^{i_0} O\left(T^{d(1 - \frac{n_j}{2})}\right)$$

(by (2))

$$= O(T^{\ell\phi/2} T^{d(i_0 - 2\ell/2)}) , \text{ since } n_1 + \dots + n_{i_0} = 2(\ell - k) = 2\ell \text{ if } k=0$$

$$<< O(T^{\ell\phi/2} T^{d(2\ell/3 - \ell)}) , \text{ since } 1 \leq i_0 \leq 2\ell/3$$

$$= O(T^{(\ell/2 - d/3)}) , \quad (4)$$

which tends to zero as $T \rightarrow \infty$ if $\phi = 1$ when $\dim R^d \geq 2$, or if $\phi = \frac{1}{2}$ when $\dim R^d = 1$.

If $k \neq 0$,

$$|T^{\ell\phi/2} \left(\frac{1}{T^\phi}\right)^\ell \sum_{s_1} \sum_{s_\ell} \prod_{i=1}^{\ell} \frac{\partial}{\partial \theta_{j_i}} \log g\left(\frac{1}{T^\phi} \lambda_{s_i} | \theta'\right) \frac{1}{f\left(\frac{1}{T^\phi} \lambda_{s_i} | \theta'\right)} C(P_{i,1}) \dots C(P_{1,k})$$

$$\cdot C(P_1) \dots C(P_{i_0})|$$

$$<< O(T^{d(i_0 - (\ell - k))} T^{\ell\phi/2} \left(\frac{1}{T^\phi}\right)^\ell \sum_{s_1} \dots \sum_{s_\ell} \prod_{i=1}^{\ell} \left| \frac{\partial}{\partial \theta_{j_i}} \log g\left(\frac{1}{T^\phi} \lambda_{s_i} | \theta'\right) \right|$$

$$\cdot |C(P_{1,1})| \dots |C(P_{1,k})| . \quad (5)$$

Since the partition is indecomposable, each $C(P_{1,i})$ refers to (s_{i_1}, s_{i_2}) where $i_1 \neq i_2$ (cf (2)). (By (2), we may write (5) as a sum of 2^k terms). A particular one of these involves j of the $O(T^{-1})$ terms and $k-j$ of the $(2\pi)^d f(\cdot | \theta') H_2(T \cdot \frac{1}{T^\phi} \lambda_{s_{i_1} - s_{i_2}})$ terms.

Consider the case $\dim R^d = 1$ and $\phi = \frac{1}{2}$.

$$H_2\left(T \cdot \frac{1}{T\phi} s\right) = \begin{cases} H_2(0) & \text{if } s = 0 \\ \frac{2 \sin(\pi \sqrt{T} s)}{\sqrt{T} \cdot s} \prod_{j=1}^d (2b_j) & s \neq 0 \end{cases}$$

If $s \neq 0$, $|H_2(T \cdot \frac{1}{T\phi} s)| \leq \frac{\text{const.}}{\sqrt{T}}$ since $s \in \mathbb{Z} \setminus \{0\}$

$$|T^{\ell/4} \left(\frac{1}{T^{\frac{1}{2}}}\right)^{\ell} \sum_{s_1} \dots \sum_{s_{\ell}} \prod_{i=1}^{\ell} \frac{\partial}{\partial \theta_{j_i}} \log g\left(\frac{1}{T\phi} s_i \mid \theta'\right) \frac{1}{f\left(\frac{1}{T\phi} s_i \mid \theta'\right)} C(P_{1,1}) \dots C(P_{1,k})$$

$$\cdot C(P_1) \dots C(P_{i_0})|$$

$$<< O(T^{-(\ell-k)/3}) T^{\ell/4} \left(\frac{1}{T^{\frac{1}{2}}}\right)^{\ell} \sum_{s_1} \dots \sum_{s_{\ell}} \prod_{i=1}^{\ell} \left| \frac{\partial}{\partial \theta_{j_i}} \log g\left(\frac{1}{T\phi} s_i \mid \theta'\right) \right|$$

$$\left(\frac{\text{const.}}{\sqrt{T}} \right)^{k-j} O(T^{-1})^j$$

$$= O(T^{-(\ell-k)/3} \cdot T^{\ell/4} \cdot T^{-k/2} \cdot T^{-j/2})$$

$$= O(T^{-\ell/12} \cdot T^{-k/6} \cdot T^{-j/2}) \rightarrow 0 \text{ as } T \rightarrow \infty.$$

Consider the case $\dim R^d \geq 2$ and $\phi = 1$. Therefore if $s \in \mathbb{N}$

$$H_2\left(T \cdot \frac{1}{T\phi} \lambda_s\right) = H_2(\lambda_s) = \begin{cases} H_2(0) & \text{if } s = 0 \\ 0 & \text{if } s \neq 0 \end{cases}$$

Then

$$\begin{aligned} & |T^{\ell/2} \left(\frac{1}{T\phi}\right)^\ell \sum_{s_1} \dots \sum_{s_\ell} \prod_{i=1}^{\ell} \frac{\partial}{\partial \theta_{j_i}} \log g\left(\frac{1}{T} \lambda_{s_i} | \theta'\right) \frac{1}{f\left(\frac{1}{T} \lambda_s | \theta'\right)} \\ & \quad \cdot C(P_{1,1}) \dots C(P_{1,k}) C(P_1) \dots C(P_{i_0})| \\ & \ll O(T^{-d(\ell-k)/3}) T^{\ell/2} \left(\frac{1}{T}\right)^\ell \sum_{s_1} \dots \sum_{s_\ell} \prod_{i=1}^{\ell} \left| \frac{\partial}{\partial \theta_{j_i}} \log g\left(\frac{1}{T} \lambda_{s_i} | \theta'\right) \right| \\ & \quad \cdot |C(P_{1,1})| \dots |C(P_{1,k})| \\ & = O(T^{-d(\ell-k)/3}) T^{\ell/2} \left(\frac{1}{T}\right)^\ell \sum_{s_1} \dots \sum_{s_\ell} \prod_{i=1}^{\ell} \left| \frac{\partial}{\partial \theta_{j_i}} \log g\left(\frac{1}{T} \lambda_{s_i} | \theta'\right) \right| \\ & \quad \prod_{1}^{k-j} H_2(\lambda_{\pm s_{i_1} \pm s_{i_2}}) O(T^{-j}), \end{aligned}$$

since $P_{1,t}$ is of the form $\{s_{i_1}, s_{i_2}\}$, $\{s_{i_1}, -s_{i_2}\}$, $\{-s_{i_1}, s_{i_2}\}$ or $\{-s_{i_1}, -s_{i_2}\}$,

and where $\prod_{1}^{k-j} H_2(\lambda_{\pm s_{i_1} \pm s_{i_2}})$ means the product over $k-j$ of the partition

elements of size 2. The partition is indecomposable, so the above sum is over only $\ell-(k-j)+1$ terms $s_1, \dots, s_\ell$, instead of all ℓ . The above term is therefore

$$O(T^{-d(\ell-k)/3}) \frac{T^{\ell/2}}{T^{(k-j)-1}} \cdot O(T^{-j}) = O(T^{-d(\ell-k)/3}) \cdot O(T^{1-\ell/2})$$

Since $\ell-k \geq 0$, and for some terms of interest, $k=\ell$, the above

terms tend to zero if and only if $1 - \frac{\ell}{2} < 0$, that is, $\ell \geq 3$.

Therefore all the third and higher order cumulants tend to zero for the case R^1 and R^d , $d \geq 2$. It only remains now to verify the second moment conditions.

$$\begin{aligned} & T^\phi \text{Cov}\{\Lambda_{j_1}^{(T)}(\theta'; \beta, \phi), \Lambda_k^{(T)}(\theta'; \beta, \phi)\} \\ &= T^\phi \left(\frac{1}{T^\phi}\right)^2 \sum_{s=1}^S \sum_{t=1}^T \frac{\partial \log g(\frac{1}{T^\phi} \lambda_s | \theta')}{\partial \theta_j} \frac{\partial \log g(\frac{1}{T^\phi} \lambda_t | \theta')}{\partial \theta_k} \\ &= \frac{\text{cov}\{\hat{I}^{(T)}(\frac{1}{T^\phi} \lambda_s), \hat{I}^{(T)}(\frac{1}{T^\phi} \lambda_t)\}}{f(\frac{2\pi}{T^\phi} \lambda_s | \theta') f(\frac{2\pi}{T^\phi} \lambda_t | \theta')} \end{aligned} \quad (6)$$

$$\begin{aligned} & \text{Cov}\{\hat{I}^{(T)}(\frac{1}{T^\phi} \lambda_s), \hat{I}^{(T)}(\frac{1}{T^\phi} \lambda_t)\} (2\pi)^{2d} H_2(0)^2 \\ &= \text{cum}_4\{T^{-d/2} \hat{I}^{(T)}(\frac{1}{T^\phi} \lambda_s), T^{-d/2} \hat{I}^{(T)}(-\frac{1}{T^\phi} \lambda_s), T^{-d/2} \hat{I}^{(T)}(\frac{1}{T^\phi} \lambda_t), T^{-d/2} \hat{I}^{(T)}(-\frac{1}{T^\phi} \lambda_t)\} \\ &+ \text{cum}_2\{T^{-d/2} \hat{I}^{(T)}(\frac{1}{T^\phi} \lambda_s), T^{-d/2} \hat{I}^{(T)}(\frac{1}{T^\phi} \lambda_t)\} \\ &+ \text{cum}_2\{T^{-d/2} \hat{I}^{(T)}(-\frac{1}{T^\phi} \lambda_s), T^{-d/2} \hat{I}^{(T)}(-\frac{1}{T^\phi} \lambda_t)\} \end{aligned}$$

$$\begin{aligned}
& + \text{cum}_2\{\tau^{-d/2}\hat{d}(\tau)\left(\frac{1}{\tau^\phi}\lambda_s\right), \tau^{-d/2}\hat{d}(\tau)\left(-\frac{1}{\tau^\phi}\lambda_t\right)\} \\
& \cdot \text{cum}_2\{\tau^{-d/2}\hat{d}(\tau)\left(-\frac{1}{\tau^\phi}\lambda_s\right), \tau^{-d/2}\hat{d}(\tau)\left(\frac{1}{\tau^\phi}\lambda_t\right)\} .
\end{aligned} \tag{7}$$

The fourth cumulant contribution to (6) is bounded by

$$\begin{aligned}
& \tau^\phi \left(\frac{1}{\tau^\phi}\right)^2 \sum_s \dots \sum_t \left| \frac{\partial \log g\left(\frac{1}{\tau^\phi}\lambda_s \mid \theta'\right)}{\partial \theta_j} \right| \left| \frac{\partial \log g\left(\frac{1}{\tau^\phi}\lambda_t \mid \theta'\right)}{\partial \theta_k} \right| = O(\tau^{-d}) \\
& = O(\tau^{\phi-d})
\end{aligned}$$

$$\ll \begin{cases} O(\tau^{-1/2}) & \text{if } \phi = 1/2 \text{ and } d = 1 \\ O(\tau^{-1}) & \text{if } \phi = 1 \text{ and } d \geq 2. \end{cases}$$

$\rightarrow 0$ as $\tau \rightarrow \infty$.

The last term in (7) contributes to (6) as follows.

(a) if $d = 1$, $\phi = 1/2$,

$$\tau^{1/2} \left(\frac{1}{\sqrt{\tau}}\right)^2 \sum_{s=1}^{\tau} \sum_{t=1}^{\tau} \left| \frac{\partial \log g\left(\frac{1}{\sqrt{\tau}}\lambda_s \mid \theta'\right)}{\partial \theta_j} \right| \left| \frac{\partial \log g\left(\frac{1}{\sqrt{\tau}}\lambda_t \mid \theta'\right)}{\partial \theta_k} \right|$$

$$\frac{\sin(\pi\sqrt{\tau}(s+t))}{\sqrt{\tau}(s+t)} + O(\tau^{-2})$$

$$= O(\tau^{-1/2}) \rightarrow 0 \quad \text{as } \tau \rightarrow \infty$$

(b) if $d \geq 2$, $\phi = 1$, so $H_2(\lambda_s) = 0$ for $s \in \mathbb{Z} \setminus \{0\}$.

$$T\left(\frac{1}{T}\right)^2 \sum_{s=1}^{S_T} \sum_{t=1}^{S_T} \frac{\partial}{\partial \theta_j} \log g\left(\frac{1}{T}\lambda_s | \theta'\right) \frac{\partial}{\partial \theta_k} \log g\left(\frac{1}{T}\lambda_t | \theta'\right) O(T^{-1})O(T^{-1})$$

$$= O(T^{-1}) \rightarrow 0 \text{ as } T \rightarrow \infty.$$

Now consider the last term in (7).

(a) If $d = 1$, $\phi = \frac{1}{2}$

$$\sqrt{T}\left(\frac{1}{\sqrt{T}}\right)^2 \sum_{s=1}^{S_T} \sum_{t=1}^{S_T} \frac{\partial}{\partial \theta_j} \log g\left(\frac{1}{\sqrt{T}}\lambda_s | \theta'\right) \frac{\partial}{\partial \theta_k} \log g\left(\frac{1}{\sqrt{T}}\lambda_t | \theta'\right) \left\{ \frac{H_2(\sqrt{T}\lambda_{s-t})}{H_2(0)} + O(T^{-1}) \right\}$$

$$\left\{ \frac{H_2(-\sqrt{T}\lambda_{t-s})}{H_2(0)} + O(T^{-1}) \right\}$$

$$= \frac{1}{\sqrt{T}} \sum \frac{\partial}{\partial \theta_j} \log g\left(\frac{1}{\sqrt{T}}\lambda_s | \theta'\right) \frac{\partial}{\partial \theta_k} \log g\left(\frac{1}{\sqrt{T}}\lambda_s | \theta'\right)$$

$$+ \sqrt{T}\left(\frac{1}{\sqrt{T}}\right)^2 \sum_{s \neq t} \sum \frac{\partial}{\partial \theta_j} \log g\left(\frac{1}{\sqrt{T}}\lambda_s | \theta'\right) \frac{\partial}{\partial \theta_k} \log g\left(\frac{1}{\sqrt{T}}\lambda_t | \theta'\right)$$

$$= \frac{\sin(\sqrt{T}\pi(s-t))}{\sqrt{T}(s-t)}^2 + O(T^{-1/2})$$

$$= \left(\frac{1}{\sqrt{T}}\right) \sum_{s=1}^{S_T} \frac{\partial}{\partial \theta_j} \log g\left(\frac{1}{\sqrt{T}} \lambda_s | \theta'\right) \frac{\partial}{\partial \theta_k} \log g\left(\frac{1}{\sqrt{T}} \lambda_s | \theta'\right) + O(T^{-1/2})$$

$$\rightarrow \int_0^{\beta} \frac{\partial}{\partial \theta_j} \log g(\lambda_x | \theta') \frac{\partial}{\partial \theta_k} \log g(\lambda_x | \theta') dx.$$

(b) If $d \geq 2$, $\phi = 1$

$$\begin{aligned} & T\left(\frac{1}{T}\right)^2 \sum_s \sum_t \frac{\partial}{\partial \theta_j} \log g\left(\frac{1}{T} \lambda_s | \theta'\right) \frac{\partial}{\partial \theta_k} \log g\left(\frac{1}{T} \lambda_t | \theta'\right) \left\{ \frac{H_2(\lambda_{s-t})}{H_2(0)} + O(T^{-1}) \right\} \\ & \cdot \left\{ \frac{H_2(\lambda_{t-s})}{H_2(0)} + O(T^{-1}) \right\} \end{aligned}$$

$$= \frac{1}{T} \sum_{s=1}^{S_T} \frac{\partial}{\partial \theta_j} \log g\left(\frac{1}{T} \lambda_s | \theta'\right) \frac{\partial}{\partial \theta_k} \log g\left(\frac{1}{T} \lambda_s | \theta'\right) + O(T^{-1})$$

$$\rightarrow \int_0^{\beta} \frac{\partial}{\partial \theta_j} \log g(\lambda_x | \theta) \frac{\partial}{\partial \theta_k} \log g(\lambda_x | \theta) dx \quad \text{as } T \rightarrow \infty.$$

(iv) Consider $\phi = 1$ and $\dim R^d = 1$. Recall (from Remark on the Proof of Theorem 3.3) that if the process is Brillinger 1-mixing (Definition 3.3),

$$\text{cum}\left\{T^{-1/2} \hat{d}_{a_1}^{(T)}\left(\frac{\lambda_{s_1}}{T}\right), \dots, T^{-1/2} \hat{d}_{a_n}^{(T)}\left(\frac{\lambda_{s_n}}{T}\right)\right\} \quad (8)$$

$$= (2\pi) f_{a_1, \dots, a_n}\left(\frac{\lambda_{s_1}}{T}, \dots, \frac{\lambda_{s_{n-1}}}{T} | \theta'\right) H_1(2\pi \lambda_{s_1} + \dots + \lambda_{s_n}) T^{(1-n/2)} + O(T^{(1-n/2)-1})$$

since $H_n(\lambda) = \int h(t/T)^n e^{-i\langle \lambda, t \rangle} dt = \int h(t/T) e^{-i\langle \lambda, t \rangle} dt = H_1(\lambda)$

for the taper under consideration. (3) is of the form

$$T^{\ell/2} T^{-\ell} \sum_{s_1} \dots \sum_{s_\ell} \prod_{j=1}^{\ell} \frac{\partial}{\partial \theta_{j_i}} \log g\left(\frac{\lambda_{s_i}}{T} | \theta'\right) \frac{1}{f\left(\frac{\lambda_{s_i}}{T} | \theta'\right)} \prod_{i=1}^k (H_n(u_n) + O(T^{-1}))$$

$$\prod_{i=1}^{i_0} (H_n(u_{n+k_i}) + O(T^{-1})) T^{(1-n/2)} \quad (9)$$

where by (8), u_n is a sum of $\pm s_i$'s. Recall $H_n(u_n) = 0$ if u_n is a non-zero integer. TABLE 1 (in (iii) above) is the special table considered in Appendix II. A typical term in (9) is

$$T^{(i_0 - (\ell - k))} T^{-\ell/2} \sum_{s_i} \dots \sum_{s_\ell} \prod_{j=1}^{\ell} \frac{\partial}{\partial \theta_{j_i}} \log g\left(\frac{\lambda_{s_i}}{T} | \theta'\right) \frac{1}{f\left(\frac{\lambda_{s_i}}{T} | \theta'\right)}$$

$$\cdot \left(\prod_{i=1}^{k_1} H_i \right) (O(T^{-(k-k_1)})) \cdot \left(\prod_{i=1}^{k_2} H_i \right) (O(T^{-(i_0-k_2)})) \quad (10)$$

where $0 \leq k_1 \leq k$, $0 \leq k_2 \leq i_0$.

By Appendix II

$(s_1, \dots, s_\ell) \mapsto (u_1, \dots, u_{k_1+k_2-1}, s_{j_{k_1+k_2}}, \dots, s_{j_\ell})$ is a 1-1

transformation. Since $H_i(u_i) = 0$ if u_i is a non-zero integer,

(10) can be bounded by

$$T^{(i_0 - (\ell - k))} T^{-\ell/2} T^{\ell - (k_1 + k_2) + 1} \cdot O(T^{-(k - k_1)}) O(T^{-(i_0 - k_2)})$$

$$= O(T^{1-\ell/2}) \rightarrow 0 \text{ as } T \rightarrow \infty \text{ if } \ell \geq 3. \quad //QED.$$

It is clear from the above proof that normal limit theorems may be obtained for many values of the parameter ϕ . One way of picking a "good" value of ϕ is to balance off a bigger normalizing rate for the second moment with a rate of approach to normality. For the quadratic functions considered, this amounts to looking at a second and third cumulant.

Definition: A separation ϕ_0 (or $T^{-\phi_0}$) is called optimal in the class of uniform separations $\{T^{-\phi} : \phi > 0\}$ if it maximizes the minimum of the second moment convergence factor $R(T)$, and the negative of the order of the normalized third cumulant, where $R^2(T) \times \text{variance}$ converges to a non-zero limit. That is

$$\sup_{\phi > 0} (\min[\text{order } R(T), -\text{order } K_3(T)]) \text{ occurs at } \phi_0.$$

Example: If $R(T) = T^{\ell/2}$ and $K_3(T) \sim KT^{(3\phi-4)/2}$ then

$$\sup_{\phi > 0} (\min[\frac{\phi}{2}, -\frac{1}{2}(3\phi-4)]) \text{ occurs at } \phi_0 = \frac{1}{2}.$$

Lemma 4.2.6

For the immigration process, the optimal separation when $\beta < \infty$ is $\phi_0 = 1$ (in general this holds for B-mixing processes which are also Brillinger 1-mixing).

Proof:

$$\begin{aligned} \text{It is easy to see } R(T) &\leq T^{\phi/2} \text{ if } 0 < \phi < 1 \\ &= T^{1/2} \quad \phi = 1 \\ &\leq T^{1/2} \quad \phi > 1 \end{aligned}$$

by the previous proof.

$K_3(T)$, a third cumulant of $R(T) \frac{\partial \Lambda^{(T)}}{\partial \theta}(\theta'; \beta, \phi)$, is $R(T)^3 O(T^{-d})$ or smaller if $\phi \neq 1$, and is $R(T)^3 O(T^{-d-1})$ if $\phi = 1$ (but not $o(T^{-d-1})$).

The optimal separation occurs at ϕ_0 a value of ϕ at which

$$\sup_{\phi > 0} \{ \min[\text{order } R(T), -\text{order } K_3(T)] \} \quad (1)$$

occurs.

$$(1) = \max \left\{ \sup_{0 < \phi < 1} \min\left(\frac{\phi}{2}, d - \frac{3\phi}{2}\right), \sup_{\phi > 1} \min\left(\frac{1}{2}, d - \frac{3}{2}\right), \min\left(\frac{1}{2}, d - \frac{1}{2}\right) \right\}$$

$$= \frac{1}{2} \text{ for } d = 1, 2, \dots$$

This occurs at $\phi_0 = 1$.

//QED.

Without Brillinger 1-mixing, the optimal ϕ_0 when $d = 1$ is $\phi_0 = \frac{1}{2}$.

$$\text{Let } \Lambda_{jk}^{(T)}(\theta; \beta, \phi) = \frac{\partial^2 \Lambda^{(T)}(\theta; \beta, \phi)}{\partial \theta_j \partial \theta_k}.$$

Lemma 4.2.7

For the immigration process

$$\begin{aligned} \text{(i)} \quad \Lambda_{jk}^{(T)}(\theta; \beta, \phi) &= \Pr \int_0^\beta \frac{\partial}{\partial \theta_j} \log g(\lambda_x | \theta) \frac{\partial}{\partial \theta_k} \log g(\lambda_x | \theta) \frac{g(\lambda_x | \theta')}{g(\lambda_x | \theta)} dx \\ &+ \int_0^\beta \frac{\partial^2}{\partial \theta_j \partial \theta_k} \log g(\lambda_x | \theta) \left(1 - \frac{g(\lambda_x | \theta')}{g(\lambda_x | \theta)} \right) dx \\ &= -\Lambda_{jk}(\theta; \beta), \quad \text{say,} \end{aligned}$$

(ii) $\forall \epsilon, \eta > 0 \exists \delta > 0$ and $T_0 > 0$ such that if $\bar{S}(\theta'; \delta) = \{\theta: \|\theta - \theta'\| \leq \delta\}$ then

$$P\left\{ \sup_{\theta \in \bar{S}(\theta', \delta)} |\Lambda_{jk}^{(T)}(\theta; \beta, \phi) - \Lambda_{jk}^{(T)}(\theta'; \beta, \phi)| \leq \epsilon \right\} \geq 1 - \eta \quad T \geq T_0.$$

Proof:

$$\text{(i)} \quad E\{\Lambda_{jk}^{(T)}(\theta; \beta, \phi)\} = \frac{1}{T^\phi} \sum_s \frac{\partial}{\partial \theta_j} \log g\left(\frac{1}{T^\phi} \lambda_s | \theta\right) \frac{\partial}{\partial \theta_k} \log g\left(\frac{1}{T^\phi} \lambda_s | \theta\right)$$

$$\frac{E\{\hat{I}^{(T)}(\frac{1}{T\phi}\lambda_s)\}}{g(\frac{1}{T\phi}\lambda_s|\theta)} + \frac{1}{T\phi} \sum_s \frac{\partial^2}{\partial\theta_j\partial\theta_k} \log g(\frac{1}{T\phi}\lambda_s|\theta) - 1 - \frac{E\{\hat{I}^{(T)}(\frac{1}{T\phi}\lambda_s)\}}{g(\frac{1}{T\phi}\lambda_s|\theta)}$$

$$+ - A_{jk}(\theta;\beta).$$

As in Proposition 4.2.3, it is shown that $\text{Var}\{\Lambda_{jk}^{(T)}(\theta;\beta;\theta)\} \rightarrow 0$,

so (i) follows by Chebyshev's Inequality.

(ii) The first and second partial derivatives of $\log g(\lambda_x|\theta)$ with respect to θ are Riemann integrable. As in the proof of Proposition 4.2.3 (iii), it can be shown that $\forall \epsilon > 0, \exists \delta_1, x_1 > 0$, if $\theta \in \bar{S}(\theta'; \delta)$, and if $x \geq x_1$, then

$$(a) \left| \frac{\partial^2}{\partial\theta_j\partial\theta_k} \log g(\lambda_x|\theta) \frac{1}{g(\lambda_x|\theta)} - \frac{\partial^2}{\partial\theta_j\partial\theta_k} \log g(\lambda_x|\theta') \frac{1}{g(\lambda_x|\theta')} \right|$$

$$\leq \epsilon \left| \frac{\partial^2}{\partial\theta_j\partial\theta_k} \log g(\lambda_x|\theta') \frac{1}{g(\lambda_x|\theta')} \right|$$

$$(b) \left| \frac{\partial}{\partial\theta_j} \log g(\lambda_x|\theta) \frac{\partial}{\partial\theta_k} \log g(\lambda_x|\theta) \frac{1}{g(\lambda_x|\theta)} - \frac{\partial}{\partial\theta_j} \log g(\lambda_x|\theta') \frac{\partial}{\partial\theta_k} \log g(\lambda_x|\theta') \frac{1}{g(\lambda_x|\theta')} \right|$$

$$\leq \epsilon \left| \frac{\partial}{\partial\theta_k} \log g(\lambda_x|\theta') \frac{1}{g(\lambda_x|\theta')} \right|$$

$$\leq \epsilon \left| \frac{\partial}{\partial\theta_j} \log g(\lambda_x|\theta') \frac{\partial}{\partial\theta_k} \log g(\lambda_x|\theta') \frac{1}{g(\lambda_x|\theta')} \right|$$

Proceeding as in (6) - (11) of the proof of Theorem 4.2.3 (iii),
the result follows. //QED.

Corollary to Lemma 4.2.7

If $\theta_T^* \xrightarrow{\text{Pr}} \theta'$ as $T \rightarrow \infty$, then (in either θ_1 or θ_2), $i = 1, 2$

$$\Lambda_{jk}^{(T)}(\theta_T^*; \beta, \phi) \xrightarrow{\text{Pr}} A_{jk}(\theta'; \beta) = \int \frac{\partial}{\partial \theta_j} \log g(\lambda_x | \theta') \frac{\partial}{\partial \theta_k} \log g(\lambda_x | \theta') dx$$

as $T \rightarrow \infty$.

Proof:

By Lemma 4.2.7 (ii), $\forall \epsilon, \eta$, choose $\delta(\epsilon)$ and $T_0(\epsilon, \eta) \rightarrow \infty$ if

$$N_{T, \delta, \epsilon} = \{\omega: \sup_{\theta \in \bar{S}(\theta; \delta)} |\Lambda_{jk}^{(T)}(\theta; \beta, \phi)(\omega) - \Lambda_{jk}^{(T)}(\theta'; \beta, \phi)(\omega)| > \epsilon\},$$

$$\text{then } P\{N_{T, \delta, \epsilon}\} < \frac{\eta}{2} \quad \forall T \geq T_0.$$

$$\text{Let } M_{T, \delta} = \{\omega: |\theta_T^*(\omega) - \theta'| > \delta\}. \text{ Choose } T_1 \geq T_0.$$

$$P\{M_{T, \delta}\} \leq \frac{\eta}{2} \quad \forall T \geq T_1.$$

If $\omega \in \Omega \setminus (M_{T, \delta} \cup N_{T, \delta, \epsilon})$, then

$$|\Lambda_{jk}^{(T)}(\theta_T^*; \beta, \phi)(\omega) - \Lambda_{jk}^{(T)}(\theta'; \beta, \phi)(\omega)| \leq \epsilon.$$

$$P\{|\Lambda_{jk}^{(T)}(\theta_T^*; \beta, \phi) - \Lambda_{jk}^{(T)}(\theta'; \beta, \phi)| > \epsilon\} \leq P\{N_{T, \delta, \epsilon}\} + P\{M_{T, \delta}\} < \eta.$$

The Corollary now follows from Lemma 4.2.7 (i).

//QED.

Theorem 4.2.8

In either parameter space θ_1 or θ_2 ,

$$(i) T^{\phi/2}(\hat{\theta}_T - \theta') \xrightarrow{D} N(0, A^{-1}(\theta'; \beta))$$

$$\text{where } \phi = \begin{cases} \frac{1}{2} & \text{if } \dim R^d = 1 \\ 1 & \text{if } \dim R^d \geq 2 \end{cases},$$

(ii) If $\dim R^d = 1$

$$T^{\frac{1}{2}}(\hat{\theta}_T - \theta') \xrightarrow{D} N(0, A^{-1}(\theta'; \beta) V(\theta'; \beta) A^{-1}(\theta'; \beta)),$$

$$\text{where } V(\theta'; \beta) = A(\theta'; \beta) + (2\pi) \frac{H_4(0)}{H_2(0)^2} \int_0^\beta \int_0^\beta \left\{ \frac{\partial}{\partial \theta_j} \log g(\lambda_x | \theta') \frac{\partial}{\partial \theta_k} \log g(\lambda_y | \theta') \right.$$

$$\left. \frac{f^{(4)}(\lambda_x, -\lambda_x, \lambda_y | \theta')}{f(\lambda_x | \theta') f(\lambda_y | \theta')} \right\} dx dy,$$

where $f^{(4)}$ is a fourth order cumulant spectral density function.

Proof: By Taylor's Theorem with remainder

$$T^{\phi/2} \frac{\partial \Lambda^{(T)}}{\partial \theta}(\theta'; \beta, \phi) = T^{\phi/2} \frac{\partial \Lambda^{(T)}}{\partial \theta}(\hat{\theta}_T; \beta, \phi) + \frac{\partial^2 \Lambda^{(T)}}{\partial \theta_j \partial \theta_k}(\theta_T^*; \beta, \phi) T^{\phi/2}(\theta' - \hat{\theta}_T)$$

where $|\theta_T^* - \hat{\theta}_T| \leq |\hat{\theta}_T - \theta'|$.

Therefore $|\theta_T^* - \theta'| \leq |\theta_T^* - \hat{\theta}_T| + |\hat{\theta}_T - \theta'| \leq 2|\hat{\theta}_T - \theta'|$

$$\therefore \theta_T^* \xrightarrow{Pr} \theta'$$

$$T^{\phi/2} \frac{\partial \Lambda^{(T)}}{\partial \theta}(\hat{\theta}_T; \beta, \phi) \xrightarrow{Pr} 0 \text{ by Theorem 4.2.4.}$$

It can be shown, $A(\theta'; \beta)$ is positive definite. Therefore

$$A^{-1}(\theta'; \beta) T^{\phi/2} \frac{\partial \Lambda^{(T)}}{\partial \theta}(\theta'; \beta, \phi) \stackrel{D}{\approx} A^{-1}(\theta'; \beta) \frac{\partial^2 \Lambda^{(T)}}{\partial \theta_j \partial \theta_k}(\theta'; \beta, \phi) T^{\phi/2}(\theta' - \hat{\theta}_T)$$

$$\stackrel{D}{\approx} T^{\phi/2}(\hat{\theta}_T - \theta')$$

where $\stackrel{D}{\approx}$ means asymptotically equal in distribution. The Theorem now follows from Theorem 4.2.5 (iii) and (iv). //QED.

Defining Fisher information as $I_T = \text{Cov}\left\{\frac{\partial \Lambda^{(T)}}{\partial \theta}(\theta'; \beta, \phi)\right\}$, that is the expected derivative of the log likelihood (cf Kendall and Stuart, Vol. 2, p. 28), from Theorem 4.2.5 it follows

$$I_T \sim T^{-\phi/2} A(\theta'; \beta).$$

We have the following asymptotic efficiency result.

Theorem 4.2.9

(i) In the sense of Fisher information, Theorem 4.2.8 (i) is asymptotically fully efficient.

(ii) In the sense of optimal separation, Theorem 4.2.8 (i) for $\dim R^d \geq 2$ and Theorem 4.2.8 (ii) for $\dim R^d = 1$, are optimal.

Proof: (i) The Cramer-Rao Lower bound is attained.

(ii) This is just Lemma 4.2.6.

//QED.

To estimate r , the immigration rate, one could use $\hat{r}_T = \frac{\hat{p}_1}{M_{(\tau, \hat{m})}}$
 in Θ_1 and $\hat{r}_T = \frac{\hat{p}_1}{M_{(\hat{\tau}, m)}}$ in Θ_2 .

The results given above use a "normalized" periodogram $\hat{I}^{(T)}(\lambda)/\hat{p}_1$. This is necessary when $\beta = \infty$. If however, $\beta < \infty$, one does not run into the same integrability problem (cf Theorem 4.2.1), so that one may use the periodogram without normalizing. In this case it is clear that all the same theorems hold, and that the parameter of immigration r , may be included in the estimation procedure. In this case

$$\frac{\partial}{\partial r} \log f(\lambda|\theta) = \frac{1}{r}$$

so that one needs $\beta < \infty$ (cf the covariance matrix in Theorem 4.2.5).

Below let $\theta = \{(m, V, \rho, r)\}$ for τ given. Let $(m, V, \rho, r) = (\theta_1, \theta_2, \theta_3, \theta_4)$,

$$\Lambda^{(T)}(\theta; \beta, \phi) = - \frac{1}{T\phi} \sum_{s=1}^{\beta T\phi} \left\{ \log f\left(\frac{\lambda_s}{T\phi} | \theta\right) + \frac{\hat{I}^{(T)}\left(\frac{\lambda_s}{T\phi}\right)}{f\left(\frac{\lambda_s}{T\phi} | \theta\right)} \right\}.$$

Theorem 4.2.10

Let the taper be $h = \chi_A$, $A = \prod_1^d [-b_j, b_j]$, $b_1 = \pi$.

(i) Then $\exists \hat{\theta}_T$, a local maximum of $\Lambda^{(T)}(\theta; \beta, \phi)$ such that

$\hat{\theta}_T \xrightarrow{\text{Pr}} \theta'$ as $T \rightarrow \infty$ and $\hat{\theta}_T \xrightarrow{\text{a.s.}} \theta'$ as $T \rightarrow \infty$ through the rationals where θ' is the true value of θ .

(ii) Let $\phi = \begin{cases} \frac{1}{2} & \text{if } d = 1 \\ 1 & \text{if } d \geq 2 \end{cases}$,

$$A(\theta'; \beta) = \int_0^\beta \frac{\partial}{\partial \theta_j} \log f(\lambda_x | \theta') \frac{\partial}{\partial \theta_k} \log f(\lambda_x | \theta') dx.$$

Then

$$T^{\phi/2}(\hat{\theta}_T - \theta') \xrightarrow{D} N(0, A^{-1}(\theta'; \beta)) \quad \text{as } T \rightarrow \infty.$$

(iii) If $d = 1$, and $\phi = 1$,

$$T^{1/2}(\hat{\theta}_T - \theta') \xrightarrow{D} N(0, A^{-1}(\theta'; \beta) + B(\theta'; \beta)) \quad \text{as } T \rightarrow \infty,$$

where $A(\theta'; \beta)B(\theta'; \beta)A(\theta'; \beta)$

$$= \frac{(2\pi)H_4(0)}{H_2(0)^2} \int_0^\beta \int_0^\beta \frac{\partial}{\partial \theta_j} \log f(\lambda_x | \theta') \frac{\partial}{\partial \theta_k} \log f(\lambda_y | \theta')$$

$$\frac{f^{(4)}(\lambda_x, -\lambda_x, \lambda_y | \theta')}{f(\lambda_x | \theta')f(\lambda_y | \theta')} dx dy.$$

Also the asymptotic Fisher efficiency and optimal separation properties given in Theorem 4.2.9 continue to hold.

§4.3 Asymptotic Likelihood Estimation for an L-Dimensional B-Mixing Process

In this section we briefly consider a multivariate extension of the methods of the previous section. By Theorem 3.4, one can obtain asymptotically independent $W_L^C(\ell, f_{XX}(\lambda))$, $\lambda \neq 0$, random variables. In order to have limiting non-singular distributions, one needs $\ell \geq L$. Let $\lambda_1(T), \dots, \lambda_L(T)$ be such that if $j \neq k$, then

$T|\lambda_j(T) - \lambda_k(T)| \rightarrow \infty$, and $\lambda_j(T) \rightarrow 0$ as $T \rightarrow \infty$. Then

$$\hat{I}_{XX}^{(T)}(\lambda|L) = (2\pi)^{-d_{H_2}(0)} \sum \hat{d}^{(T)}(\lambda + \lambda_j(T)) \hat{d}^{(T)*}(\lambda + \lambda_j(T)), \quad (1)$$

is asymptotically $W_L^C(L, f_{XX}(\lambda))$ and these are asymptotically independent as in the Corollary to Theorem 3.4.

Suppose we now proceed as in the previous section with

$$\Lambda^{(T)}(\theta; \beta, \phi) = - \frac{1}{T\phi} \sum_{s=1}^{S_T} \{L \log |f_{XX}(\frac{\lambda_s}{T\phi} | \theta)| - \text{tr } f_{XX}(\frac{\lambda_s}{T\phi} | \theta)^{-1} I^{(T)}(\frac{\lambda_s}{T\phi} | L)\}, \quad (2)$$

(using the complex Wishart density (1.6) in Goodman (1963)).

Again consider the special taper $h = \chi_A$, $A = \Pi[-b_j, b_j]$ with $b_1 = \pi$.

Recall if $j \neq k$,

$$T|\lambda_j(T) - \lambda_k(T)| \rightarrow \infty \quad (3)$$

Thus none of $\lambda_j(T)$ is $o(T^{-1})$, if $L \geq 2$.

In the previous section, it is seen that $\phi = 1$ is optimal as well as the largest ϕ so that we obtain normal limit theorems with covariance involving only integrals of spectral densities. In order to obtain that, only frequencies of the form $\frac{\lambda_s}{T}$, $s \in \mathbb{Z}$ are considered in order to take advantage of the fact that $H_2(T \cdot \frac{\lambda_s}{T}) = 0$ for $s \neq 0$. Now consider (2) with $\phi = 1$. It is seen that this involves frequencies of the form $\frac{\lambda_s}{T} + \lambda_j(T)$. It is necessary that $\lambda_j \rightarrow 0$, but also that $T|\lambda_j(T)|$ cannot remain bounded if $L \geq 2$. In particular λ_j is not of the form $\frac{\lambda_t}{T}$, $t \in \mathbb{Z}$. Then we cannot take the same advantage of the structure of H_2 . If $\phi < 1$, we can take $\lambda_j(T) = \frac{c(j)}{T^\phi}$ for some appropriate constants $c(j)$. In this way, at least for $\phi = \frac{1}{2}$, we can obtain a normal limit theorem with rate $R(T) = T^{\phi/2}$. These results are clearly not as good as in the case $L = 1$.

Suppose now we abandon our view of "asymptotic likelihood". Apart from Fisher efficiency, the main advantage of this approach is that it gives rise to an estimating equation. Instead of using (1) as a sum of L periodograms, just consider, as before

$$\hat{I}^{(T)}(\lambda) = (2\pi)^{-d} H_2(0)^{-1} \hat{d}^{(T)}(\lambda) * \hat{d}^{(T)}(\lambda). \quad (4)$$

It has a singular limiting distribution ($\hat{I}$ is of rank 1). Despite this, consider the following which is motivated by (2):

$$\Lambda^{(T)}(\theta; \beta, \phi) = - \frac{1}{T^\phi} \sum_{s=1}^{\beta T} \left\{ \log |f_{XX}(\frac{\lambda_s}{T^\phi} | \theta)| - \text{tr} f_{XX}(\frac{\lambda_s}{T^\phi} | \theta)^{-1} \hat{I}^{(T)}(\frac{\lambda_s}{T^\phi}) \right\}. \quad (5)$$

Now if $\phi = 1$, we no longer have the technical problem of (3).

In fact, using (5), we can obtain results similar to those of the previous section (except the Fisher information result of course).

$$\text{Let } \Lambda(\theta; \beta) = - \int_0^\beta \{ \log f_{XX}(\lambda_x | \theta) - \text{tr } f_{XX}^{-1}(\lambda_x | \theta) f_{XX}(\lambda_x | \theta') \} dx .$$

$\Lambda(\theta; \beta)$ has a unique maximum at θ' . It can be shown $\Lambda^{(T)}(\theta; \beta) \xrightarrow{\text{Pr}} \Lambda(\theta; \beta)$ and also locally uniform convergence analogous to that in Theorem 4.2.3. It then follows that there exists a local maximum of $\Lambda^{(T)}(\theta; \beta, \phi)$ such that $\hat{\theta}_T \xrightarrow{\text{Pr}} \theta'$ where θ' is the true value of the parameter. Under the conditions of Theorem 4.2.3 (iv), it also follows that $\hat{\theta}_T \rightarrow \theta'$ a.s. as $T \rightarrow \infty$ through the rationals. Then one can write down the following normal limit Theorem.

Theorem 4.3.1

Suppose $h = \chi_A$, $A = \Pi[-b_j, b_j]$, $b_1 = \pi$,

$$(i) \ E \left\{ \frac{\partial \Lambda^{(T)}}{\partial \theta} (\theta'; \beta, \phi) \right\} = O(T^{-1}) .$$

$$(ii) \text{ Let } \phi = \begin{cases} \frac{1}{2} & , \dim R^d = 1 \\ 1 & , \dim R^d \geq 1 \end{cases} ,$$

$$-A_{jk}(\theta'; \beta) = \text{pr. lim}_{T \rightarrow \infty} \frac{\partial^2 \Lambda(T)}{\partial \theta_j \partial \theta_k}(\theta'; \beta, \phi) = \int_0^\beta \text{tr} \left(\frac{\partial f}{\partial \theta_k}(\lambda_x | \theta') \frac{\partial f^{-1}}{\partial \theta_j}(\lambda_x | \theta') \right) dx.$$

Then $A(\theta'; \beta) T^{\phi/2} (\hat{\theta}_T - \theta') \xrightarrow{D} N(0, A(\theta'; \beta))$ as $T \rightarrow \infty$.

Proof: (Sketch)

If A is an invertible matrix then (Birkhoff and MacLane, p. 284)

$$\frac{\partial}{\partial x} |A| = |A| \text{tr} \left(A^{-1} \frac{\partial A}{\partial x} \right).$$

Therefore $E \left\{ \frac{\partial \Lambda(T)}{\partial \theta_j}(\theta'; \beta, \phi) \right\}$

$$= - \frac{1}{T^\phi} \sum \left\{ \frac{|f(\frac{\lambda_s}{T^\phi} | \theta')|}{|f(\frac{\lambda_s}{T^\phi} | \theta')|} \text{tr} \left(f^{-1} \left(\frac{\lambda_s}{T^\phi} | \theta' \right) \frac{\partial f}{\partial \theta_j} \left(\frac{\lambda_s}{T^\phi} | \theta' \right) \right. \right. \\ \left. \left. + \text{tr} \left(\frac{\partial f^{-1}}{\partial \theta_j} \left(\frac{\lambda_s}{T^\phi} | \theta' \right) E \hat{I}^{(T)} \left(\frac{\lambda_s}{T^\phi} \right) \right) \right\}$$

$$= - \frac{1}{T^\phi} \sum \left\{ \text{tr} \left(f^{-1} \frac{\partial f}{\partial \theta_j} + \frac{\partial f^{-1}}{\partial \theta_j} f \right) + O(T^{-1}) T^\phi O(T^{-\phi}) \right\}$$

$$= O(T^{-1}),$$

$$\text{since } f^{-1} \frac{\partial f}{\partial \theta_j} + \frac{\partial f^{-1}}{\partial \theta_j} f = \frac{\partial}{\partial \theta_j} I = 0.$$

Part (ii) is proven in an analogous manner to the proof in §4.2. //QED.

If we do use the asymptotic likelihood via (1) and (2), then Theorem 4.3.1 holds with $\phi = \frac{1}{2}$ for all $\dim R^d$. In this case A has the Fisher information interpretation and the estimates are asymptotically fully efficient. However, one has to give up a faster rate of convergence by using (2) instead of (5). Using (5) even allows one to obtain a normal limit Theorem with $\phi = 1$ in case $\dim R^d = 1$. This involves fourth order cumulant spectra as in Theorem 4.2.5.

§4.4 Asymptotic Minimum Squares Estimation

The asymptotic likelihood procedure requires that the second order cumulant spectra $f_{\chi\chi}(\lambda|\theta)$ be invertible for all $\lambda \neq 0, \forall \theta$. Actually this is only necessary for $0 < |\lambda| < \beta$ where β can be determined independently of θ . Asymptotic likelihood gives an intuitively appealing estimating equation. In this section we consider just starting with a particular estimating equation without a likelihood motivation. Brillinger's Lemma, especially (iv), describes what an estimating equation should be; that is, it should identify the parameter value θ' of interest in a unique way.

Consider an L-time slice version of the immigration process and the taper $h = \chi_A$, $A = \Pi[-b_j, b_j]$, $b_1 = \pi$. Let $0 < \beta < \infty$, $\phi > 0$ and

$$\Gamma^{(T)}(\theta; \beta, \phi) = \frac{1}{T^\phi} \sum_{s=1}^{\beta T^\phi} \left| F\left(\frac{\lambda_s}{T^\phi} | \theta\right) - \hat{I}^{(T)}\left(\frac{\lambda_s}{T^\phi}\right) \right|^2, \quad (1)$$

where F is the $L \times L$ second order spectral density function,

$$\hat{I}^{(T)}(\lambda) = (2\pi)^{-d} H_2(0)^{-1} \hat{d}^{(T)}(\lambda) \star \hat{d}^{(T)}(\lambda),$$

and for an $L \times L$ matrix, $A = (a_{ij})$

$$\|A\|^2 = \sum_{i,j} |a_{ij}|^2.$$

We obtain our estimate as a local minimum of (1), and so we refer to this procedure as asymptotic minimum squares.

We apply Brillinger's Lemma. To be specific, we consider an L -time slice of the immigration process with $L = 1$ and $L = 2$.

The parameter spaces we use are

$$\Theta_2 = \{ (m, \rho, V, r, \tau) : -V < m < V, V, \rho, r, \tau \in \mathbb{R}^+ \},$$

when $L = 1$ and $L = 2$ respectively. Let θ' be the true value of the parameter, the appropriate parameter space determined by context.

When $L = 2$, the time slices are at τ and $\tau + t$, where $t > 0$ is an experimental parameter. It determines the time scale.

The second order spectral density matrices are given by

$$(i) \quad L = 1 \quad F(\lambda|\theta) = f_{\tau, \tau}(\lambda|\theta),$$

$$(ii) \quad L = 2 \quad F(\lambda|\theta) = \begin{bmatrix} f_{\tau, \tau}(\lambda|\theta) & f_{\tau+t, \tau}(\lambda|\theta) \\ f_{\tau, \tau+t}(\lambda|\theta) & f_{\tau+t, \tau+t}(\lambda|\theta) \end{bmatrix},$$

where (cf §2)

$$f_{\tau, \tau}(\lambda|\theta) = \begin{cases} (2\pi)^{-d} [rM_{(\tau, m)} + 2V\alpha r \int_0^\tau se^{-ms} ds] & \text{if } m + \rho|\lambda|^2 = 0 \\ (2\pi)^{-d} [rM_{(\tau, m)} + 2V\alpha r \frac{(M_{(\tau, m)} - M_{(\tau, 2m+\rho|\lambda|^2)})}{m + \rho|\lambda|^2}] & \text{if } m + \rho|\lambda|^2 \neq 0 \end{cases}$$

$$\text{and } f_{\tau, \tau+t}(\lambda|\theta) = e^{-\frac{1}{2}(2m + \rho|\lambda|^2)t} f_{\tau, \tau}(\lambda|\theta).$$

$$\text{Let } \Gamma(\theta; \beta) = \int_0^\beta ||F(\lambda_x|\theta) - F(\lambda_x|\theta')||^2 dx. \quad (2)$$

It can be shown that for the immigration process, when $L = 1$ or 2 Γ has a unique minimum at θ' . It is clear that conditions (i) - (iv) of Brillinger's Lemma holds. Proofs are similar to §4.2, and so are omitted.

Proposition 4.4.1

For the immigration process, for $0 < \beta < \infty$, and $L = 1$ or 2

$$(i) \Gamma^{(T)}(\theta; \beta, \phi) \xrightarrow{\text{Pr}} \Gamma^{(T)}(\theta; \beta, \phi) \text{ as } T \rightarrow \infty,$$

$$(ii) \forall \theta_1, \forall \epsilon > 0, \exists \delta(\epsilon, \theta_1) \ni \forall \eta > 0, \exists T_0(\epsilon, \eta, \theta_1),$$

$$P\left\{ \sup_{\bar{S}(\theta_1, \delta)} |\Gamma^{(T)}(\theta; \beta, \phi) - \Gamma^{(T)}(\theta_1; \beta, \phi)| > \epsilon \right\} \leq \eta \quad \forall T \geq T_0,$$

(iii) For the special taper χ_A , $A = \Pi[-b_j, b_j]$, if $\phi = \frac{1}{2}$ or 1 , $\forall$ compact $K \subset \Theta_1$ (or Θ_2)

$$\sup_K |\Gamma^{(T)}(\theta; \beta, \phi) - \Gamma(\theta; \beta)| \rightarrow 0 \text{ a.s.}$$

as $T \rightarrow \infty$ through the rationals.

Proof: Similar to Proposition 4.2.3

Proposition 4.4.2

Let $\theta' \in U_1 \subset U_2$, U_1 open, U_2 compact. Consider the L-time slice immigration process $L = 1$; or 2 . Then $\hat{\theta}_T \in U_2$.
 $\Gamma^{(T)}(\hat{\theta}_T; \beta, \phi) \leq \Gamma^{(T)}(\theta; \beta, \phi)$ for all $\theta \in U_2$ and such that $\hat{\theta}_T \in U_1$ with probability tending to 1 and $T \rightarrow \infty$. Also $\hat{\theta}_T \in U_1$ almost surely as $T \rightarrow \infty$ through the rationals, if $\phi = 1$ or $\frac{1}{2}$.

Proof: Essentially the same as that of Theorem 4.2.4. //QED.

Corollary to Proposition 4.4.2

For an L-time slice immigration process, $L = 1$ or 2 , since $\Gamma^{(T)}$ is differentiable with respect to θ (in either parameter space), there exists a root $\hat{\theta}_T$ of $\frac{\partial \Gamma^{(T)}}{\partial \theta}(\theta; \beta, \phi)$ such that $\hat{\theta}_T \rightarrow \theta'$ as $T \rightarrow \infty$. Also $\hat{\theta}_T \rightarrow \theta'$ a.s. as $T \rightarrow \infty$ through the rationals provided $\phi = \frac{1}{2}$ or 1 .

Theorem 4.4.3

$$(i) E\left\{\frac{\partial \Gamma^{(T)}}{\partial \theta}(\theta'; \beta, \phi)\right\} = O(T^{-1})$$

$$(ii) \text{ If } \phi = \begin{cases} \frac{1}{2} & \text{if } \dim R^d = 1 \\ 1 & \text{if } \dim R^d \geq 2 \end{cases}, \text{ then}$$

$$T^{\phi/2} \left(\frac{\partial \Gamma^{(T)}}{\partial \theta}(\theta'; \beta, \phi) - E\left(\frac{\partial \Gamma^{(T)}}{\partial \theta}(\theta'; \beta, \phi)\right) \right) \xrightarrow{D} N(0, 4V) \text{ as } T \rightarrow \infty,$$

$$\text{where } v_{jk}(\theta'; \beta) = \int_0^\beta \text{tr} \left(\left[\frac{\partial F}{\partial \theta_j}(\lambda_x | \theta') \right] F(\lambda_x | \theta') \left[\frac{\partial F}{\partial \theta_k}(\lambda_x | \theta') \right] F(\lambda_x | \theta') \right) dx .$$

(iii) If $\phi = 1$ and $\dim R^d = 1$,

$$T^{\frac{1}{2}} \left(\frac{\partial \Gamma(T)}{\partial \theta}(\theta'; \beta, \phi) - E \frac{\partial \Gamma(T)}{\partial \theta}(\theta'; \beta, \phi) \right)$$

$$\stackrel{D}{\rightarrow} N(0, 4V + \frac{4H_4(0)}{2\pi H_2(0)^2} \int_0^\beta \int_0^\beta \left(\sum_a \sum_b \frac{\partial f_{ab}}{\partial \theta_j}(\lambda_x | \theta') \frac{\partial f_{ab}}{\partial \theta_k}(\lambda_y | \theta') \right.$$

$$\left. \cdot f_{abab}(\lambda_x, -\lambda_x, \lambda_y | \theta') dx dy \right) .$$

Proof: (i) Straightforward.

(ii) and (iii): $\frac{\partial \Gamma(T)}{\partial \theta}(\theta'; \beta, \phi)$ is a quadratic function of Fourier

transforms which is essentially the same as that considered in

Theorem 4.2.5. Thus the normality follows in exactly the same way

as in Theorem 4.2.5. The second moment limits are easily verified. //QED.

Theorem 4.4.4

$$T^{\phi/2}(\hat{\theta}_T - \theta') \stackrel{D}{\rightarrow} N(0, A^{-1}VA^{-1}) , \quad \text{if } \phi = 1, \dim R^d \geq 2$$

$$\stackrel{D}{\rightarrow} N(0, A^{-1}VA^{-1}) , \quad \text{if } \phi = \frac{1}{2}, \dim R^d = 1$$

$$\stackrel{D}{\rightarrow} N(0, A^{-1}(V+V_1)A^{-1}), \text{ if } \phi = 1, \dim R^d = 1 ,$$

where

$$\lim_{T \rightarrow \infty} \text{pr} \frac{\partial^2 \Gamma(T)}{\partial \theta_j \partial \theta_k} (\theta'; \beta, \phi) = 2^{\int_0^{\beta}} \left(\text{tr} \left(\frac{\partial F}{\partial \theta_j} (\lambda_x | \theta') \frac{\partial F}{\partial \theta_k} (\lambda_x | \theta) \right) \right) dx$$

$$= 2A(\theta'; \beta).$$

Proof: Similar to Theorem 4.2.8.

//QED.

§5 Concluding Remarks

The method of approximate likelihood is interesting since the periodograms are not asymptotically identically distributed. A simple example of independent but not identically distributed random variables with common parameter $\theta \in \mathbb{R}$ such that the mle is not consistent is given. It is a version of the well known dilution series experiment. Let $X_n \sim \text{Poisson, mean } \theta a^{-n}$, $a > 1$ known.

The mle is $\hat{\theta}_n = \left(\sum_{i=1}^n X_i \right) \frac{(a-1)}{1-a^{-n}}$. Then $\frac{\hat{\theta}_n}{a-1} \xrightarrow{D} \text{Poisson, mean } \frac{\theta}{a-1}$

as $n \rightarrow \infty$. The problem here of course is bounded Fisher information.

In the class of tapers considered, Theorem 4.2.10, the asymptotic normality of the estimates, is best in that a faster rate of convergence cannot be obtained.

For a vector valued point process, the "approximate likelihood" method is not useful in general, since positive definiteness of the spectral density function is needed. The "asymptotic least squares" method is introduced to overcome this difficulty.

The methods of estimation used are for isotropic processes. In principle they can be extended to non-isotropic processes since a line or set in the frequency space which distinguishes parameter values is needed.

The results involve an experimental parameter β . When $\beta < \infty$, no frequency λ such that $|\lambda| > \beta$ is used. If higher frequencies are negligible, then it should be possible to extend the estimation

procedure to data recorded as population totals in small squares or blocks. Such a result would be similar to Shannon's Sampling Theorem (Koopmans (1974)) for real time series.

If the B-mixing assumption is not met, but only holds for the first $k \geq 4$ cumulants, then the normal limit theorems of chapters 3 and 4 may no longer hold. However the first k moments and first $\lfloor \frac{k}{2} \rfloor$ moments of the limiting random variables in chapters 3 and 4 respectively would still coincide with the normal moments. This is the same type of approximate normal limit one has when estimating spectral densities via weighted periodograms (Theorem 3.5) under similar conditions. This type of approximation is often used in real time series practice.

The method used to prove B-mixing of the simple branching diffusion process should easily extend to Poisson branching or n -ary branching (as opposed to binary branching). Other branching mechanisms should also be investigated.

A special case of the estimation problem considered in chapter is $m = 0$. In this case, better results can be obtained. It can be shown that if $f(\lambda|\theta_1) = f(\lambda|\theta_2)$ for $\lambda = 0$ and $|\lambda|^2 = 1, 2, 4$ and 8 , then $\theta_1 = \theta_2$. Consider

$$L^{(T)}(\theta) = \sum_{\sigma}^4 |f(\lambda_i|\theta_i) - \hat{f}^{(T)}(\lambda_i)|^2,$$

where $\lambda_0 = 0$, $|\lambda_1|^2 = 1$, $|\lambda_2|^2 = 2$, $|\lambda_3|^2 = 4$, $|\lambda_4|^2 = 8$, and where $\hat{f}^{(T)}$ is given in Theorem 3.6. If $\hat{\theta}_T$ is the least squares estimate from $L^{(T)}$, it will follow that $(B_T^{-1}(B_T T)^d)(\hat{\theta}_T - \theta')$ is asymptotically

normal. The rate of convergence to normality, $B_T^{-1/2}(B_T T)^{d/2}$ is much faster than the rate $T^{1/2}$ obtained in section 4.

The last comment to be made concerns the problem of goodness of fit. Csörgő, Sheshadri and Yalovsky (1975) obtain a characterization of an exponential distribution and proceed to give a test of fit. Such characterizations seem to be ideally what one would want in a test of fit. Short of this one would like to derive a necessary (non-parametric) consequence and test this. This process is similar to what some philosophers call verification or truth or falsehood, which is discussed in Carnap (1935). We refer to such tests of fit as a partial test of fit. A test of fit should not be too sensitive, since a stochastic model is usually an approximation of reality and so should also model "nearby systems". For the simple branching diffusion process, we propose a statistic of a partial test of fit.

Consider a two time slice version, times $\tau_1 = \tau$ and $\tau_2 = \tau + t$. Let $\lambda \neq 0$, $h = X_{\Pi[-b_j, b_j]}$. Then by Theorem 3.3,

$$T^{d/2} \operatorname{Re} \left(\frac{\hat{d}^{(T)}(\lambda)}{T^d} \right) \xrightarrow{D} N(0, (2\pi)^d H_2(0) \begin{bmatrix} f_{\tau_1, \tau_1}(\lambda|\theta) & f_{\tau_2, \tau_1}(\lambda|\theta) \\ f_{\tau_1, \tau_2}(\lambda|\theta) & f_{\tau_2, \tau_2}(\lambda|\theta) \end{bmatrix})$$

Thus

$$T^{-d/2} (\operatorname{Re} \frac{\hat{d}^{(T)}(\lambda)}{T^d} + \operatorname{Re} \frac{\hat{d}^{(T)}(\lambda)}{T^d}) \xrightarrow{D} N(0, (2\pi)^d H_2(0) (f_{\tau_1, \tau_1}(\lambda|\theta) + 2f_{\tau_1, \tau_2}(\lambda|\theta) + f_{\tau_2, \tau_2}(\lambda|\theta))) \quad (1)$$

The variance term is positive. Then by Chung (1974, Theorem 4.4.6)

$$Z_T(\lambda) = \frac{T^{-d/2} (\operatorname{Re} \hat{d}_{\tau_1}^{(T)}(\lambda) + \operatorname{Re} \hat{d}^{(T)}(\lambda))}{(2\pi)^d H_2(0)^{1/2} [f_{\tau_1, \tau_1}(\lambda | \hat{\theta}_T) + 2f_{\tau_1, \tau_2}(\lambda | \hat{\theta}_T) + f_{\tau_2, \tau_2}(\lambda | \hat{\theta}_T)]^{1/2}}$$

$$\begin{matrix} D \\ \rightarrow N(0,1) \end{matrix} \text{ as } T \rightarrow \infty.$$

where $\hat{\theta}_T$ is the estimate constructed in chapter 4. $Z_T(\lambda)$ is a statistic. If the process is B-mixing, but the covariance in (1) is not that of the simple branching diffusion process, then $Z_T(\lambda)$ is still asymptotically normal, but its variance is not 1. Thus $Z_T(\lambda)$ tests the question, "Do we have the proper covariance structure for the bivariate point process to be a simple branching diffusion process?"

By Theorem 3.3, one could use several frequencies, and thus use $Z_T(\lambda_1), \dots, Z_T(\lambda_k)$ to construct a partial test of fit.

References

- [1] Billingsley, P. (1968) Convergence of Probability Measures, John Wiley and Sons.
- [2] Brillinger, D.R. (1970) The frequency analysis of relations between spatial series. Proceedings of the Twelfth Biennial Seminar of the Canadian Mathematical Congress, Montreal.
- [3] Brillinger, D.R. (1974) Statistical inference for stationary point processes, in Stochastic Processes and Related Topics, edited by M.L. Puri, Academic Press, New York.
- [4] Brillinger, D.R. (1975) Time Series: Data Analysis and Theory, Holt, Rinehart and Winston, Inc., Toronto.
- [5] Carnap, R. (1935) Philosophy and Logical Syntax, first chapter reprinted in The Age of Analysis, edited by Morton White, Mentor, 1955.
- [6] Chung, K.L. (1974) A First Course in Probability Theory, Academic Press, New York.
- [7] Csörgö, M., V. Sheshadri and M. Yalovsky (1975) Applications of characterizations in the area of goodness of fit, in Statistical Distributions in Scientific Work, edited by G. Patil, S. Kotz and J.K. Ord.
- [8] Dawson, D. and G. Ivanoff (1977) Branching Diffusions and Random Measures, to appear in Advances in Probability, edited by P. Ney
- [9] Gel'fand, I.M. and N. Ya Vilenkin (1964) Generalized Functions, Academic Press, New York.

- [10] Goodman, N.R. (1963) Statistical analysis based upon a certain multivariate complex Gaussian distribution (an introduction), *Ann. Math. Stat.*, 34, 152-177.
- [11] Ivanoff, G. (1976) Branching Diffusion Processes, Ph.D. Thesis, Department of Mathematics, Carleton University.
- [12] Jagers, P. (1973) Aspects of random measures and point processes, *Z. Wahr. Verw. Geb.* 179-238.
- [13] Jordon, C. (1960) Calculus of Finite Differences, Chelsea Publishing Co., New York.
- [14] Keiding, N. (1975) Maximum likelihood estimation in the birth and death process, *Ann. Stat.*, 3, 363-372, Correction, *Ann. Stat.*, 6, 472.
- [15] Kendall, M. and A. Stuart. The Advanced Theory of Statistics, Vol. I (1969), Vol. II (1973). Hafner Publishing Co., New York.
- [16] Koopmans, L.H. (1974) The Spectral Analysis of Time Series, Academic Press, New York.
- [17] Leanov, V.P. and A.N. Shiryaev (1959) On a method of calculation of semi-invariants, *Theory Prob. and Appl.*, IV, 319-329.
- [18] Parzen, E. (1960) Modern Probability Theory and Its Applications, John Wiley and Sons, New York.
- [19] Puri, P.S. (1968) Some further results on the birth-and-death process and its integral, *Proc. Camb. Phil. Soc.*, 64, 141-154.
- [20] Royden, H.L. (1968) Real Analysis, Second Edition, MacMillan, Toronto.
- [21] Sawyer, S. (1976) Branching diffusion processes in population genetics, *Adv. Appl. Prob.*, 8, 650-689.

Appendix I: Recursive Formulae Relating Joint Cumulants and Joint Moments.

Let $\{X_n\}$ be a sequence of random variables,

$$\mu_n(X_1, \dots, X_n) = E(X_1 \dots X_n),$$

$$\kappa_n(X_1, \dots, X_n) = \text{cum}(X_1, \dots, X_n).$$

The functionals μ_n and κ_n are joint moments and cumulants.

Theorem

$$\begin{aligned} \mu_n(X_1, \dots, X_n) \\ = \kappa_n(X_1, \dots, X_n) \end{aligned} \quad (1)$$

$$+ \sum_{j=1}^{n-1} \sum_{\substack{\rho \in S_j \\ \rho(1)=1}} \kappa_j(X_1, X_{\rho(2)}, \dots, X_{\rho(j)}) \mu_{n-j}(X_{\rho(j+1)}, \dots, X_{\rho(n)}),$$

where $\sum_{\substack{\rho \in S_j \\ \rho(1)=1}}^{n-1}$ is the sum over the $\binom{n-1}{j-1}$ partitions of $\{1, \dots, n\}$ into A_1

and A_2 , where $1 \in A_1$.

Proof: (i) Case where m.g.f's exist.

$$E\{\exp[X_1 t_1 + \dots + X_n t_n]\}$$

$$= \exp\{\kappa_1(x_1)t_1 + \dots + \kappa_1(x_n)t_n + \kappa_2(x_1, x_2) \frac{t_1 t_2}{2!} + \dots\}.$$

Formally differentiating with respect to the κ_L 's, and comparing coefficients of the t_i 's, one obtains for distinct $i_1, \dots, i_L$,

$$\frac{\partial \mu_n(x_1, \dots, x_n)}{\partial \kappa_L(x_{i_1}, \dots, x_{i_L})} = \begin{cases} \mu_{n-L}[(x_1, \dots, x_n) \setminus (x_{i_1}, \dots, x_{i_L})], & \text{if } \{i_1, \dots, i_L\} \subseteq \{1, \dots, n\} \\ 0, & \text{otherwise} \end{cases}$$

where $\mu_{n-L}[(x_1, \dots, x_n) \setminus (x_{i_1}, \dots, x_{i_L})] = \mu_{n-L}(x_{j_1}, \dots, x_{j_{n-L}})$,

where $\{j_1, \dots, j_{n-L}\} = \{1, \dots, n\} \setminus \{i_1, \dots, i_L\}$.

(ii) It is known that μ_n is a polynomial in the cumulants, with zero constant coefficient. By induction on n and L , it can be shown that (1) is the unique solution of the system, in (i), of differential equations, in the class of polynomials with zero constant coefficient.

(iii) In the case of only finitely many cumulants, it is known that moments and cumulants, as far as they exist, satisfy the usual relation in terms of partitions of positive integers (Leonov and Shiryaev (1959)). Since (1) is just a rewriting of this, the theorem still holds. //QED.

Corollary (Kendall and Stuart (1969), p. 90)

If the random variables $X_1, \dots, X_n$ are equal, a.e., (1) becomes

$$\mu_n = \kappa_n + \sum_{j=1}^{n-1} \binom{n-1}{j-1} \kappa_j \mu_{n-j},$$

where $\mu_n = \mu_n(X_1, \dots, X_n)$, etc.

Appendix II: Leonov and Shiryaev's Theorem

Let $\{X_{i,j}: 1 \leq j \leq m_i, 1 \leq i \leq n\}$ be a collection of random variables with joint cumulants existing. Let

$$Y_1 = X_{1,1} X_{1,2} \cdots X_{1,m_1},$$

$$\vdots$$

$$Y_n = X_{n,1} X_{n,2} \cdots X_{n,m_n}.$$

Consider the corresponding TABLE 1 of indices, with rows $R_1, \dots, R_n$.

TABLE 1

(1,1)	(1,2)	...	(1,m ₁)
(2,1)	(2,2)	...	(2,m ₂)
⋮			
(n,1)	(n,2)	...	(n,m _n)

Consider a partition $\mathcal{P} = \{P_1, \dots, P_t\}$ of the elements in TABLE 1.

$\mathcal{P}$ is said to be indecomposable if for all $J = \{j_1, \dots, j_\ell\} \subsetneq \{1, \dots, n\}$ and $I = \{i_1, \dots, i_s\} \subsetneq \{1, \dots, t\}$,

$$\bigcup_{i \in I} P_i \neq \bigcup_{j \in J} R_j.$$

Example: $n = 4$, $m_i = 2$ for all i .

$$(i) P_1 = \{(1,1), (2,1)\}, P_2 = \{(3,1), (4,2)\}, P_3 = \{(1,2)\}, P_4 = \{(2,2)\}, \\ P_5 = \{(3,2), (4,1)\}.$$

$\mathcal{P} = \{P_1, P_2, P_3, P_4, P_5\}$ is not indecomposable since

$$P_1 \cup P_3 \cup P_4 = R_1 \cup R_2.$$

$$(ii) P_1 = \{(1,1), (2,1), (3,1)\},$$

$$P_2 = \{(4,1), (1,2)\},$$

$$P_3 = \{(2,2), (3,2)\},$$

$$P_4 = \{(4,2)\},$$

$\mathcal{P} = \{P_1, P_2, P_3, P_4\}$ is indecomposable.

Let $C(P)$ be the joint cumulant of the $X_{i,j}$ with indices in P .

Theorem (Leonov and Shiryaev (1959))

$$\overline{\text{cum}\{Y_1, \dots, Y_n\}} = \sum^* C(P_1) \dots C(P_t)$$

$$\mathcal{P} = \{P_1, \dots, P_t\}$$

where $\sum^*$ is the sum over all indecomposable partitions of TABLE 1.

This Theorem is also given in Brillinger (1975) with proof.

The most important case in this work is the case $m_i = 2$ for all i and where indecomposable partitions contain no sets of size 1. Also in this special case, $(i,1) + (i,2) = 0$. TABLE 1 may now be written as

TABLE 2

s_1	$-s_1$
s_2	$-s_2$
$\vdots$	
s_n	$-s_n$

This case has some special properties which makes the "rectangular taper" so useful in Chapter 4.

Let $\mathcal{P} = \{P_1, \dots, P_t\}$ be an indecomposable partition of TABLE 2 with $|P_i| \neq 1 \quad \forall i$. Let

$$u_1 = \sum P_1 \equiv \text{sum over all the elements in } P_1,$$

$$\vdots$$

$$u_t = \sum P_t.$$

Then $(s_1, \dots, s_n) \mapsto (u_1, \dots, u_{t-1}, s_{i_t}, \dots, s_{i_n})$ is a 1-1 map for some appropriate $\{i_t, \dots, i_n\}$, which is non-empty. *

Example: $n = 4, \mathcal{P} = \{P_1, P_2, P_3\}$,

$$P_1 = \{s_1, -s_2\}, P_2 = \{s_3, s_4, -s_4\}, P_3 = \{-s_1, s_2, -s_3\}.$$

$$u_1 = s_1 - s_2, u_2 = s_3,$$

$(s_1, s_2, s_3, s_4) \mapsto (u_1, u_2, s_2, s_4)$ is 1-1.

Proof of *:

Formally consider $V = V_n(R)$, an n -dimensional vector space over R , with basis $s_1, \dots, s_n$. To show *, it suffices to show $\{u_1, \dots, u_{t-1}\}$ is linearly independent, and complete the basis from the s_i 's. * is proved by induction on t .

$t = 2$ is trivial (since $t-1 = 1$)

Suppose the result is true for $2, \dots, t-1$. Consider t and

$$a_1 u_1 + \dots + a_{t-1} u_{t-1} = 0. \quad (1)$$

(i) Suppose there exists $s_i \in P_1 \cup \dots \cup P_{t-1}$, $-s_t \in P_t$, say $s_i \in P_k$ (or similarly $s_i \in P_k^*$, $-s_i \in P_1 \cup \dots \cup P_{t-1}$). In this special case, then (1) gives $a_k = 0$.

Notice $\{P_\ell, 1 \leq \ell \leq t-1, \ell \neq k\} \cup \{P_k \cup P_t\}$ is an indecomposable partition, of size $t-1$. By the induction hypothesis, then (1) holds only if

$a_1 = \dots = a_{t-1} = 0$, that is, $\{u_1, \dots, u_{t-1}\}$ is linearly independent.

(ii) For all $s_i \in P_1 \cup \dots \cup P_{t-1}$, $-s_i \in P_1 \cup \dots \cup P_{t-1}$.

Since the partition is indecomposable, $\exists s_i, P_{j_1}, P_{j_2}, j_1 \neq j_2 \Rightarrow s_i \in P_{j_1}, -s_i \in P_{j_2}$. Then (1) holds only if $a_{j_1} = a_{j_2}$ (since $a_{j_1} - a_{j_2}$ is the coefficient of s_i). Without loss of generality, $j_1 = 1, j_2 = 2$.

Then (1) can be rewritten as

$$a_1(u_1 + u_2) + a_3u_3 + \dots + a_{t-1}u_{t-1} = 0, \quad (2)$$

which arises from the indecomposable partition $\{P_1 \cup P_2, P_3, \dots, P_t\}$ of size $t-1$. By the induction hypothesis, $a_1 = \dots = a_{t-1} = 0$. //QED.

END

06

00

79

IN