UNIVERSITY OF CALIFORNIA
RIVERSIDE

Modeling Almost Periodicity in Point Processes

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Applied Statistics

by

Nan Shao

December 2010

Dissertation Committee:

Dr. Keh-Shin Lii, Chairperson
Dr. Barry C. Arnold
Dr. Tae-Hwy Lee

The Dissertation of Nan Shao is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I would like to express my greatest appreciation to my advisor, Dr. Keh-Shin Lii, for his inspiring guidance and constant support. Thank you for sharing with me your brilliant expertise and insights and showing me the way of thinking, working and living.

I am sincerely grateful to my committee members, Dr. Barry C. Arnold, who is always available when I need his advice, and Dr. Tae-Hwy Lee for his thoughtful questions. My profuse thanks are extended to Dr. Daniel Jeske, Dr. James Flegal, Dr. Subir Ghosh, Dr. Jun Li, Dr. Vanamamalai Seshadri, Dr. Aman Ullah and Dr. Gloria Gonzalez-Rivera for their knowledge and influence on my practise of statistics. Additionally, I would like to thank Dr. Linda Penas and Paula Lemire for their friendship and support in my daily life.

My special thanks go to my husband, Bushi Wang, for always being there whenever and wherever I need him. I also would like to acknowledge the encouragement and supportiveness of my parents, Bei Wang and Xinmin Shao, and my grandparents, Niancui Ding and Huaijie Wang.

To my husband, Bushi Wang; my parents, Xinmin Shao and Bei Wang, for all the support

ABSTRACT OF THE DISSERTATION

Modeling Almost Periodicity in Point Processes

by

Nan Shao

Doctor of Philosophy, Graduate Program in Applied Statistics
University of California, Riverside, December 2010
Dr. Keh-Shin Lii, Chairperson

We propose a model for the analysis of non-stationary point processes with almost periodic rate of occurrence. The model deals with the arrivals of events which are unequally spaced and show a pattern of periodicity or almost periodicity, such as the rate of financial transactions or customer/phone calls arrivals. The concept of almost periodicity is described and the purely periodic process is just a special case of the almost periodic process. We consider a non-homogeneous Poisson process and model its rate of occurrence as the sum of sinusoidal functions plus a base line. Given the number of sinusoidal functions which is denoted as K, a set of simple and consistent estimates of frequencies, phases and amplitudes which form the sinusoidal functions are constructed mainly by the Bartlett periodogram. The estimates are shown to be asymptotically normally distributed. Computational issues are discussed and it is shown that the frequency estimates have to be resolved with order $o(T^{-1})$ to guarantee the asymptotic unbiasedness and consistency of the estimates of phases and amplitudes, where T is the length of the observation period. The prediction of the next occurrence is also discussed. The proposed model is a finite approximation of the almost periodic process in terms of a finite value of K. In practice, the value of K is usually unknown, and we suggest to use the model selection criteria to determine K. Two criteria AIC and BIC are reviewed and discussed in the frame work of our

model. Simulation and real data examples are used to illustrate the theoretical results and the utility of the model.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis presents a general model for the point processes which show certain pattern of periodicity. In Chapter 1, we introduce the basic concept of point processes, including its definition, intensity function, finite Fourier transform and the periodogram of a point process. The thesis is on the modeling of almost periodic point processes, and the motivation of studying such processes and the concept of almost periodicity are introduced in Chapter 2. The new model is presented and studied in Chapter 2 and Chapter 3. The procedures to implement the proposed methodology in the data analysis are summarized and presented in Chapter 3, section 3.4. A discussion of potential future research directions is in Chapter 4.

1.1 Basic concept of point processes

Point processes are very common in daily life. A temporal point process is the random occurrence of a series of events. So the arrivals of customers in a restaurant, the initiations of phone calls made to a customer service center, the times of stock transactions, and the occurrences of earthquakes in certain area are all temporal point processes. There are also spatial point processes which are the random locations where the events of interest occur, such as the

location of the wild fire. In this dissertation, we only consider temporal point processes, and use temporal point processes and point processes interchangeably.

Studying point processes is meaningful, as it helps us understand the structure of the process which enables us to forecast. For example, the rate of stock transactions carries information about the market which influences the stock market value and volume of shares as well as other financial products. In this case, it is natural to model the timing of transactions. Moreover, a rigorous understanding of how the market moves provides investors with vital information to help them make financially wise investments. Moreover, in the study of phone calls made to a customer service center, a good understanding of the pattern of phone call arrivals is important to the scheduling of operators and thus could increase the efficiency of the call center.

Point processes have been extensively studied in the literatures, such as Bartlett (1957, 1963, 1967, 1978), Brillinger (1972, 1978, 1982, 1983, 2003, 2008), Cox and Lewis (1966), Lewis (1970, 1972), Cox and Isham (1980), Imoto et al. (1999), Vere-Jones (1982), Vere-Jones and Ozaki (1982), and Hassan and Lii (2006). The research of point processes usually focuses on three aspects, namely, the counting processes which is another way to express point processes, the intensity and/or the spectral of point processes, and the inter-arrival times of point processes. The concept of counting processes and intensity will be reviewed later in this section. There are also literatures in the random sampling of a continuous stochastic process, such as Lii and Masry (1994).

In the research the point process is defined on the half real line $\mathbb{R}^+$ for convenience. One specification of the point process is from its event times. Let $\{t_j, j = 1, 2, \cdots\}$ be a sequence of nonnegative random variables with $0 \leq t_j \leq t_{j+1}$, then the sequence $\{t_j\}$ is the **point process** on $[0, \infty)$. If there is no multiple occurrence, namely, $t_j < t_{j+1}$ for any j, the process is called a **simple point process**. We will restrict our consideration to simple point processes.

The point process can also be specified by the counting process $N(t)$ where $N(t)$ repre-

sents the total number of points or "events" that have occurred up to time t ($t \geq 0$), so the counting process $N(t)$ must satisfy:

(i) $N(t) \geq 0$.

(ii) $N(t)$ is integer valued.

(iii) If $s < t$, then $N(s) \leq N(t)$.

(iv) For $s < t$, $N(t) - N(s)$ equals the number of events that have occurred in the interval $(s, t]$.

A counting process is said to possess **independent increments** if the numbers of events that occur in disjoint time intervals are independent. That means, for example, the number of events that occur by time 10 and between times 10 and 15 must be independent. Considering customer arrivals, if $N(t)$ is the number of customers who enter a particular store at or prior to time t, then $N(t)$ is a counting process, and it would be reasonable to assume that the number of customers who enter the store during one period of time is independent of the number of customers who enter the store during another disjoint period of time.

A counting process is said to possess **stationary increments** if the distribution of the number of events that occur in any interval of time depends only on the length of the time interval. In other words, the process has stationary increments if the number of events in the interval $(s, s+t)$ has the same distribution for all s. A point process with stationary increments is called a stationary point process. There are other mathematical definitions of a stationary stochastic process, and we refer to Priestley (1981) for more information. In this dissertation, the processes we consider are not stationary, and they are non-homogeneous Poisson processes which will be discussed in the next section.

The definition and properties of the counting process above are from Ross (2007) Chapter 5.

In point processes, the only observations are the points, or the occurrence time of events.

We can describe the pattern of occurrence as how frequently the events occur, and it is the idea of the intensity function. The **intensity function** $\lambda(t)$ of a point process, also called the jumping rate, or the mean rate of occurrence in Cox and Isham (1980), is defined as

$$\begin{aligned}\lambda(t) &= \lim_{\delta \to 0+} \frac{P\{N(t, t+\delta) > 0\}}{\delta} \\ &= \lim_{\delta \to 0+} \frac{E\{N(t, t+\delta)\}}{\delta}.\end{aligned}$$

So $E\{dN(t)\} = \lambda(t)dt$ with $dN(t) = N(t+dt) - N(t)$. The integral of the intensity function over a time period $(s, t]$, that is, $\int_s^t \lambda(t)dt$, is the expected number of events which would occur in $(s, t]$. And thus the intensity function describes the first-order moment property of the unconditional counting measure. When the intensity is large in a period of time, we should expect more events to occur during that period of time, and vice versa. In addition, it is easy to see that if the process is stationary, the intensity function is constant since a stationary process will have the same expected number of events in any time intervals with the same interval length.

The intensity function is an important concept in point processes, and this thesis is on the modeling of the intensity function of a non-homogeneous Poisson process. We introduce the Poisson process in the next section.

1.2 Poisson processes

Poisson processes are the most commonly used point processes in practice. It is defined as follows:

Definition 1 *The counting process* $\{N(t), t \geq 0\}$ *is said to be a* ***Poisson process*** *having rate or the intensity function* $\lambda(t)$ *with* $\lambda(t) \geq 0$*, if*

(i) $N(0) = 0$.

(ii) The process has independent increments.

(iii) $P\{N(t+h) - N(t) \geq 2\} = o(h)$.

(iv) $P\{N(t+h) - N(t) = 1\} = \lambda(t)h + o(h)$.

We state the following result and refer the proof to Ross (2007).

The number of events in the interval $(0, t]$ is Poisson distributed with mean $\Lambda(t) = \int_0^t \lambda(s)ds$. That is, for all $t > 0$,

$$P\{N(t) = n\} = e^{\Lambda(t)} \frac{\Lambda^n(t)}{n!}, \quad n = 0, 1, \ldots$$

Note that when the intensity function is constant, that is, $\lambda(t) = \lambda > 0$, the Poisson process is called the homogeneous Poisson process and it is stationary. Otherwise, the process is called the non-homogeneous Poisson process, and it is not stationary.

For a homogeneous Poisson process with intensity λ, its inter-arrival times, $\{x_j : x_j = t_j - t_{j-1}, \text{if } j \geq 2; x_1 = t_1\}$, are independent and identically distributed (i.i.d.) exponential random variables having mean $1/\lambda$. So to generate a homogeneous Poisson process, we just need to generate a series of i.i.d. exponential random variables $\{x_j\}$ with mean $1/\lambda$, and the data points or event times are $\{t_j = \sum_1^j x_i, j \geq 1\, t_0 = 0\}$.

Denote the time to stop taking observations as T (assume we always taking observations start at time 0). In the point process literatures, the sample size $n = N(T)$ is usually regarded as a random quantity for mathematical convenience, and the observation length T is fixed.

1.3 Simulation of a non-homogeneous Poisson process

If the intensity function is bounded, that is, $\lambda(t) \leq \lambda$ for all $t \geq 0$ ($\lambda > 0$), a non-homogeneous Poisson process can be generated by thinning method (Lewis and Shedler (1979)). The procedures are as follows:

(i) For each $j \geq 1$, generate an exponential random variable x_j with rate $1/\lambda$, and x_j is independent of the previously generated $\{x_i, i < j\}$. Denote $t_j = \sum_1^j x_i$, so $\{t_1, t_2, \ldots\}$ is a homogeneous Poisson process with rate λ.

(ii) Generate an uniformly distributed random variables $u_j \sim U(0,1)$, and u_j is independent of the homogeneous Poisson process $\{t_1, t_2, \ldots\}$ and the previously generated $\{u_i, i < j\}$. If $u_j \leq \frac{\lambda(t_j)}{\lambda}$, keep t_j, otherwise, delete t_j.

(iii) Repeat step (ii) until the desired observation length T of the "thinned" sequence is obtained.

Denote the "thinned" sequence as $\{t_j^*\}$, then it is a non-homogeneous Poisson process with the intensity function $\lambda(t)$.

1.4 Finite Fourier transform and periodogram of point processes

The **finite Fourier transform** of a point process is defined by Bartlett (1963) as

$$d_T(\omega) = \frac{1}{\sqrt{2\pi T}} \int_0^T e^{-i\omega t} dN(t),$$

and the **periodogram** of a point process is defined by the squared norm of the finite Fourier transform

$$\begin{aligned} I_T(\omega) &= \frac{1}{2\pi T} \Big| \int_0^T e^{-i\omega t} dN(t) \Big|^2 \\ &= \frac{1}{2\pi T} \Big| \sum_{t_j \leq T} e^{-i\omega t_j} \Big|^2, \end{aligned}$$

where T is the observation length. Again, we consider T as a fixed quantity and the sample size $n = N(T)$ as a random number.

Periodogram can be calculated directly from the data $\{t_j : j = 1, 2, \ldots, t_j \leq T\}$.

The intensity function of a point process is unobservable. However, as illustrated by Figures 1.1 and 1.2 and the discussion afterwards, the periodogram is closely related to the intensity function when the point process follows a non-homogeneous Poisson process and the intensity function takes the form

$$\lambda(t) = \sum_{k=1}^{K} A_k \cos(\omega_k t + \phi_k) + B,$$

where $\lambda(t) \geq 0$ and $A_k > 0$ and $\omega_k > 0$, $k = 1, \ldots, K$.

Figures 1.1 and 1.2 show an example of the intensity function of a non-homogeneous Poisson process and the corresponding periodogram calculated from one realization of the non-homogeneous Poisson process.

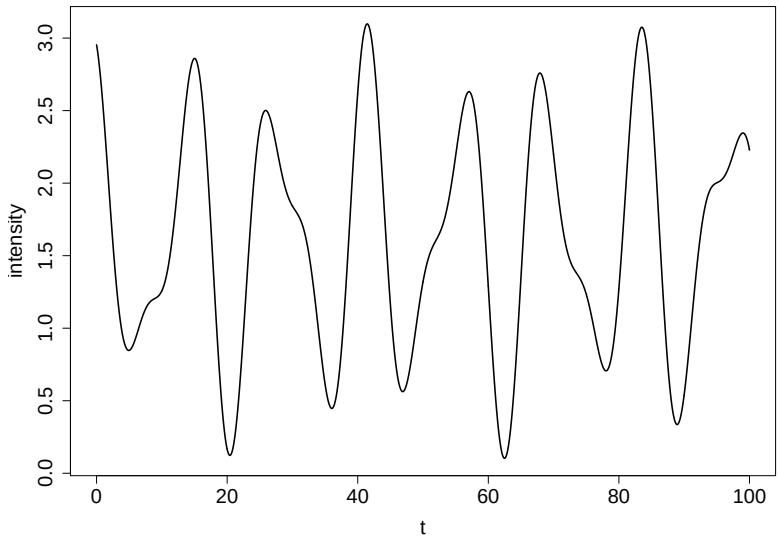


Figure 1.1: Intensity function of a non-homogeneous Poisson process $\lambda(t) = \cos(\frac{\pi}{4\sqrt{3}}t) + 0.5\cos(\frac{\pi}{3\sqrt{2}}t + \frac{\pi}{4}) + 1.6$.

The location of the two largest peaks of the periodogram correspond to the two frequencies $\frac{\pi}{4\sqrt{3}}$ and $\frac{\pi}{3\sqrt{2}}$ in the intensity function, and the height of the peaks is closely related to the magnitude of the two cosine functions in the intensity function. This relationship between periodogram and the intensity function of a non-homogeneous Poisson process will be investigated in more details in Chapter 2.

The properties of the periodogram are studied by Bartlett (1963) when the periodogram is derived from a stationary process. Brillinger (1972) studied the spectral analysis of stationary interval functions and gives the asymptotic distribution of the periodogram. We refer to these

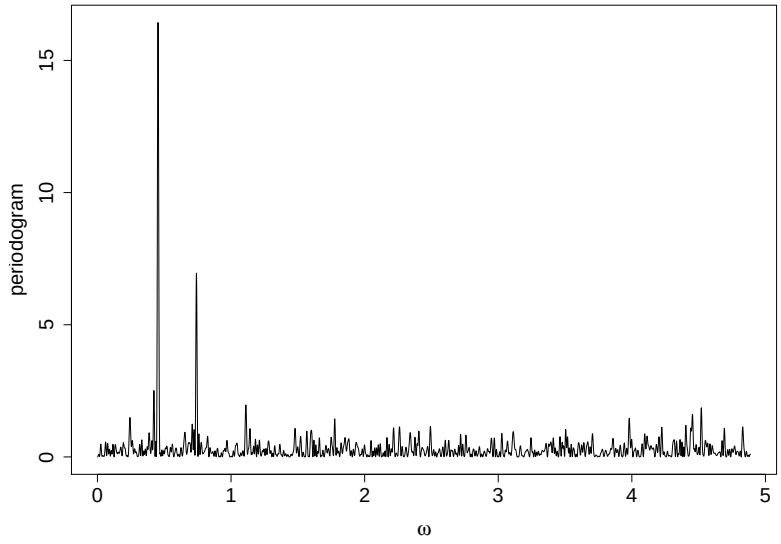


Figure 1.2: The periodogram of a non-homogeneous Poisson process with the aforementioned intensity function $\lambda(t) = \cos(\frac{\pi}{4\sqrt{3}}t) + 0.5\cos(\frac{\pi}{3\sqrt{2}}t + \frac{\pi}{4}) + 1.6$. The process consists 778 observations with observation length $T = 500$.

two references for further discussion of the periodogram. In this dissertation, our main interest is in the non-homogeneous Poisson process, and the process is not stationary.

Chapter 2

Almost Periodic Poisson Processes

2.1 Introduction

The main objective of this chapter is to present a model for non-homogeneous Poisson processes with periodic or aperiodic but almost periodic rate of occurrence. The work in this chapter has been summarized in the paper 'Modelling non-homogeneous Poisson processes with almost periodic intensity functions' which will appear in *Journal of the Royal Statistical Society Series B* and now can be downloaded at http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9868.2010.00758.x/pdf. This chapter and Appendix A provide more details than the paper, especially in the proof.

There are many real life examples of point processes which show a pattern of periodicity, such as higher volume of customers coming to a restaurant during Friday night and weekend than weekdays, stock transaction with high activity at the beginning and the end of the day (Engle (2000)), more phone calls made to a customer service center on Monday morning than Friday evening, and even in seismicity where periodic pattern of earthquakes has been observed in certain area (Vere-Jones and Ozaki (1982) and Imoto et al. (1999)). Many applications arise in areas such as healthcare and medical sciences (Lewis (1970, 1972) and Kuhl et al. (1995)), me-

teorology (Lee et al. (1991)), and some others are noted in Helmers et al. (2003) and references therein.

As introduced in Chapter 1, the rate of occurrence is also called the intensity function, and more events in a certain period of time is equivalent to larger value of intensity function during that time period. So when a point process shows a pattern of periodicity, its intensity function is modeled as a periodic function. Many references have discussed the modeling of periodic intensity functions of Poisson processes, such as Lewis (1970, 1972), Vere-Jones (1982), Helmers et al. (2003, 2005), but none of them is in the almost periodic context. The novelty of our work is in the presentation of constructing a much more general model which includes the purely periodic Poisson processes as a special case. The concept of 'almost periodicity' and our model will be introduced in the next section. We will give a literature review on the parametric and non-parametric approaches of modeling periodic Poisson process as follows.

Point processes with single or multiple periodicities of which the corresponding frequencies are integer multiples of a fundamental frequency have been studied in the parametric context. Lewis (1970, 1972) established the estimation and detection of a cyclic varying rate of a non-homogeneous Poisson process when the frequency is known *a priori*. The rate function took the form

$$\lambda(t) = A \exp\{\rho \cos(\omega t + \phi)\} \tag{2.1}$$

which would lead to simple results based on sufficient statistics and guarantee the positivity of the intensity function. Vere-Jones (1982) considered and established the asymptotic properties of a consistent estimate for the unknown frequency ω in (2.1), which is based on the maximum of the Bartlett periodogram (Bartlett (1963)) over certain frequency range and coincides with the maximum likelihood estimate. As introduced in Chapter 1, the periodogram of a point process

is defined as

$$I_T(\omega) = (2\pi T)^{-1}\Big|\int_0^T e^{-i\omega t}dN(t)\Big|^2 = (2\pi T)^{-1}\Big|\sum_{t_j\le T} e^{-i\omega t_j}\Big|^2, \tag{2.2}$$

where $(0, T)$ is the observation interval, $N(t)$ is the counting process which is the number of events by time t, and $\{t_j : t_j \le T\}$ are the random occurrences of events. However, (2.1) is limited in its parametric form, that is within each cycle the intensity function has a sharp peak and a flat trough, and the model may not perform well when the true intensity function departs from this parametric form, such as a flat peak and a steep trough or several peaks in one cycle. Kuhl (1994) and Kuhl et al. (1995, 1997) extended Lewis's model where the rate of Poisson processes was an exponential function with the sum of several sinusoidal functions plus a baseline constant in the exponent. An approximated likelihood was used in calculating the parameter estimates numerically, but this approximated likelihood may differ from the true likelihood severely when the baseline in the exponent is relatively small compared to the amplitudes of the sinusoidal functions in the exponent. In addition, there is no justification of the goodness of the initial values of the frequencies which is usually of most interest, nor the statistical properties of the final parameter estimates. A series of paper by J. Garrido and Y. Lu have proposed and investigated several parametric doubly periodic Poisson models and apply the models to the hurricane data to include the El Niño and La Niña effect. We refer to Lu and Garrido (2005) and the reference therein. The intensity function is expressed as a product of two periodic functions that represent short-term and long-term trend. In their paper, both short- and long-term periods are assumed to be known and the long-term period is an integer multiple of the short-term period, so this doubly periodic Poisson model is a special case of the purely periodic Poisson model.

The period or, equivalently, frequency of the periodic Poisson process has been studied non-parametrically as well, we refer to Mangku (2001) and Bebbington and Zitikis (2004). An updated version of Mangku (2001) is Helmers and Mangku (2003). The main idea in Helmers

and Mangku (2003) is that if the intensity function is periodic with period τ, then the counts of events within each time interval of length τ should share the same expected value, so one can find out the estimate of τ by searching for the proper interval length which minimizes the sum of squared difference between counts within each time interval and the averaged counts. However, this approach is flawed in the following sense: the expected value of the number of counts in a time interval is the integrated value of the intensity function in that time interval, so if the intensity function is periodic and the integrated intensity within the first half period is the same as the integrated intensity within the latter half period, the approach in Helmers and Mangku (2003) would find a period estimate around half of the true period, not the complete true period. An example for this case is that the intensity function is $\lambda(t) = 0.1 + |\cos(2\pi t)|$ with period 1 and it is symmetric within each period, so the integrated intensity in the first half period is the same as the one in the second half period, and thus the approach in Helmers and Mangku (2003) would estimate the period around 0.5 instead of 1. Follow a similar argument, a smaller divisor of the period could be found as the period estimate if the integrated intensities are the same within each sub-divisor of the period. In addition, the search range for the period estimate is not well defined, or relies on the prior knowledge of the length of the period which is sometimes unavailable in practice. Bebbington and Zitikis (2004) constructed a family of non-parametric estimators for the period of a cyclic Poisson process, and one candidate is the twice of the length of the interval which maximizes the sum of squared difference between counts in adjacent time intervals. The authors generalized this estimator to a family of non-parametric estimators and used simulations to demonstrate the utility of the results. From the discussion in the paper, the proposed estimator may find the period estimate to be half of or even smaller divisor of the true period. In addition, the asymptotic and statistical properties of these non-parametric estimators are not shown in the paper though the authors made some remarks on the derivation of the asymptotic properties. One needs to pay attention to the claim made in Bebbington and

Zitikis (2004) that the periodogram cannot handle multimodal cycles, namely, multiple peaks in one cycle. This is not true, and this claim is made solely based on the simulation results without theoretical proof. In fact, in their simulation, the periodogram estimator is searched too sparsely, the searching grid may not be fine enough to have the length $o(T^{-1})$ where T is the observation length. The computational issues on searching the periodogram estimator will be discussed in this chapter.

There are a series of papers on non-parametric estimation of the intensity function of a cyclic Poisson process. Helmers et al. (2003, 2005) constructed and investigated a consistent kernel-type non-parametric estimator of the intensity function, where the unknown period τ is first estimated by periodogram-based estimator in Vere-Jones (1982) or the non-parametric estimator in Helmers and Mangku (2003), and then the intensity function at time s is estimated by using the data points near $s + k\hat{\tau}$, where k is non-negative integer and $\hat{\tau}$ is the estimator of τ. The idea in estimating the shape of the intensity function is essentially similar as the 'folding' technique in Hall et al. (2000) although we believe that it is a coincidence. Hall et al. (2000) introduced a general framework of nonparametric estimation of a periodic function when observations were made at irregularly spaced time. The main difference between the estimation of general periodic function and the estimation of periodic intensity function of a Poisson process is that the value of the general periodic function is observable (as the response), but the intensity function of a point process is unobservable. The only observations in a point process are the points, or the occurrence time of the series of events.

In addition, motivated by J. Garrido and Y. Lu's work, Helmers et al. (2007) constructed and analyzed a non-parametric estimator for the doubly periodic Poisson intensity function under the same assumptions as in Lu and Garrido (2005). Specifically, the long-term periodic function is modeled as a step function which takes positive constant value over the whole short-term period. So within one long-term period, each short-term periodic effect is a constant multiples

of the first short-term effect. This structure of long-term periodic function is used in one of the parametric models in Lu and Garrido (2005). The estimation of the short-term periodic function is in essence similar to the technique in Helmers et al. (2003, 2005).

Moreover, the non-parametric prediction upper bound for a future observation of a cyclic Poisson process has been studied in Helmers and Mangku (2009) where the period is assumed to be known.

In the preceding non-parametric estimation and prediction, the assumption of a purely periodic intensity function is critical so that the intensity function repeats itself exactly, and thus can be extended to the future for prediction or produces 'replicates' of the process even when there is only one realization of the process. For the case when the intensity function is almost periodic but not periodic then the intensity function does not repeat itself exactly and in this case it is difficult to implement the non-parametric technique for estimation and prediction by these non-parametric methods. We will give more information on almost periodic function in the next section.

The rest of this chapter is organized as follows. In section 2.2, we introduce the concept of almost periodic function and our model. In section 2.3, we introduce the notations and assumptions, and state the main results in estimating the parameters in the intensity function. In section 2.4, we discuss some computational issues in estimating the frequencies. Prediction of the next occurrence is discussed in section 2.5. Simulation studies are carried out in section 2.6 and a real life data set is analyzed in section 2.7. We conclude with some further discussion in section 2.8. Appendix A contains most of the proofs of the results.

2.2 Our model

2.2.1 Almost periodic function

There are various different but equivalent definitions of almost periodic function. We use the definition in Corduneanu (1989).

Definition 2 *A complex valued function $f(x)$ defined for $-\infty < x < \infty$ is called almost periodic, if for any $\varepsilon > 0$ there exists a trigonometric polynomial $T_\varepsilon(x)$, such that*

$$|f(x) - T_\varepsilon(x)| < \varepsilon, \quad -\infty < x < \infty,$$

where the trigonometric polynomial is in the form of $T(x) = \sum_{k=1}^{n} c_k e^{i\lambda_k x}$, and c_k are complex numbers and λ_k are real numbers.

Thus almost periodic functions are those functions defined on the real line, which can be uniformly approximated by trigonometric polynomials.

An equivalent definition of almost periodic function is given by Bohr (1947) as follows.

Definition 3 *For any $\varepsilon > 0$, if there exists a number $l(\varepsilon) > 0$ with the property that any interval of length $l(\varepsilon)$ of the real line contains at least one point with abscissa ξ, such that*

$$|f(x + \xi) - f(x)| < \varepsilon, \quad -\infty < x < \infty,$$

then $f(x)$ is almost periodic.

We refer to Corduneanu (1989) for the proof of equivalence of above two definitions and more details of almost periodic functions.

From definition 2 it follows that any trigonometric polynomial is an almost periodic function. In addition, any periodic function which has a Fourier series representation, such as continuous periodic function, is also almost periodic, in other words, such periodic functions are just special cases of almost periodic functions.

Figure 1.1 shows an almost periodic function. The function has certain pattern of periodicity, but there is no period τ such that the function at any time t has exactly the same value as it is at time $t + \tau$. The function allows some deviation from the purely periodic case.

2.2.2 Almost periodic point processes

In practice, almost periodic process is much more general than the periodic one, since any particular configuration that occurs once may recur not exactly, but within some level of accuracy. And it is more reasonable to assume an almost periodic intensity function when the point process shows a patter of periodicity. In the point process case, the intensity function is defined on the non-negative real line, and by definition 2 that any almost periodic function can be uniformly approximated by the trigonometric polynomials, we model the almost periodic point process with the following intensity function,

$$\lambda(t) = \sum_{k=1}^{K} A_k \cos(\omega_k t + \phi_k) + B, \tag{2.3}$$

where A_k, B, ω_k, ϕ_k are unknown parameters with $A_1 > A_2 > \cdots > A_K > 0$, $-\pi/2 \leq \phi_k < 3\pi/2$ and $\omega_k > 0$, $k = 1, \ldots, K$. The baseline B is a constant such that $\lambda(t)$ is non-negative for any $t > 0$. A sufficient condition to guarantee the non-negativity of $\lambda(t)$ is $B \geq \sum_{k=1}^{K} A_k$.

By Fourier expansion, the sum of sinusoidal functions can capture most of the variation of a periodic function in any shape, and when the frequencies are not integer multiples of a fundamental frequency, (2.3) is not periodic but almost periodic. Chen (2006) first used this model to test the existence of the periodic components when the frequencies $\{\omega_k : k = 1, \ldots, K\}$ are known. He also proposed a method of detection of hidden periodicity. There, Poisson process assumption was relaxed, while the process is assumed to have stationary increment under null hypothesis.

2.3 Assumptions and Estimation

We consider simple point process in which at most one event occurs at any given time. We also assume the point process is a non-homogeneous Poisson process.

Assumption 4 *$N(t)$, $t \geq 0$, is a non-homogeneous Poisson process with a periodic or almost periodic intensity function* (2.3)*, observed over the time interval $[0, T]$, and the number of periodic components K is given.*

The estimation of K will be discussed in Chapter 3.

Inspired by Vere-Jones (1982), we construct the estimates of all unknown parameters in (2.3) mainly by the periodogram (2.2) for non-homogeneous Poisson processes. The estimates do not coincide with the maximum likelihood estimates, but they can be very good initial values in finding the MLE's. Denote $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_K)'$ in (2.3). We determine $\hat{\boldsymbol{\omega}}_T = (\hat{\omega}_{1,T}, \ldots, \hat{\omega}_{K,T})$ as the estimate of $\boldsymbol{\omega}$ which are frequencies corresponding to the K largest local maxima of the periodogram (2.2) under a certain minimum separation condition that is explained below. Analogous to the estimation of several harmonic components in the ordinary time series analysis (Walker (1971)), a minimum separation condition on the $\boldsymbol{\omega}$ must be imposed to keep the ω_k from being too close together and thus prevent the estimates of two angular frequencies from converging to the same value in probability. We have the following assumption.

Assumption 5 $T \min_{k \neq k'}(|\omega_k - \omega_{k'}|) \to \infty$*, as* $T \to \infty$.

The choice of minimum separation depends on the data length T and the rate of occurrence $\lambda(t)$. If the number of observations is large, in other words, T is large and/or $\lambda(t)$ is large, then we should be able to estimate frequencies which are closer to each other.

In addition, the maxima of the periodogram (2.2) should be searched in an appropriate range. For a fixed data set, the periodogram (2.2) is an almost periodic function of ω, and

it follows from the properties of almost periodic functions that, there will be arbitrarily large values of ω for which the periodogram approaches within any given accuracy of its absolute maximum $N(T)^2/2\pi T$ (Vere-Jones (1982)). So the maxima of the periodogram over too large a range will be unrelated to the periodic effects. The 'average' Nyquist frequency $\pi N(T)/T$ is suggested to be the upper bound in Vere-Jones and Ozaki (1982), while a wider condition is shown in Vere-Jones (1982) that the upper bound only needs to increase sufficiently slowly with T. As for the lower bound, a neighborhood of 0 has to be excluded to remove the peak of the periodogram at the origin and the peaks near the origin caused by leakage, so that the frequency estimates will not converge to 0. We have the following assumption on the search range.

Assumption 6

$$O(T^{\delta'-1}) \leq \omega_k \leq \Omega_T, \quad k = 1, \ldots, K,$$

where $0 < \delta' < 1$ *and* Ω_T *is the upper bound, possibly determined by observations on the process in the interval* $(0, T)$ *with* $E(\Omega_T) = O(T^{1-\delta})$, $\delta > 0$.

There is another way to eliminate the peaks of the periodogram around the origin by using the centralized periodogram

$$I_T^*(\omega) = (2\pi T)^{-1} \Big| \int_0^T e^{-i\omega t} \big[dN(t) - \frac{N(T)}{T} dt \big] \Big|^2, \tag{2.4}$$

which has an asymptotically negligible effect at other frequencies. So if the frequency estimates are determined by (2.4), the lower bound in the search range can be 0. In the following discussion, we use the unmodified periodogram (2.2) to derive the asymptotic results, but all the results hold equally for the centralized periodogram (2.4).

In summary, to estimate the frequencies, we only consider local maxima of the periodogram which occur in the range defined in Assumption 6, and the corresponding frequencies are well separated where their shortest distance cannot decrease faster than $O(T^{-1})$ as $T \to \infty$.

So for a given set of local maxima of the periodogram, to search for the next local maximum, one has to exclude the peaks that occur in the neighborhoods of the frequencies which correspond to the given local maxima, and the width of each neighborhood is $O(T^{-1+\gamma})$ with $\gamma > 0$. Then frequencies corresponding to the largest K such local maxima of the periodogram $I_T(\omega)$ are defined to be the frequency estimates $\hat{\boldsymbol{\omega}}_T = (\hat{\omega}_{1,T}, \ldots, \hat{\omega}_{K,T})$.

Vere-Jones (1982) confirmed that some detailed results for the asymptotic behavior of the periodogram-based frequency estimate in a periodic signal in Gaussian noise (Hannan (1973)) would at least have some counterpart in the point process context by using the decomposition

$$dN(t) = \lambda(t)dt + dZ(t), \tag{2.5}$$

where $\lambda(t)$ is the intensity function, and $dZ(t)$ is a process with mean 0, and independent but non-stationary increments. In addition, $\text{cov}(dZ(t), dZ(\tau)) = \text{var}(dZ(t))\delta_{t,\tau}$ with $\delta_{t,\tau}$ equals 1 if $t = \tau$ and 0 otherwise, and $\text{var}(dZ(t)) = \text{var}(N(t)) = \lambda(t)dt - \lambda^2(t)(dt)^2$. So $dZ(t)$ is mean-squared bounded.

The lemma and its variants in Vere-Jones (1982) play an important role in showing the similarities between results in ordinary time series with Gaussian noises and non-homogeneous Poisson processes. We state a similar result here.

Lemma 7 *Let $N(t)$ be a Poisson process with a bounded intensity function $\lambda(t)$, observed over the time interval $[0, T]$, let Ω_T be a frequency upper bound, determined possibly by observations on the process in the interval $(0, T)$, and set*

$$dZ(t) = dN(t) - \lambda(t)dt.$$

Then if Ω_T satisfies the condition in Assumption 6, for every $m \geq 1$, as $T \to \infty$,

$$T^{-m} \sup_{0 \leq \omega \leq \Omega_T} \left| \int_0^T t^{m-1} e^{-i\omega t} dZ(t) \right| \to 0 \quad (\textit{almost surely}).$$

We refer to Vere-Jones (1982) for the original proof of the case $m = 1$. In the appendix we correct the typo in the original proof and generalize it to the case $m \geq 1$.

Lemma 7 implies that in the decomposition

$$\begin{aligned} J_T(\omega) &:= T^{-m} \int_0^T t^{m-1} e^{-i\omega t} dN(t) \\ &= T^{-m} \int_0^T t^{m-1} e^{-i\omega t} \lambda(t) dt + T^{-m} \int_0^T t^{m-1} e^{-i\omega t} dZ(t) \\ &= J_T^{(\lambda)}(\omega) + J_T^{(Z)}(\omega), \end{aligned} \tag{2.6}$$

$J_T^{(Z)}(\omega)$ converges to 0 uniformly for $\omega \in (0, \Omega_T]$ or, equivalently, $J_T^{(\lambda)}(\omega)$ is the dominant term in $J_T(\omega)$.

We use the above results to establish the asymptotic properties of the parameter estimates, starting with the 'super-efficiency' in the frequency estimates.

Proposition 8 *Under Assumptions 4, 5 and 6, $\hat{\boldsymbol{\omega}}_T$ is a consistent estimate of $\boldsymbol{\omega}$, and*

$$(\hat{\omega}_{k,T} - \omega_k) = o(T^{-1}), \quad (a.s.), \quad k = 1, \ldots, K. \tag{2.7}$$

We leave the proof in the appendix. The result here is similar to the 'supper-efficiency' of frequency estimates in the ordinary time series with convergence rate $o(n^{-1})$ where n is the sample size. One may wonder why frequency estimates require such an accuracy which is not only better than $O(T^{-1/2})$ but also $O(T^{-1})$. Our intuitive answer is that most of the time, the estimation of the amplitude of the sinusoid relies on the frequency estimates, and if the estimated frequency is not within the desired precision of the true frequency, corresponding amplitude estimate is not consistent. In the ordinary time series literature, Rice and Rosenblatt (1988) discussed the use of periodogram on frequency estimation in ordinary time series with a periodic signal in stationary noise sequence. They pointed out that in order to determine the frequency estimate which maximizes the periodogram, one would have to sample the periodogram deterministically by a grid mesh with the mesh length $o(n^{-1})$ where n is the sample size. Otherwise, if the distance between the frequency estimate and the true frequency is greater than

$o(n^{-1})$, the corresponding amplitude and phase estimates are inconsistent. These results extend when there are several harmonic components in the time series. Such results have a counter part in the point process case. We will give a detailed discussion about computational issue on frequency estimation in the framework of our model in section 2.3.

To investigate the asymptotic behavior of the parameter estimates, we define the following random variables

$$\begin{gathered} U := T^{-\frac{1}{2}} \int_0^T dZ(t), \\ V_k := T^{-\frac{1}{2}} \int_0^T \cos(\omega_k t) dZ(t), \quad W_k := T^{-\frac{1}{2}} \int_0^T \sin(\omega_k t) dZ(t) \\ X_k := T^{-\frac{3}{2}} \int_0^T t \cos(\omega_k t) dZ(t), \quad Y_k := T^{-\frac{3}{2}} \int_0^T t \sin(\omega_k t) dZ(t), \\ k = 1, \ldots, K. \end{gathered} \tag{2.8}$$

For simplicity, we use the following notations

$$\delta_{k,k'} = I\{\omega_k = \omega_{k'}\}, \quad \delta_{j,k+k'} = I\{\omega_j = \omega_k + \omega_{k'}\},$$

$$\delta_{j,k-k'} = I\{\omega_j = \omega_k - \omega_{k'}\}, \quad \delta_{j,k'-k} = I\{\omega_j = \omega_{k'} - \omega_k\},$$

where $I\{.\}$ is the indicator function.

From the bounded variance and independent increment property of the process $dZ(t)$, a central limit theorem can be applied in determining the asymptotic distribution of the variables defined in (2.8). The proof is in the Appendix.

Proposition 9 *Under Assumptions 4 and 5,* $(U, V_1, \ldots, V_K, W_1, \ldots, W_K, X_1, \ldots, X_K, Y_1, \ldots, Y_K)'$ *is asymptotically normally distributed as* $T \to \infty$*, with mean* $\mathbf{0}$*, and variance-covariance*

$$\lim_{T\to\infty} \mathit{cov}[(U, V_k, W_k, X_k, Y_k)', (U, V_{k'}, W_{k'}, X_{k'}, Y_{k'})'] =$$

$$\begin{pmatrix} B & (A_{k'}/2)\cos(\phi_{k'}) & -(A_{k'}/2)\sin(\phi_{k'}) & (A_{k'}/4)\cos(\phi_{k'}) & -(A_{k'}/4)\sin(\phi_{k'}) \\ (A_k/2)\cos(\phi_k) & E_1(k,k') & E_3(k,k') & \frac{1}{2}E_1(k,k') & \frac{1}{2}E_3(k,k') \\ -(A_k/2)\sin(\phi_k) & E_3(k',k) & E_2(k,k') & \frac{1}{2}E_3(k',k) & \frac{1}{2}E_2(k,k') \\ (A_k/4)\cos(\phi_k) & \frac{1}{2}E_1(k',k) & \frac{1}{2}E_3(k,k') & \frac{1}{3}E_1(k,k') & \frac{1}{3}E_3(k,k') \\ -(A_k/4)\sin(\phi_k) & \frac{1}{2}E_3(k',k) & \frac{1}{2}E_2(k',k) & \frac{1}{3}E_3(k',k) & \frac{1}{3}E_2(k,k') \end{pmatrix}$$

where

$$E_1(k,k') := \lim_{T\to\infty} \mathit{cov}(V_k, V_{k'}) = \sum_{j=1}^{K} \frac{A_j}{4}\cos(\phi_j)(\delta_{j,k+k'} + \delta_{j,k-k'} + \delta_{j,k'-k}) + \frac{B}{2}\delta_{k,k'},$$

$$E_2(k,k') := \lim_{T\to\infty} \mathit{cov}(W_k, W_{k'}) = \sum_{j=1}^{K} \frac{A_j}{4}\cos(\phi_j)(-\delta_{j,k+k'} + \delta_{j,k-k'} + \delta_{j,k'-k}) + \frac{B}{2}\delta_{k,k'},$$

$$E_3(k,k') := \lim_{T\to\infty} \mathit{cov}(V_k, W_{k'}) = \sum_{j=1}^{K} \frac{A_j}{4}\sin(\phi_j)(-\delta_{j,k+k'} + \delta_{j,k-k'} - \delta_{j,k'-k}),$$

$k, k' = 1, \ldots, K$. *The rows of the matrix correspond to random variables with subscript* k *and the columns correspond to random variables with subscript* k'. *Other covariance in the matrix among* $V_{k'}, W_{k'}, X_k, X_{k'}, Y_k$ *and* $Y_{k'}$ *are given similarly.*

Now we show the asymptotic behavior of $\hat{\omega}_{\mathbf{T}}$.

Since the periodogram is a twice continuously differentiable function of ω_k, and $\hat{\omega}_{k,T}$ is the kth largest local maximum, a Taylor series expansion about $\hat{\omega}_{k,T}$ yields

$$I'_T(\omega_k) = I'_T(\hat{\omega}_{k,T}) + (\omega_k - \hat{\omega}_{k,T})I''_T(\bar{\omega}_{k,T}) = (\omega_k - \hat{\omega}_{k,T})I''_T(\bar{\omega}_{k,T}),$$

where $0 \le |\bar{\omega}_{k,T} - \omega_k| \le |\hat{\omega}_{k,T} - \omega_k|$. So

$$T^{\frac{3}{2}}(\hat{\omega}_{k,T} - \omega_k) = -2\pi T^{-\frac{3}{2}} I'_T(\omega_k)/2\pi T^{-3} I''_T(\bar{\omega}_{k,T}). \tag{2.9}$$

The results in Proposition 9 will help describe the asymptotic behavior of the numerator and denominator of (2.9), starting with the decomposition of $I_T(\omega)$

$$(2\pi T)I_T(\omega) = \Big[\int_0^T \sin(\omega t)dN(t)\Big]^2 + \Big[\int_0^T \cos(\omega t)dN(t)\Big]^2$$
$$= \Big[\int_0^T \sin(\omega t)\lambda(t)dt + \int_0^T \sin(\omega t)dZ(t)\Big]^2 + \Big[\int_0^T \cos(\omega t)\lambda(t)dt + \int_0^T \cos(\omega t)dZ(t)\Big]^2. \tag{2.10}$$

We call the integral Z-integral if it is with respect to $dZ(t)$, and λ-integral if the integral is with respect to $\lambda(t)dt$.

In the first derivative of $(2\pi T)I_T(\omega)$ in (2.10) at $\omega = \omega_k$, the products of λ-integrals in the expansion with $\lambda(t)$ given in (2.3) are $O(T^3)$ plus some $O(T^2)$ terms, but the $O(T^3)$ terms are canceled out eventually, and the products of Z-integrals are $O(T^2)$ in distribution according to Proposition 9. Meanwhile, the products of one λ-integral and one Z-integral are $O(T^{5/2})$ in distribution, so they are the leading terms. Expressing the Z-integrals in terms of the random variables defined in (2.8) and evaluating the λ-integrals by (A.1) in the Appendix, we find

$$2\pi T^{-\frac{3}{2}} I_T'(\omega_k) = A_k\{\frac{1}{2}V_k \sin(\phi_k) + \frac{1}{2}W_k \cos(\phi_k) - X_k \sin(\phi_k) - Y_k \cos(\phi_k)\} + o(1), \quad k = 1, \ldots, K.$$

As for the denominator of (2.9), break it in the similar way. All terms involving Z-integrals uniformly (in ω) converge to 0 (a.s.) by Lemma 7, and the leading terms are the products of λ-integrals. The result $|\bar{\omega}_{k,T} - \omega_k| = o(T^{-1})$ implied by Proposition 8 is needed in evaluating the limiting value of λ-integrals by (A.1) and (A.2). We find that

$$2\pi T^{-3} I_T''(\bar{\omega}_{k,T}) \to -\frac{1}{24}A_k^2, \quad \text{almost surely}, \quad k = 1, \ldots, K. \tag{2.11}$$

So

$$T^{\frac{3}{2}}(\hat{\omega}_{k,T} - \omega_k) = \frac{24}{A_k}\{\frac{1}{2}V_k \sin(\phi_k) + \frac{1}{2}W_k \cos(\phi_k) - X_k \sin(\phi_k) - Y_k \cos(\phi_k)\} + o(1), \quad k = 1, \ldots, K. \tag{2.12}$$

Now we obtain the asymptotic behavior of $\hat{\boldsymbol{\omega}}_T$ which is shown in Theorem 10.

Theorem 10 *Propositions 8 and 9 imply that as* $T \to \infty$, $T^{\frac{3}{2}}(\hat{\boldsymbol{\omega}}_{\boldsymbol{T}} - \boldsymbol{\omega})$ *is asymptotically normally distributed, with mean* $\mathbf{0}$*, and variance-covariance*

$$\lim_{T\to\infty} cov(T^{\frac{3}{2}}(\hat{\omega}_{k,T} - \omega_k), T^{\frac{3}{2}}(\hat{\omega}_{k',T} - \omega'_k)) = \frac{12}{A_k A_{k'}}\Big[2B\delta_{k,k'} + \sum_{j=1}^{K} A_j \times$$

$$\Big(-\cos(\phi_j - \phi_k - \phi_{k'})\delta_{j,k+k'} + \cos(\phi_j - \phi_k + \phi_{k'})\delta_{j,k-k'} + \cos(\phi_j + \phi_k - \phi_{k'})\delta_{j,k'-k}\Big)\Big],$$

where $k, k' = 1, \ldots, K$*. In particular, the variance is*

$$\lim_{T\to\infty} var(T^{\frac{3}{2}}(\hat{\omega}_{k,T} - \omega_k)) = \frac{12}{A_k^2}\Big(2B - \sum_{j=1}^{K} A_j \cos(\phi_j - 2\phi_k)\delta_{j,k+k}\Big).$$

We proceed to define $\hat{A}_{k,T}$ and $\hat{\phi}_{k,T}$ as the estimates of A_k and ϕ_k, respectively. We denote $(i)_k$ and $(ii)_k$ by

$$(i)_k := -T^{-1}\int_0^T \sin(\hat{\omega}_{k,T}t)dN(t), \quad \text{and } (ii)_k := T^{-1}\int_0^T \cos(\hat{\omega}_{k,T}t)dN(t),$$

where $k = 1, \ldots, K$. Then define $\hat{A}_{k,T}$ and $\hat{\phi}_{k,T}$ by

$$\hat{A}^2_{k,T} = 4[(i)_k^2 + (ii)_k^2] = (8\pi/T)I_T(\hat{\omega}_{k,T}), \quad \text{so } \hat{A}_{k,T} = \sqrt{\hat{A}^2_{k,T}},$$

and

$$\tan\hat{\phi}_{k,T} = (i)_k/(ii)_k, \quad \text{if } (ii)_k \neq 0,$$

$$\text{and } \hat{\phi}_{k,T} = \begin{cases} \arctan\tan\hat{\phi}_{k,T}, & \text{if } (ii)_k > 0, \\ \arctan\tan\hat{\phi}_{k,T} + \pi, & \text{if } (ii)_k < 0, \\ \pi/2\,\mathrm{sgn}((i)_k), & \text{if } (ii)_k = 0, \end{cases}$$

where $k = 1, \ldots, K$.

To establish the asymptotic property of $\hat{A}^2_{k,T}$, we use the following representation

$$\begin{aligned} T^{\frac{1}{2}}(\hat{A}^2_{k,T} - A_k^2) &= -(8\pi)T^{-\frac{1}{2}}\Big[I_T(\omega_k) - I_T(\hat{\omega}_{k,T})\Big] + T^{\frac{1}{2}}\Big[(8\pi/T)I_T(\omega_k) - A_k^2\Big] \\ &= -\frac{8\pi}{2}T^{-\frac{1}{2}}(\omega_k - \hat{\omega}_{k,T})^2 I''_T(\bar{\omega}_{k,T}) + 4A_k\{V_k\cos(\phi_k) - W_k\sin(\phi_k)\} + o(1) \\ &= 4A_k\{V_k\cos(\phi_k) - W_k\sin(\phi_k)\} + o(1), \end{aligned} \tag{2.13}$$

where $\bar{\omega}_{k,T}$ in the second last equation satisfies $0 \leq |\bar{\omega}_{k,T} - \omega_k| \leq |\hat{\omega}_{k,T} - \omega_k|$. The first term in the second last equation is obtained by taking Taylor series expansion of $I_T(\omega)$ about $\hat{\omega}_{k,T}$, and since $I_T''(\bar{\omega}_{k,T})$ is $O(T^3)$ by (2.11), and $(\omega_k - \hat{\omega}_{k,T})^2$ is $O(T^{-3})$ in distribution by Theorem 10, this term is $O(T^{-\frac{1}{2}})$ and converges to 0 as $T \to \infty$. The second term in the second last equation is obtained by the decomposition in (2.10) and expressing Z-integral in terms of the random variables defined in (2.8). Following Proposition 9, the last equation in (2.13) implies the asymptotic normality of $T^{\frac{1}{2}}(\hat{A}_{k,T}^2 - A_k^2)$. And δ-method yields the asymptotic normality of $\hat{\boldsymbol{A}}_{\boldsymbol{T}}$.

Turning to $\tan\phi_k$, we first discuss the case when $\phi_k \neq \pi/2$ or $-\pi/2$, then $\cos(\phi_k) \neq 0$. Splitting $(ii)_k$ into the sum of one λ-integral and one Z-integral as in (2.5), Lemma 7 implies that the leading term in $(ii)_k$ is the λ-integral and it is $(A_k/2)\cos(\phi_k) \neq 0$ by (A.1), so $(ii)_k$ is not 0 asymptotically. We break $T^{\frac{1}{2}}(\tan\hat{\phi}_{k,T} - \tan\phi_k)$ as follows,

$$T^{\frac{1}{2}}(\tan\hat{\phi}_{k,T} - \tan\phi_k) = -T^{\frac{1}{2}}\Big[\frac{(i)_k}{(ii)_k} - \frac{T^{-1}\int_0^T \sin(\omega_k t)dN(t)}{T^{-1}\int_0^T \cos(\omega_k t)dN(t)}\Big] + T^{\frac{1}{2}}\Big[-\frac{T^{-1}\int_0^T \sin(\omega_k t)dN(t)}{T^{-1}\int_0^T \cos(\omega_k t)dN(t)} - \tan\phi_k\Big] = (iii)_k + (iv)_k.$$

In addition, in the decomposition of the first derivative of $(i)_k/(ii)_k$ at $\bar{\omega}_{k,T}$ ($0 \leq |\bar{\omega}_{k,T} - \omega_k| \leq |\hat{\omega}_{k,T} - \omega_k|$), the leading term is the products of λ-integrals, and it is $O(T)$ in distribution. By making use of (A.1), a Taylor series expansion of $(iii)_k$ about ω_k yields

$$\begin{aligned}(iii)_k &= -T^{\frac{1}{2}}\Big[\frac{d}{d\omega}\Big(\frac{T^{-1}\int_0^T \sin(\omega t)dN(t)}{T^{-1}\int_0^T \cos(\omega t)dN(t)}\Big)\Big|_{\omega=\bar{\omega}_{k,T}}\Big](\hat{\omega}_{k,T} - \omega_k)\\ &= -T^{\frac{1}{2}}\Big[T\frac{1+o(1)}{2\cos^2(\phi_k + o(1))}\Big](\hat{\omega}_{k,T} - \omega_k)\\ &= -\frac{12}{A_k\cos^2(\phi_k)}\{\frac{1}{2}V_k\sin(\phi_k) + \frac{1}{2}W_k\cos(\phi_k) - X_k\sin(\phi_k) - Y_k\cos(\phi_k)\} + o(1),\end{aligned}$$

where the last equation follows by (2.12). Now we turn to $(iv)_k$ and break integrals in a similar way; we find that

$$(iv)_k = -\frac{2}{A_k\cos^2(\phi_k + o(1))}\{V_k\sin(\phi_k) + W_k\cos(\phi_k)\} + o(1).$$

So

$$T^{\frac{1}{2}}(\tan\hat{\phi}_{k,T}-\tan\phi_k)=-\frac{4}{A_k\cos^2(\phi_k)}\{2V_k\sin(\phi_k)+2W_k\cos(\phi_k)-3X_k\sin(\phi_k)-3Y_k\cos(\phi_k)\}+o(1), \tag{2.14}$$

and we obtain the asymptotic normality of $T^{\frac{1}{2}}(\tan\hat{\phi}_{k,T}-\tan\phi_k)$ by Proposition 9. Again, δ-method yields the asymptotic normality of $\hat{\boldsymbol{\phi}}_{\boldsymbol{T}}$ when $\phi_k\neq\pi/2$ or $-\pi/2$.

Consider the case when $\phi_k=\pi/2$ or $-\pi/2$. It is easy to see that $(ii)_k\to 0$ as $T\to\infty$ and $(i)_k\to(A_k/2)\sin(\phi_k)=(A_k/2)\,\mathrm{sgn}\{\sin(\phi_k)\}$ by Lemma 7 and (A.1). So as $T\to\infty$ and $\phi_k\to\pi/2$, $\hat{\phi}_{k,T}=\arctan\tan\hat{\phi}_{k,T}=\arctan[(i)_k/(ii)_k]\to\pi/2$, and a similar result applies to $\phi_k\to-\pi/2$. So $\hat{\phi}_{k,T}$ is a consistent estimate of ϕ_k when $\phi_k=\pm\pi/2$ and the asymptotic normality of $\hat{\phi}_{k,T}$ obtained above extends to the case when $\phi_k=\pm\pi/2$.

The cross-covariance of $\hat{\boldsymbol{A}}_{\boldsymbol{T}}$ and $\hat{\boldsymbol{\phi}}_{\boldsymbol{T}}$ will be given in Theorem 13.

Theorem 11 *Propositions 8 and 9 imply that, as $T\to\infty$, $T^{\frac{1}{2}}(\hat{\boldsymbol{A}}_{\boldsymbol{T}}-\boldsymbol{A})$ and $T^{\frac{1}{2}}(\hat{\boldsymbol{\phi}}_{\boldsymbol{T}}-\boldsymbol{\phi})$ are asymptotically normally distributed, with mean* $\mathbf{0}$*, and variance-covariance*

$$\lim_{T\to\infty}\mathit{cov}(T^{\frac{1}{2}}(\hat{A}_{k,T}-A_k),T^{\frac{1}{2}}(\hat{A}_{k',T}-A_{k'}))=2B\delta_{k,k'}+\sum_{j=1}^{K}A_j\times$$

$$\Big(\cos(\phi_j-\phi_k-\phi_{k'})\delta_{j,k+k'}+\cos(\phi_j-\phi_k+\phi_{k'})\delta_{j,k-k'}+\cos(\phi_j+\phi_k-\phi_{k'})\delta_{j,k'-k}\Big),$$

$$\lim_{T\to\infty}\mathit{cov}(T^{\frac{1}{2}}(\hat{\phi}_{k,T}-\phi_k),T^{\frac{1}{2}}(\hat{\phi}_{k',T}-\phi_{k'}))=\frac{4}{A_kA_{k'}}\Big[2B\delta_{k,k'}+\sum_{j=1}^{K}A_j\times$$

$$\Big(-\cos(\phi_j-\phi_k-\phi_{k'})\delta_{j,k+k'}+\cos(\phi_j-\phi_k+\phi_{k'})\delta_{j,k-k'}+\cos(\phi_j+\phi_k-\phi_{k'})\delta_{j,k'-k}\Big)\Big].$$

where $k,k'=1,\ldots,K$. In particular, the variances are

$$\lim_{T\to\infty}\mathit{var}(T^{\frac{1}{2}}(\hat{A}_{k,T}-A_k))=2B+\sum_{j=1}^{K}A_j\cos(\phi_j-2\phi_k)\delta_{j,k+k},$$

$$\lim_{T\to\infty}\mathit{var}(T^{\frac{1}{2}}(\hat{\phi}_{k,T}-\phi_k))=\frac{4}{A_k^2}\Big(2B-\sum_{j=1}^{K}A_j\cos(\phi_j-2\phi_k)\delta_{j,k+k}\Big).$$

Now we consider the estimate of B. Since B represents the mean occurrence rate, $N(T)/T$ should be a good estimate. Define $\hat{B}_T$ as

$$\hat{B}_T = N(T)/T.$$

It is shown in the Appendix that $\hat{\lambda}_T(t) = \lambda(t) + O(T^{-\frac{1}{2}})$, so, for large T, $\hat{\lambda}_T(t)$ is asymptotically non-negative. In practice, for a finite sample, the estimate $\hat{B}_T = N(T)/T$ may lead to negative $\hat{\lambda}_T(t)$ at some t, so, if one is willing to make the assumption that $B \geq \sum_{k=1}^{K} A_k$, then $\hat{B}_T = \max(\sum_{k=1}^{K} \hat{A}_{k,T}, N(T)/T)$ is a good choice, and in this case, such an estimate of B is asymptotically equal to $N(T)/T$. Otherwise, this latter estimate of B might have a positive bias. For simplicity, in the simulation study later, we use the latter estimate of B. However, we note that the difference in using either estimate of B is small in the simulation examples. There are other ways to guarantee the non-negativity of the estimated intensity function, such as $\tilde{\lambda}_T(t) = \max(0, \hat{\lambda}_T(t))$.

Theorem 12 *Proposition 9 implies that as $T \to \infty$, $T^{\frac{1}{2}}(\hat{B}_T - B)$ is asymptotically normally distributed, with mean 0, and variance B.*

The correlation of all above estimates are given in the following theorem.

Theorem 13 *It follows from Propositions 9 and 8 that as $T \to \infty$, $T^{\frac{3}{2}}(\hat{\boldsymbol{\omega}}_{\boldsymbol{T}} - \boldsymbol{\omega})$, $T^{\frac{1}{2}}(\hat{\boldsymbol{A}}_{\boldsymbol{T}} - \boldsymbol{A})$, $T^{\frac{1}{2}}(\hat{\boldsymbol{\phi}}_{\boldsymbol{T}} - \boldsymbol{\phi})$, and $T^{\frac{1}{2}}(\hat{B}_T - B)$ are jointly normally distributed, with mean $\mathbf{0}$, and variance-*

covariance

$$\lim_{T\to\infty} cov(T^{\frac{3}{2}}(\hat{\omega}_{k,T}-\omega_k), T^{\frac{1}{2}}(\hat{\phi}_{k',T}-\phi_{k'})) = \frac{6}{A_k A_{k'}}\sum_{j=1}^{K} A_j \times$$

$$\Big(\cos(\phi_j-\phi_k-\phi_{k'})\delta_{j,k+k'} - \cos(\phi_j-\phi_k+\phi_{k'})\delta_{j,k-k'} - \cos(\phi_j+\phi_k-\phi_{k'})\delta_{j,k'-k}\Big),$$

$$\lim_{T\to\infty} cov(T^{\frac{1}{2}}(\hat{A}_{k,T}-A_k), T^{\frac{1}{2}}(\hat{\phi}_{k',T}-\phi_{k'})) = \frac{1}{A_{k'}}\sum_{j=1}^{K} A_j \times$$

$$\Big(\sin(\phi_j-\phi_k-\phi_{k'})\delta_{j,k+k'} - \sin(\phi_j-\phi_k+\phi_{k'})\delta_{j,k-k'} + \sin(\phi_j+\phi_k-\phi_{k'})\delta_{j,k'-k}\Big),$$

$$\lim_{T\to\infty} cov(T^{\frac{1}{2}}(\hat{A}_{k,T}-A_k), T^{\frac{1}{2}}(\hat{B}_T-B)) = B,$$

$$\lim_{T\to\infty} cov(T^{\frac{3}{2}}(\hat{\omega}_{k,T}-\omega_k), T^{\frac{1}{2}}(\hat{A}_{k',T}-A_{k'})) = 0,$$

$$\lim_{T\to\infty} cov(T^{\frac{3}{2}}(\hat{\omega}_{k,T}-\omega_k), T^{\frac{1}{2}}(\hat{B}_T-B)) = 0,$$

$$\lim_{T\to\infty} cov(T^{\frac{1}{2}}(\hat{\phi}_{k,T}-\phi_k), T^{\frac{1}{2}}(\hat{B}_T-B)) = 0.$$

where $k, k' = 1, \ldots, K$.

2.4 Computational issues

Given the similarities between results on frequency estimation in ordinary time series and point processes, one should expect similar concern in using the periodogram as discussed in Rice and Rosenblatt (1988) for ordinary time series. There are many local maxima and minima in the periodogram, and the usual optimization algorithms are not suitable in searching for the K largest local maxima under Assumption 5 unless very good initial values are given and the search range for each local maximum is greatly narrowed down. Proposition 8 suggests that if we wish to determine the K largest local maxima effectively under Assumption 5, we would have to sample the periodogram more finely than $O(T^{-1})$ or, namely, $o(T^{-1})$.

We will first discuss what happens to the amplitude and phase estimates when the frequency estimates are not within $o(T^{-1})$ from the corresponding true frequencies. For simplicity and economy, we assume $K = 1$, and denote the true frequency as ω_0. The following results extend

when $K > 1$.

Suppose that we had obtained an estimate $\tilde{\omega}$ of ω_0, and the estimates of A_0 and ϕ_0 ($\phi_0 \neq \pm\pi/2$) are based on $\tilde{\omega}$, so

$$\tilde{A}^2 = (8\pi/T)I_T(\tilde{\omega}), \quad \text{and } \tan\tilde{\phi} = -T^{-1}\int_0^T \sin(\tilde{\omega}t)dN(t)\Big/T^{-1}\int_0^T \cos(\tilde{\omega}t)dN(t).$$

The discussion of (2.13) implies that if $\tilde{A}^2$ is consistent then $T^{-1}I_T(\tilde{\omega}) - T^{-1}I_T(\omega_0) \to 0$ as $T \to \infty$. Assuming that $|\tilde{\omega} - \omega_0| = cT^{-1}$ for $c \neq 0$, it then follows from Lemma 7 with $m = 1$ that the leading terms in $T^{-1}I_T(\tilde{\omega})$ and $T^{-1}I_T(\omega_0)$ are the λ-integrals. Further calculation by (A.1) shows that

$$T^{-1}I_T(\tilde{\omega}) - T^{-1}I_T(\omega_0) = \frac{A_0^2}{8\pi c^2}(2 - 2\cos(c) - c^2) + o(1). \tag{2.15}$$

Since $2 - 2\cos(c) - c^2 \leq 2 - 2 - c^2 < 0$ for $c \neq 0$, so the leading term on the right-hand side of (2.15) is always negative and is of order $O(1)$ for $c \neq 0$. So $\tilde{A}^2$ is not a consistent estimate of A_0^2. Likewise, since

$$\begin{aligned}
&\frac{T^{-1}\int_0^T \sin(\tilde{\omega}t)dN(t)}{T^{-1}\int_0^T \cos(\tilde{\omega}t)dN(t)} - \frac{T^{-1}\int_0^T \sin(\omega_0 t)dN(t)}{T^{-1}\int_0^T \cos(\omega_0 t)dN(t)} \\
=&\frac{\sin(\phi_0)\sin(c) - \cos(\phi_0)(1-\cos(c)) + o(1)}{\sin(\phi_0)(1-\cos(c)) + \cos(\phi_0)\sin(c) + o(1)} - \frac{\sin(\phi_0) + o(1)}{\cos(\phi_0) + o(1)},
\end{aligned}$$

where the leading term is a cyclic function of c with period 2π, and it takes value 0 only at $c = 0$ and is of order of $O(1)$ elsewhere for $c \in (-\pi, \pi]$, so $\tan\tilde{\phi}$ is not consistent either.

If $|\tilde{\omega} - \omega_0| = cT^{-1+\gamma}$, for $\gamma > 0$, $c \neq 0$, then $T^{-1}I_T(\tilde{\omega})$ is of order of $o(1)$, so $\tilde{A}^2$ is not consistent. As for $\tan\tilde{\phi}$, its numerator and denominator are both of order of $o(1)$, and it may not even have a limiting value as $T \to \infty$, so it is not consistent.

We thus see that if the frequency estimates are not within $o(T^{-1})$ of the corresponding true frequencies, the amplitude and phase estimates are not consistent. Note that the fast Fourier transform cannot be used in calculating the periodogram of a point process because the points $\{t_j : t_j \leq T\}$ are not equally spaced.

So for a given data set, we should search the periodogram on a grid mesh with the mesh length $o(T^{-1})$, such as $2\pi T^{-3/2}$, and determine the initial values corresponding to the K largest ordinates subject to the minimum separation condition, and then do a more refined search in the neighborhood of the initial values. Problems would arise in choosing the minimum separation for a particular set of data and in determining the neighborhood of the initial values in the more refined search. We discuss these two problems as follows.

The choice of minimum separation is $O(T^{-1+\beta})$ with $\beta > 0$. Practically, we may use $O(T^{-1/2})$, but it might be too wide to exclude frequencies with distance larger than $O(T^{-1})$ but smaller than $O(T^{-1/2})$. As indicated in (2.15), if we set $\tilde{\omega} = \omega_k + cT^{-1}$ and $\omega_0 = \omega_k$, then we see that the leading term in $(8\pi/T)I_T(\omega)$ at $\omega = \omega_k + cT^{-1}$ is $A_k^2(2 - 2\cos(c))/c^2$, which is around 0 when $|c| > 6\pi$ and A_k is not too large, so the periodogram $I_T(\omega)$ may go down to the noise level when it is outside $[\omega_k - 6\pi/T, \omega_k + 6\pi/T]$. So in practice, we determine $\hat{\omega}_{1,T}$ by maximizing the periodogram, and may determine $\hat{\omega}_{2,T}$ by maximizing the periodogram outside $[\hat{\omega}_{1,T} - 6\pi/T, \hat{\omega}_{1,T} + 6\pi/T]$, and so on. The suggestion of $6\pi/T$ here is the smallest or the most aggressive choice since we assume that the true frequencies are well separated with minimum distance greater than $O(T^{-1})$. The minimum separation can also be determined by prior knowledge of how far apart the frequencies are. Note that when the amplitude A_k is large and the periodogram has a large dynamic range, the choice of $6\pi/T$ is too small. Since the order of A_k may be unknown for a given data set, it is safer to exclude a neighborhood that is wider than $12\pi/T$. With the above discussion, there is always the issue of detecting a small signal in the midst of a strong signal, i.e., the ratio A_1/A_k is very large with a finite sample size.

As for the second problem regarding the neighborhood of the initial values in the more refined search, since the random fluctuation of $I_T(\omega)$ generated by Z-integrals is at most of order of $O(T^{1/2})$ following a similar discussion to that in the proof of Proposition 9, we need to find the smallest $h > 0$, denoted as h_0, such that for every ω outside $[\omega_k - h, \omega_k + h]$, the

deterministic part of $I_T(\omega)$, namely the λ-integral, drops by $O(T^{1/2+\gamma})$ (for any $\gamma > 0$) from its local maximum at $\omega = \omega_k$, and thus the random fluctuation will not cause a local maximum that is of the same order as $I_T(\hat{\omega}_{k,T})$. In this case, the neighborhood of the initial value in the more refined search can have a width with the same order as h_0.

For every $\omega = \omega_k + h$ ($h \neq 0$), it is easy to show that the leading term in $I'_T(\omega)$ is the products of λ-integrals with order $O(T^2)$, and it is $T^2 A_k^2(-2 + 2\cos(hT) + hT\sin(hT))/2(hT)^3$ which is positive if $-2\pi/T \leq h < 0$ and negative if $0 < h \leq 2\pi/T$. Moreover, the leading term in $I''_T(\omega)$ is the products of λ-integrals with order $O(T^3)$, and it is $T^3 A_k^2(6 - 6\cos(hT) - 4hT\sin(hT) + (hT)^2\cos(hT))/2(hT)^4$ which is negative for $-1/T \leq h \leq 1/T$. So the deterministic part of the periodogram $I_T(\omega)$ has only one maximum at $\omega = \omega_k$ and no local minimum in $[\omega_k - 2\pi/T, \omega_k + 2\pi/T]$. Therefore the neighborhood of the initial value for $\hat{\omega}_{k,T}$, [(*initial value*) $- 2\pi/T$, (*initial value*) $+ 2\pi/T$], must contain the true frequency ω_k since this initial value for $\hat{\omega}_{k,T}$ is resolved with order $o(T^{-1})$. In addition, the λ-integral part or the deterministic part of $I_T(\omega)$ is $TA_k^2(2 - 2\cos(hT))/(8\pi h^2 T^2)$, which drops by

$$\begin{aligned}\frac{A_k^2}{8\pi}T\Big[1 - \frac{2 - 2\cos(hT)}{(hT)^2}\Big] &= \frac{A_k^2}{8\pi}T\frac{(hT)^2 - 2 + 2[1 - \frac{(hT)^2}{2!} + \sum_{n=2}^{\infty}(-1)^n\frac{(hT)^{2n}}{(2n)!}]}{(hT)^2} \\ &= \frac{A_k^2}{8\pi}T\Big[\frac{1}{12}(hT)^2 + \sum_{n=3}^{\infty}(-1)^n\frac{(hT)^{2n-2}}{(2n)!}\Big],\end{aligned}$$

from its local maximum obtained at $\omega = \omega_k$ with value $A_k^2 T/(8\pi)$. Setting this equation to $O(T^{1/2})$, we have $(hT)^2 = O(T^{-1/2})$, so $h = O(T^{-5/4})$. Any further departure from ω_k than $O(T^{-5/4})$ will bring down the deterministic part of $I_T(\omega)$ by more than $O(T^{1/2})$, so h_0 mentioned above is $O(T^{-5/4})$. And the search range in the more refined search for $\hat{\omega}_{k,T}$ is (*initial value*) $\pm\, O(T^{-5/4})$. In practice, we may use (*initial value*) $\pm\, T^{-5/4}\log T$ as a conservative choice for the more refined search range.

2.5 Prediction

In this section, we investigate the problem of predicting the next occurrence conditional on the current observations. The number of observations $N(T) = n$ is given in the condition.

Denote

$$\Lambda(t) = \int_0^t \lambda(s)ds.$$

It is easy to see that $\Lambda(t)$ is a non-decreasing function. The following theorem provides the point forecast and its standard error for the next occurrence.

Theorem 14 *Under Assumption 4, the one-step prediction $\hat{T}_{n+1}$ is defined to minimize the mean-squared error (MSE), and is given by*

$$\begin{aligned} \hat{T}_{n+1} &= E(T_{n+1}|T_n = t_n, \ldots, T_1 = t_1) \\ &= t_n + \int_{t_n}^{\infty} e^{-\Lambda(s)+\Lambda(t_n)}ds, \quad n \geq 1, \end{aligned} \tag{2.16}$$

and its MSE ν_n without considering the uncertainty brought by the parameter estimates is given by

$$\begin{aligned} \nu_n &= E(T_{n+1} - \hat{T}_{n+1})^2 \\ &= E_{T_n}\Big[\int_{T_n}^{\infty} 2(s - T_n)e^{-\Lambda(s)+\Lambda(T_n)}ds - \Big\{\int_{T_n}^{\infty} e^{-\Lambda(s)+\Lambda(T_n)}ds\Big\}^2\Big], \quad n \geq 1. \end{aligned} \tag{2.17}$$

Since $N(t)$ follows a Poisson distribution with mean $\int_0^t \lambda(s)ds = \Lambda(t)$, the marginal distribution of T_n is

$$\begin{aligned} f_{T_n}(t_n) &= -\frac{\partial}{\partial t_n}P(T_n > t_n) = -\frac{\partial}{\partial t_n}P(N(t_n) < n) \\ &= -\frac{\partial}{\partial t_n}\sum_{j=0}^{n-1}\frac{[\Lambda(t_n)]^j}{j!}e^{-\Lambda(t_n)} \\ &= \frac{[\Lambda(t_n)]^{n-1}}{(n-1)!}\lambda(t_n)e^{-\Lambda(t_n)}, \quad t_n > 0, \end{aligned}$$

so $\Lambda(T_n)$ follows a Gamma distribution with $\alpha = n$, $\beta = 1$.

The calculation of (2.17) is carried out by Monte Carlo integration with the following two steps.

1. Generate a large sample of $\Psi_n = \Lambda(T_n) \sim$ Gamma($\alpha = n$, $\beta = 1$). Solve for T_n, namely, t_n as the root of $\Lambda(t_n) - \psi_n = 0$.

2. Calculate $2\int_{t_n}^{\infty}(s - t_n)e^{-\Lambda(s)+\Lambda(t_n)}ds - [\int_{t_n}^{\infty} e^{-\Lambda(s)+\Lambda(t_n)}ds]^2$, and average over t_n. We obtain ν_n.

Suppose we predict T_{n+1} by a homogeneous Poisson process with $\lambda = B$ whereas the true model is given by Assumption 4. Such a prediction of T_{n+1} is denoted by $\tilde{T}_{n+1}$ and is given by $1/B$. The improvement in the prediction in terms of the reduction of the MSE is

$$\mu_n := \mathrm{E}(T_{n+1} - \tilde{T}_{n+1})^2 - \mathrm{E}(T_{n+1} - \hat{T}_{n+1})^2 = \mathrm{E}_{T_n}\Big[\int_{t_n}^{\infty} e^{-\Lambda(s)+\Lambda(t_n)}ds - \frac{1}{B}\Big]^2,$$

and it can be calculated by Monte Carlo integration with procedures similar to those discussed before.

It is easy to see that the improvement μ_n is 0 if and only if all A_k's are zero, namely, the true model is a homogeneous Poisson process.

We note that with the functional form of the intensity function, one can also easily obtain the prediction bounds, i.e., the interval which includes the arrival of the next event with certain probability. For example, a $100(1-\alpha)\%$ upper prediction bound a is defined by $\Pr\{T_{n+1} < a | T_n = t_n, \ldots, T_1 = t_1\} = 1 - \alpha$, where $a > t_n$ and $100(1-\alpha)\%$ is the confidence level . The conditional distribution of T_{n+1} given the past events is given by (A.4) in the Appendix, and by using (A.4) we can obtain a as the solution to $\Lambda(a) = \Lambda(t_n) - \log(\alpha)$.

2.6 Simulation study

In this section, we show the simulation results in evaluating the performance of our model. We consider four different periodic or almost periodic intensity functions of non-homogeneous

Poisson processes. They are respectively,

$$\text{Case 1:} \quad \lambda(t) = 1.6 + \cos\Big(\frac{\pi}{4\sqrt{3}}t\Big) + 0.5\cos\Big(\frac{\pi}{3\sqrt{2}}t + \frac{\pi}{4}\Big),$$

$$\text{Case 2:} \quad \lambda(t) = \sqrt{3.1 + 3\cos\Big(\frac{\pi}{3\sqrt{2}}t\Big)},$$

$$\text{Case 3:} \quad \lambda(t) = 0.1 + 0.5\text{Mod}[t, 2\pi],$$

$$\text{Case 4:} \quad \lambda(t) = 1.3\exp\Big\{\cos\Big(\frac{\pi}{3\sqrt{2}}t + \frac{\pi}{4}\Big)\Big\}.$$

The intensity function in Case 1 is almost periodic but not periodic since the ratio of the two frequencies $\pi/4\sqrt{3}$ and $\pi/3\sqrt{2}$ is not rational, and the function never repeats itself exactly. The intensity function in Case 2, 3 and 4 are periodic functions, and they are not, but can be approximated by the sum of sinusoidal functions. In particular, Case 4 has the same function form as (2.1) which has been discussed in Lewis (1970, 1972) and Vere-Jones (1982) and so on.

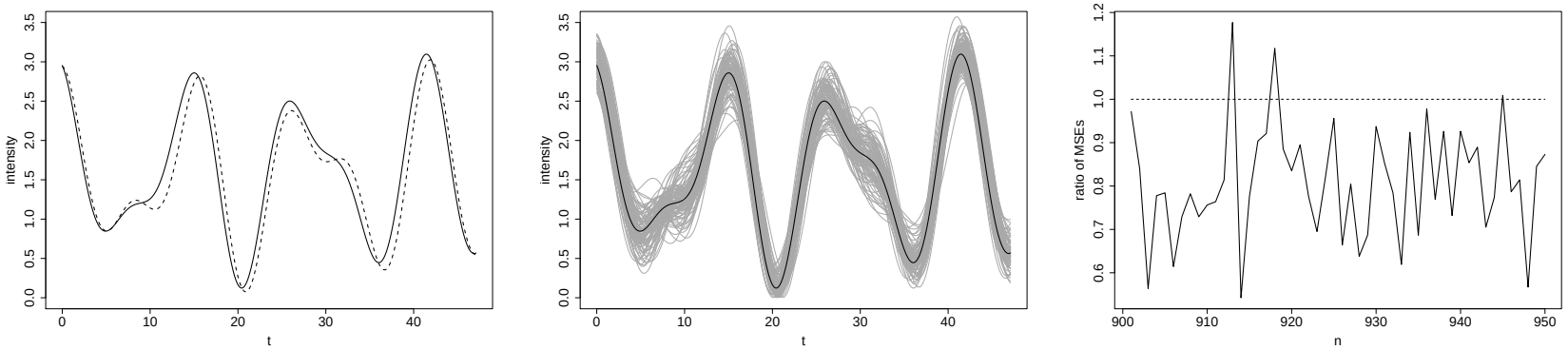
We generate the non-homogeneous Poisson processes accordingly, and there are 100 independent replicates in each case with the observation length $T = 500$, and approximate 700 to 900 data points are used for the estimation in each replicate. We estimate the intensity function in the framework of our model, and compare the estimated intensity function with the true intensity function. In each of the following figures, plot (a) displays the true intensity function (solid line) and one estimated intensity function from a single replicate (dashed line), and plot (b) displays the true intensity function (dark solid line) and 100 estimated intensity function from 100 replicates (light solid lines), and it gives a rough idea of the distribution of the estimated intensity function at time t. Figure 2.1 presents Case 1. The fitting is quite good because the intensity function takes the same function form as our model. Figure 2.2, 2.3 and 2.4 present Case 2, 3 and 4, respectively. We take two sinusoidal terms in our model ($K = 2$) to estimate the intensity function in Case 2, and the estimated intensity function captures the shape precisely with steep troughs and relatively flat peaks. Case 3 has a discontinuous intensity function which is usually difficult to estimate; but our model with $K = 3$ sinusoidal terms captures the sharp

jumps at the discontinuous points and the general shape of the true intensity function. Case 4 takes the same functional form as (2.1), and our model with $K = 2$ sinusoidal terms does a good approximation.

We also conduct the 'out-of-sample' one-step-ahead prediction using the estimated intensity function, and compare the MSE with the MSE under the homogeneous Poisson process model by taking their ratio, namely $\frac{1}{100}\sum_{i=1}^{100}(t_{n+1}^i - \hat{t}_{n+1}^i)^2 / \frac{1}{100}\sum_{i=1}^{100}(t_{n+1}^i - \tilde{t}_{n+1}^i)^2$, where t_{n+1}^i is the $(n+1)$th time point in the ith replicate, and $\hat{t}_{n+1}^i$ and $\tilde{t}_{n+1}^i$ are the 'out-of-sample' one-step-ahead prediction of t_{n+1}^i under our model and under homogeneous Poisson process model respectively. Here the intensity function of the homogeneous Poisson process model is estimated by $N(T)/T$. Plot (c) in each figure shows the ratio, and the prediction is carried out for the $901st$ to $950th$ data points ($901 \leq n+1 \leq 950$); the dashed line represents ratios equal to 1. Asymptotically, the ratio is expected to be below 1 for every $n \geq 1$. The few peaks in the ratio plots which are greater than 1 are caused by random variation of the processes. On average, the reduction in MSE by using our model is 19.1% in Case 1, 11.2% in Case 2, 9.6% in Case 3 and 20.7% in Case 4.

Notice that the choice of K is subjective in above simulations, and it will be discussed in Chapter 3. Overall, the simulation demonstrates that our model is very general and it can capture most variations in the intensity function of any periodic or aperiodic but almost periodic non-homogeneous Poisson process.

We also calculate the sample means and standard errors of the parameter estimates from the 100 replicates in Case 1 and show them in Table 2.1. This is to verify the theoretical results in section 2.3. We see that the sample means are close to the true parameter values, and the sample standard deviations are close to the asymptotic standard deviations and so are the sample

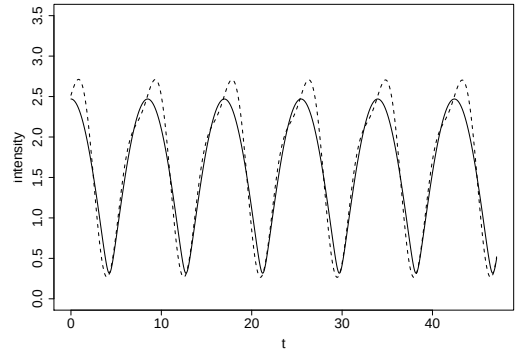


(a) Solid line: true intensity function. Dashed line: estimated intensity function from one replicate.

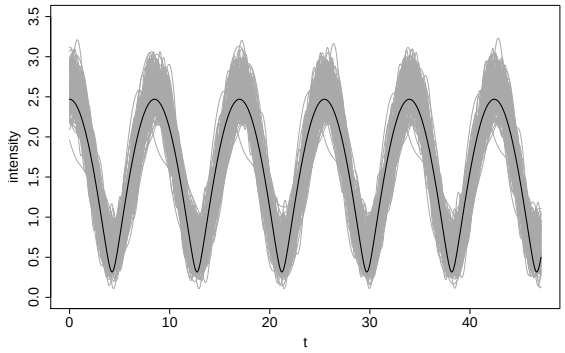
(b) Dark solid line: true intensity function. Light solid lines: estimated intensity function from 100 replicates.

(c) Ratio of MSE under our model and MSE under the homogeneous Poisson process model.

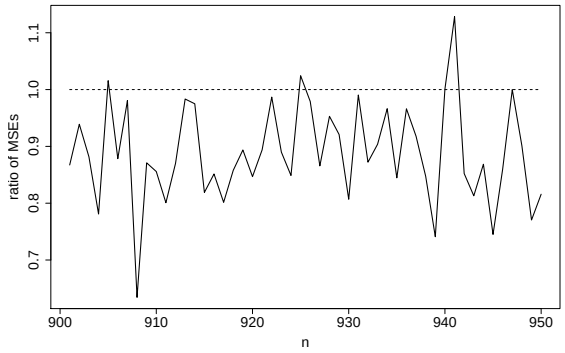
Figure 2.1: Case 1. The number of data points used for estimation in 100 replicates ranges from 717 to 866. The observation length $T = 500$. The 'out-of-sample' one-step-ahead prediction is carried out for the $901st$ to $950th$ data points.



(a) Solid line: true intensity function. Dashed line: estimated intensity function from one replicate.

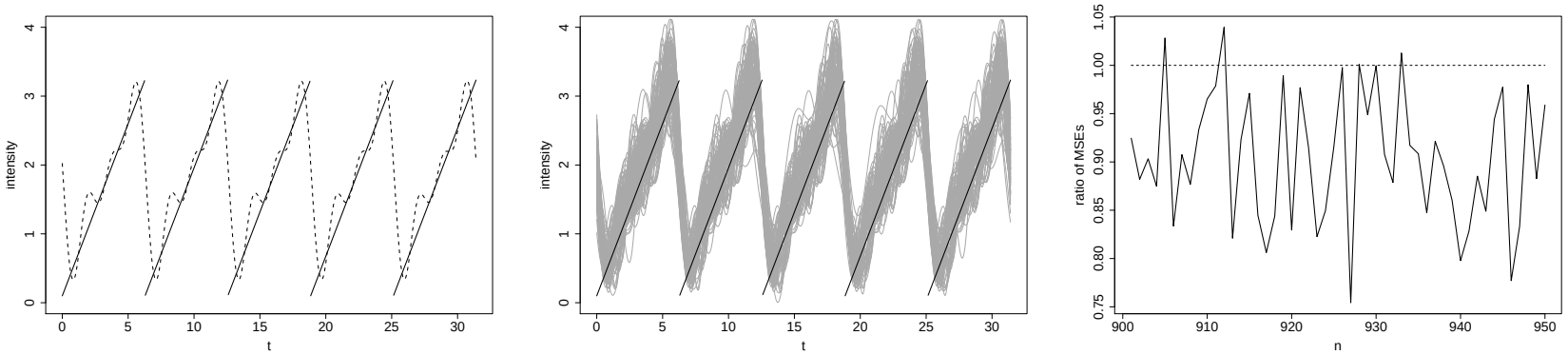


(b) Dark solid line: true intensity function. Light solid line: estimated intensity function from 100 replicates.



(c) Ratio of MSE under our model and MSE under the homogeneous Poisson process model.

Figure 2.2: Case 2. The number of data points used for estimation in 100 replicates ranges from 749 to 872. The observation length $T = 500$. The 'out-of-sample' one-step-ahead prediction is carried out for the $901st$ to $950th$ data points.

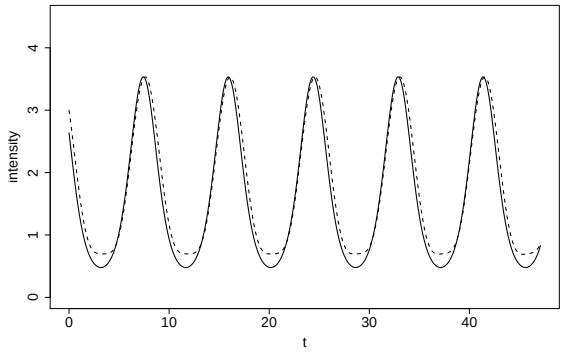


(a) Solid line: true intensity function. Dashed line: estimated intensity function from one replicate.

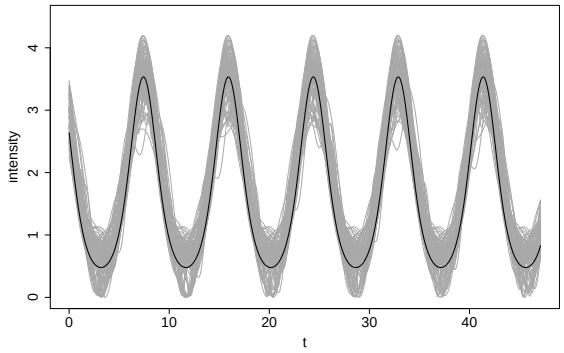
(b) Dark solid line: true intensity function. Light solid line: estimated intensity function from 100 replicates.

(c) Ratio of MSE under our model and MSE under the homogeneous Poisson process model.

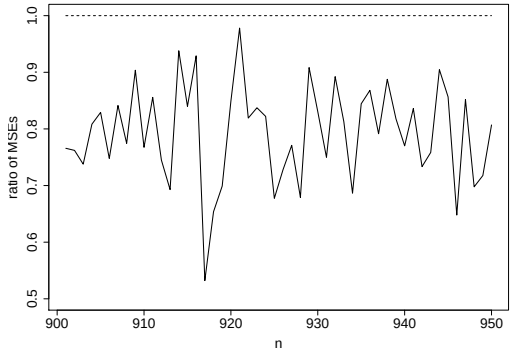
Figure 2.3: Case 3. The number of data points used for estimation in 100 replicates ranges from 733 to 906. The observation length $T = 500$. The 'out-of-sample' one-step-ahead prediction is carried out for the $901st$ to $950th$ data points.



(a) Solid line: true intensity function. Dashed line: estimated intensity function from one replicate.



(b) Dark solid line: true intensity function. Light solid line: estimated intensity function from 100 replicates.



(c) Ratio of MSE under our model and MSE under the homogeneous Poisson process model.

Figure 2.4: Case 4. The number of data points used for estimation in 100 replicates ranges from 754 to 879. The observation length $T = 500$. The 'out-of-sample' one-step-ahead prediction is carried out for the $901st$ to $950th$ data points.

covariances.

Table 2.1: The means and standard errors of the parameter estimates from the 100 replicates in Case 1

	ω_1	ω_2	A_1	A_2	ϕ_1	ϕ_2	B
true value	0.45345	0.74048	1	0.5	0	0.78540	1.6
sample mean	0.45374	0.74116	1.01223	0.50671	-0.07806	0.59168	1.60616
asymptotic sd	0.00055	0.00111	0.08000	0.08000	0.16000	0.32000	0.05657
sample sd	0.00052	0.00112	0.07507	0.07838	0.16499	0.33445	0.05613

2.7 IBM stock transaction example

Stock transactions are unequally spaced and the occurring time of the transactions is a point process. Since there are usually many transactions during the opening and closing time and much less during lunch time, and this pattern repeats every day, there must be a day-effect in the stock transaction process. We studied the times of occurrence of IBM stock transaction over a 4 week period beginning November 26, 1990 and ending December 21, 1990 (measured in second)[1]. The NYSE opens at 9:30am and closes at 4:00pm but there are 135 transactions occurred after 4:00pm and no transaction before 9:30am. These observations were deleted from the analysis and there are 17077 data points in the analysis. The overnight waiting time was ignored and the observation length is 468000 seconds or 20 days with 6.5 hours each day. There are 1401 distinct transaction time with multiple transactions, and 1264 of them have 2 simulta-

[1]The data is available from the NYSE as the TORQ database which stands for Trades, Orders Reports, and Quotes. The TORQ database is distributed freely for academic study. The IBM data we used here is part of the data set which has been used in Engle (2000), and it also can be downloaded at http://faculty.chicagobooth.edu/ruey.tsay/teaching/fts2/ibm.txt.

neous transactions with possibly different volume. For such transactions which occurred at the same time, we assume that it is because the precision of the recording is 1 second and the actual transaction time is the recorded time plus some infinitesimal time. So for simultaneous transaction time we replace them (while keeping the first transaction time) by adding 0.05 second to the previous transaction time. In this way, we do not lose much information in the counting process and approximately keep the same intensity function of the point process.

Since our observation length is 4 weeks, we assume the largest observable long trend period is not longer than 2 weeks or 234000 seconds. And we assume that the shortest financial cycle in the transaction process is no less than one hour or 3600 seconds, so the search range for periodicity is $[2\pi/234000, 2\pi/3600]$. The intensity function of the stock transaction process is estimated by using our model with $K = 5$, and the centralized periodogram (2.4) is used in finding the estimates of the parameters. Figure 2.5 displays the centralized periodogram in the search range $[2\pi/234000, 2\pi/3600]$, and the peaks corresponding to the frequency estimates $\hat{\omega}_1, \ldots, \hat{\omega}_5$ are marked by the solid dots. The peak to the left of the first dot is not included because it is too close to the first dot with distance less than $6\pi/T$. If the observation length were longer than 4 weeks and more data points were included in the analysis, the peak to the left of the first dot should stand out and should be considered in estimating the frequency $\boldsymbol{\omega}$. The estimates of the parameters are given by

$$\begin{aligned}
\boldsymbol{\omega} &= (\omega_1, \omega_2, \omega_3, \omega_4, \omega_5) = 10^{-4}(2.6775, 0.5772, 8.0805, 1.2024, 5.3790), \\
\boldsymbol{A} &= (A_1, A_2, A_3, A_4, A_5) = 10^{-2}(1.043, 0.667, 0.489, 0.442, 0.438), \\
\boldsymbol{\phi} &= (\phi_1, \phi_2, \phi_3, \phi_4, \phi_5) = (-0.0209, 0.2773, -0.9046, 0.9588, 0.2861),
\end{aligned} \tag{2.18}$$

and $B = 0.03649$.

The periods corresponding to the frequencies are calculated by $2\pi/\omega$, and they are 6.52 hours, 30.24 hours or almost one week (4.65 days), 2.16 hours, 14.52 hours and 3.24 hours

respectively. So ω_1 represents the day-effect and ω_2 represents the week-effect in the stock transaction process. In addition ω_3 and ω_5 are multiples of ω_1, and ω_4 is almost twice of ω_2.

We also estimate the intensity function by using model (2.1) in Lewis (1970, 1972) and Vere-Jones (1982). This model can capture only one single period, namely, the day-effect in the transaction process. The estimated intensity function by using this model is

$$\lambda(t) = 0.03574 \exp\{0.2888 \cos(0.00026775t - 0.0209)\},$$

where the frequency and phase estimates are the same as ω_1 and ϕ_1 in (2.18).

Non-parametric smoothing was used to provide a general idea of the shape of the intensity function in the transaction process for IBM. For a given time t, the non-parametric estimate of the intensity function was given by $\sum_{i=1}^{N(T)} g((t - t_i)/h)/h$, where t_i are the data points and $g(.)$ is the kernel function. We use standardized normal density for the kernel function, and the bin width h is taken to be 20 minutes. Figure 2.6 displays the non-parametric estimate of the intensity function, the estimated intensity function by using our model (2.3), and the estimated intensity function by using model (2.1) in four weeks. Comparing with the non-parametric estimate of the intensity function, our model (2.3) captures the variation in the intensity function very well except two spikes on Thursday in week 2, but model (2.1) does a poorer job and it cannot capture the difference between weeks.

To obtain a quantitative comparison for the predictions of our model (2.3), model (2.1), and the homogeneous Poisson model, we computed the 'out-of-sample' one-step-ahead prediction squared errors for the day right after the 4-week observation period, namely, December 24, 1990. There are 324 transactions on this day so we have 324 squared errors under each model. The averaged squared errors are used for the comparison. They are 1.09240, 1.16546 and 1.20233 (min^2) respectively, and the reductions in the averaged squared error compared to the homogeneous Poisson model are 9.14% and 3.07% for our model (2.3) and model (2.1),

respectively.

From this example, we see that our model (2.3) performed better in both capturing the periodicity of the process and prediction than model (2.1) which seems to be the only widely used parametric model for periodic non-homogeneous Poisson process.

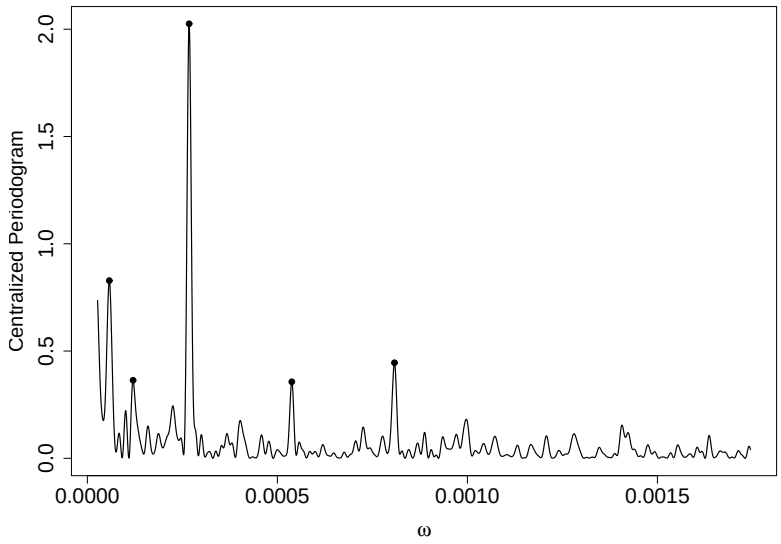
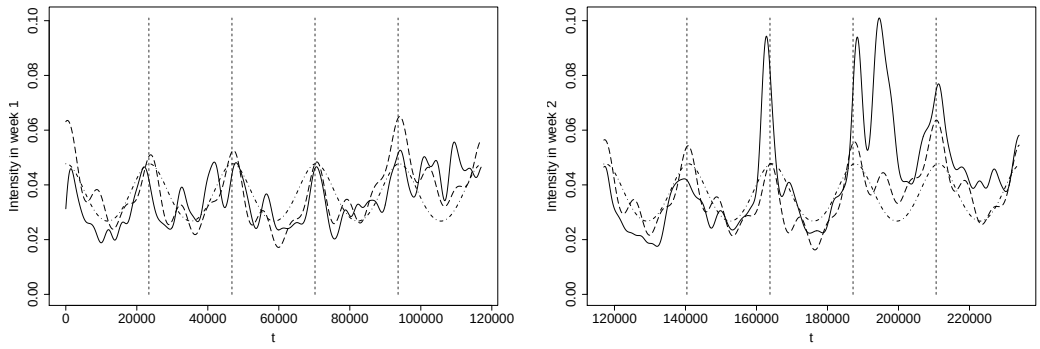


Figure 2.5: Solid line: the centralized periodogram of IBM transaction time from November 26, 1990 to December 21, 1990. The five solid dots are the peaks corresponding to five periodic components in the model.

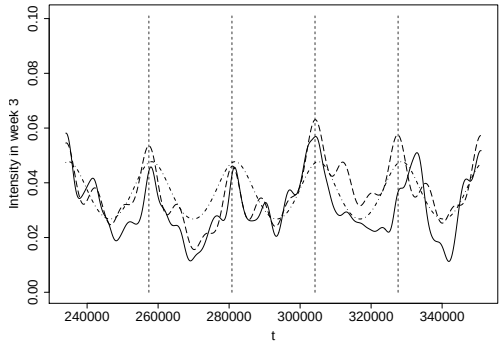
2.8 Discussion and Conclusion

In this chapter we have proposed a very general model for the rate of a non-homogeneous Poisson process which can be almost periodic or periodic in any shape. We have illustrated the usefulness of the proposed model in both simulation and real data example, and the real data example demonstrated that our model performed much better than the existing model in modeling

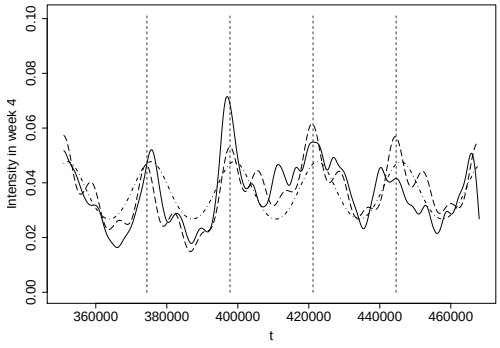


(a) Week 1

(b) Week 2



(c) Week 3



(d) Week 4

Figure 2.6: Solid line: the non-parametric estimate of the intensity function. Long dashed line ($- - -$): the estimated intensity function using our model (2.3). Dotdashed lines ($\cdot - \cdot -$): the estimated intensity function using model (2.1). The vertical light dashed lines separate days in each week.

periodic non-homogeneous Poisson process. The periodogram is used in the model parameter estimation, and a detailed discussion on frequency estimation is provided. Other methods such as maximum likelihood estimation can also be considered but may be more computationally intensive. The parameter estimates proposed in this paper can be used as very good initial values in other estimation method which requires iteration, such as MLE. The selection of K, the number of periodic components in the model, will be discussed in Chapter 3.

Chapter 3

Determining K by Model Selection Criteria

3.1 Introduction

In this chapter, we propose to use model selection criteria to determine K, the number of sinusoidal components in the almost periodic intensity function of a non-homogeneous Poisson process.

There are mainly two classes of model selection criteria (Burnham and Anderson (2002)): (I) criteria that are estimates of Kullback-Leibler information or distance, such as Akaike information criterion (AIC), the bias corrected version of AIC when sample size is small (AICC) (Hurvich and Tsai (1989)), and a generalized version of AIC derived by Takeuchi (1976) (TIC); (II) criteria which are consistent estimators of the dimension of the model, such as Bayesian information criterion (BIC) derived by Schwarz (1978) and CAICF proposed by Bozdogan (1987) (C denoting "consistent" and F denoting the use of the Fisher information matrix). The criteria in the first class is usually not consistent, but they assume that the true data generating model

is usually too complex to formulate, and one could only approximate the true model from a given set of candidate models. For a given set of data and the sample size, such criteria select the best finite-dimensional approximating model which has the shortest "information distance" from the true model. As for the second class of model selection criteria, they are based on the assumptions that a "true model" exists, and the model selection goal is to select the true model, and the probability of selecting this true model approaches 1 as sample size increases and thus the model selection criteria are consistent.

In the biological and social sciences and medical sciences, Burnham and Anderson (2002) argue that the first class of model selection criteria (AIC-type) are reasonable for the analysis of empirical data since in these fields, the increased sample size may stem from the addition of new geographic field sites or laboratories, and thus the number of factors in the model may also increase. In this case, the data-generating model may not remain fixed as the sample size increases. As for the consistent criteria, they may be useful in some physical sciences where a true model might exist and remain the same as sample size increases. Both classes of criteria have good interpretation on the model they select: the model being selected is either the "best" model to approximate the true model in terms of "information distance" or the model which is identified as true model with an asymptotic probability of 1 if one believes that the true model belongs to the set of candidate models.

The selection of K is similar to the selection of the order of ARMA process in time series: there is no unique and "best" or "correct" selection method and practitioners usually consider multiple model selection criteria. To fit the ARMA model to a time series data, after narrowing down the model candidates to a few by the model selection criteria, we also exam the whiteness of the residuals from different models, and/or exam the predictability of the model to determine the best model for the data. Analogous to the time series analysis, we propose to consider different model selection criteria to produce the candidates for K in the framework of our model,

and then exam the model fitting and select the best K. In this chapter, we will restrict our attention to AIC and BIC since they are the representatives of the two classes of model selection criteria.

We notice that in the almost periodic Poisson process, $T \to \infty$ is used in the asymptotic derivation instead of $n \to \infty$ as in the usual random sample setting where T is the observation length, and the convergence rate of the MLE of the frequency is 'super-efficient' which is similar to the periodogram estimate, so the AIC and BIC need some modification in the framework of our model.

In this chapter, the maximum likelihood estimate is denoted as $\hat{\theta}$, somewhat similar to the notation of the periodogram estimates in chapter 2. But we do not discuss the periodogram estimates in this chapter, so it should be easy for the readers to distinguish the notation of two estimates. In addition, the true parameter is denoted as θ_0.

The rest of this chapter is organized as follows. We give an overview of AIC and BIC in section 3.2. Then we discuss the likelihood function and the MLE of the almost periodic Poisson process and derive the AIC and BIC in the framework of our model in section 3.3. In section 3.4, we provide procedures on how to implement the proposed methodology. In section 3.5 and 3.6, we show the simulation results and re-analyze the IBM data used in Chapter 2.

3.2 An overview of AIC and BIC

3.2.1 Akaike information criterion

We give a review of Akaike information criterion (AIC) in this section based on the original paper of Akaike (1973) and the book by Burnham and Anderson (2002). We refer to Burnham and Anderson (2002) for more detailed discussion on the information-based model selection criteria and their applications.

Akaike (1973) proposed the use of Kullback-Leibler (K-L) distance as a fundamental basis for model selection, where the K-L distance between the models f and g is defined for continuous functions (discrete case can be defined similarly) as the integral (usually multidimensional)

$$\begin{aligned} I(f,g) &= \int f(x)\log\Big(\frac{f(x)}{g(x|\theta)}\Big)dx \\ &= \int f(x)\log(f(x))dx - \int f(x)\log(g(x|\theta))dx \\ &= \int f(x)\log(f(x))dx - \mathrm{E}_x[\log(g(x|\theta))]. \end{aligned} \tag{3.1}$$

The K-L distance relates to the "information" loss when g is used to approximate f. Here f is the true model which reflects the complex measuring or sampling process that generated the observed data, and it might not even have parameters in a sense that would be analogous to θ in a modeling framework. Sometimes it is useful to think of f as full reality and let it have conceptually an infinite number of parameters. On the other hand, $g(x|\theta)$ is a class of models that are used to approximate the truth; and among this class of models, $g(x|\theta_0)$ minimizes the K-L distance, and hence is the best model selected. In the last line of (3.1), we see that $\int f(x)\log(f(x))dx$ is a constant (although unknown), so to minimize K-L distance is equivalent to minimize $-\mathrm{E}_x[\log(g(x|\theta))]$, i.e.the relative K-L distance, over the candidate models.

However, even the relative K-L distance cannot be computed without full knowledge of both f and the parameters θ in each of the candidate models $g_i(x|\theta)$. And in practice, the model parameters must be estimated, and the difference between having θ or θ_0 and having the estimate $\hat{\theta}$ affects how we must use K-L distance as a basis for model selection. And that basically causes us to change our model selection criterion to that of minimizing relative *expected* estimated K-L distance rather than minimizing known (relative) K-L distance over the set of models considered. That is, we select the model g to minimize

$$-\mathrm{E}_y\mathrm{E}_x[\log(g(x|\hat{\theta}(y)))], \tag{3.2}$$

where x and y are independent random samples from the same distribution and $\hat{\theta}(y)$ is the

maximum likelihood estimate (MLE) of θ based on the sample y. In addition, both statistical expectations are taken with respect to truth f. So now the critical issue for getting an applied K-L model selection criterion was to estimate (3.2). Akaike (1973) found out that under certain conditions, $-\log(\mathcal{L}(\hat{\theta}|data)) + p$ is an asymptotically unbiased estimate of the selection target (3.2) where $\log(\mathcal{L}(\hat{\theta}|data))$ is the maximized log-likelihood for model g given the data and p is the number of parameters in the model g. And thus the AIC criterion is given by multiplying $-\log(\mathcal{L}(\hat{\theta}|data)) + p$ by 2 (for some historical reason) as follows:

$$\textbf{AIC} = -2\log(\mathcal{L}(\hat{\theta}|data)) + 2p.$$

The model which gives the smallest AIC value is the best model to approximate the truth f among the candidate models.

We now show how to get the unbiased estimate of (3.2). This is the critical part which connects the relative expected K-L distance and the maximized log-likelihood.

The first step is the Taylor series expansion of $\log(g(x|\hat{\theta}(y)))$ around θ_0 for any given x. Denote $\hat{\theta}(y)$ as $\hat{\theta}$. Since x and y (vectors) are independent random samples, then $\hat{\theta}$ is independent of x. Note that both x and y are of size n, and θ is a $p-$dimensional vector.

$$\begin{aligned}\log(g(x|\hat{\theta})) \approx & \log(g(x|\theta_0)) + \Big[\frac{\partial \log(g(x|\theta_0))}{\partial \theta}\Big]'[\hat{\theta} - \theta_0] \\ & + \frac{1}{2}[\hat{\theta} - \theta_0]'\Big[\frac{\partial^2 \log(g(x|\theta_0))}{\partial \theta^2}\Big][\hat{\theta} - \theta_0].\end{aligned} \tag{3.3}$$

Truncation at the quadratic term entails an unknown degree of approximation, but it is an error of approximation that goes to zero as $n \to \infty$. Here $\partial^2 \log(g(x|\theta_0))/\partial\theta^2$ is a $p \times p$ matrix. For simplicity, we shall write the matrix $\theta\theta'$ as θ^2.

Take the expected value of (3.3) with respect to x:

$$\begin{aligned}\mathrm{E}_x\big[\log(g(x|\hat{\theta}))\big] \approx & \mathrm{E}_x[\log(g(x|\theta_0))] + \mathrm{E}_x\Big[\frac{\partial \log(g(x|\theta_0))}{\partial \theta}\Big]'[\hat{\theta} - \theta_0] \\ & - \frac{1}{2}[\hat{\theta} - \theta_0]'\Big[-\mathrm{E}_x\frac{\partial^2 \log(g(x|\theta_0))}{\partial \theta^2}\Big][\hat{\theta} - \theta_0].\end{aligned} \tag{3.4}$$

The linear term above vanishes, because $\mathrm{E}_x\left[\frac{\partial \log(g(x|\theta_0))}{\partial \theta}\right] = 0$. The proof is as follows:

Since θ_0 is the solution to $\min_\theta I(f,g) = \int f(x)\log\left(\frac{f(x)}{g(x|\theta)}\right)dx$, then

$$\frac{\partial}{\partial \theta}\int f(x)\log\left(\frac{f(x)}{g(x|\theta)}\right)\Big|_{\theta=\theta_0} dx = 0,$$

that is

$$\frac{\partial}{\partial \theta}\int f(x)\log(f(x))dx = \frac{\partial}{\partial \theta}\int f(x)\log(g(x|\theta))\big|_{\theta=\theta_0} dx, \tag{3.5}$$

Because θ is not involved in $f(\cdot)$, the first term of (3.5) is 0. And under certain regularity condition, the second term is

$$\int f(x)\Big[\frac{\partial}{\partial \theta}\log(g(x|\theta))\Big]\Big|_{\theta=\theta_0} dx = \mathrm{E}_x\Big[\Big\{\frac{\partial}{\partial \theta}\log(g(x|\theta))\Big\}\Big|_{\theta=\theta_0}\Big] = 0.$$

Now we come back to (3.4), and again since the expectation is taken with respect to f, and g and f may not be the same, $-\mathrm{E}_x\frac{\partial^2 \log(g(x|\theta_0))}{\partial \theta^2}$ is not exactly equal to the Fisher's information matrix $\mathcal{I}(\theta_0)$; however, Fisher's information matrix is a good approximation to it if g is a good approximation to f. And in fact, we can consider to take the expectation with respect to g as an approximation to the case when the expectation is taken with respect to the unknown f as long as g is just a good model for f (Shibata (1989)). As for the situation when g is too restrictive to be good, the term $-2\log(g(x|\hat{\theta}))$ in the AIC will be much inflated, and then we will not select that model.

Now equation (3.4) can be written as

$$\mathrm{E}_x\big[\log(g(x|\hat{\theta}))\big] \approx \mathrm{E}_x[\log(g(x|\theta_0))] - \frac{1}{2}[\hat{\theta}-\theta_0]'\mathcal{I}(\theta_0)[\hat{\theta}-\theta_0].$$

We take the expectation of the above equation with respect to $\hat{\theta}$ (i.e. with respect to y). Note that $\mathrm{E}_x[\log(g(x|\theta_0))]$ is a constant and can be written as $\mathrm{E}_y[\log(g(y|\theta_0))]$. So we have

$$\begin{aligned}\mathrm{E}_y\mathrm{E}_x\big[\log(g(x|\hat{\theta}))\big] &\approx \mathrm{E}_y[\log(g(y|\theta_0))] - \frac{1}{2}\mathrm{E}_y\{[\hat{\theta}-\theta_0]'\mathcal{I}(\theta_0)[\hat{\theta}-\theta_0]\} \\ &= \mathrm{E}_y[\log(g(y|\hat{\theta}))] + \mathrm{E}_y[\log(g(y|\theta_0)) - \log(g(y|\hat{\theta}))] \\ &\quad - \frac{1}{2}\mathrm{E}_y\{[\hat{\theta}-\theta_0]'\mathcal{I}(\theta_0)[\hat{\theta}-\theta_0]\}.\end{aligned} \tag{3.6}$$

In (3.6), the term $\mathrm{E}_y[\log(g(y|\theta_0))-\log(g(y|\hat{\theta}))]$ can be approximated by Taylor series expansion of $\log(g(y|\theta_0))$ around $\hat{\theta}$, and again the truncation at the quadratic term entails an error of approximation that goes to zero as $n \to \infty$.

$$\begin{aligned}
&\mathrm{E}_y[\log(g(y|\theta_0)) - \log(g(y|\hat{\theta}))] \\
\approx&\mathrm{E}_y\Big\{\Big[\frac{\partial \log(g(y|\hat{\theta}))}{\partial\theta}\Big]'[\hat{\theta}-\theta_0]\Big\} + \frac{1}{2}\mathrm{E}_y\Big\{[\hat{\theta}-\theta_0]'\Big[\frac{\partial^2 \log(g(y|\hat{\theta}))}{\partial\theta^2}\Big][\hat{\theta}-\theta_0]\Big\} \qquad (3.7)\\
=&\frac{1}{2}\mathrm{E}_y\Big\{[\hat{\theta}-\theta_0]'\Big[\frac{\partial^2 \log(g(y|\hat{\theta}))}{\partial\theta^2}\Big][\hat{\theta}-\theta_0]\Big\}.
\end{aligned}$$

The linear term in (3.7) vanishes because the MLE $\hat{\theta}$ satisfies $\frac{\partial \log(g(y|\hat{\theta}))}{\partial\theta} = 0$.

So now $\mathrm{E}_y\mathrm{E}_x\big[\log(g(x|\hat{\theta}))\big]$ is approximated by

$$\begin{aligned}
\mathrm{E}_y\mathrm{E}_x\big[\log(g(x|\hat{\theta}))\big] \approx \mathrm{E}_y[\log(g(y|\theta_0))] &- \frac{1}{2}\mathrm{E}_y\Big\{[\hat{\theta}-\theta_0]'\Big[-\frac{\partial^2 \log(g(y|\hat{\theta}))}{\partial\theta^2}\Big][\hat{\theta}-\theta_0]\Big\} \qquad (3.8)\\
&- \frac{1}{2}\mathrm{E}_y\{[\hat{\theta}-\theta_0]'\mathcal{I}(\theta_0)[\hat{\theta}-\theta_0]\}.
\end{aligned}$$

Under certain regularity condition, $[\hat{\theta}-\theta_0]'\Big[-\frac{\partial^2 \log(g(y|\hat{\theta}))}{\partial\theta^2}\Big][\hat{\theta}-\theta_0]$ and $[\hat{\theta}-\theta_0]'\mathcal{I}(\theta_0)[\hat{\theta}-\theta_0]$ both asymptotically follow χ^2 distribution with degrees of freedom p, the dimension of θ. We refer to Roatgi and Saleh (2001) page 419 for the regularity conditions for the i.i.d. sample case. Since $\mathrm{E}\chi^2_p = p$, then

$$-\mathrm{E}_y\mathrm{E}_x\big[\log(g(x|\hat{\theta}))\big] \approx \mathrm{E}_y[-\log(g(y|\theta_0))] + p.$$

So $-\log(g(y|\theta_0)) + p$ is an asymptotically unbiased estimate of $-\mathrm{E}_y\mathrm{E}_x\big[\log(g(x|\hat{\theta}))\big]$. We complete the proof here.

In the point process case, the observation length T is fixed while the sample size n is random, but since the Poisson process has independent increment, the derivation of AIC in the Poisson process case should be essentially similar as the derivation of AIC when we have independent observations. We will give some insight of the AIC in the Poisson process case in section 3.2.

3.2.2 Bayesian information criterion

Bayesian information criterion (BIC) was established by Schwarz (1978). It is a consistent model selection criterion which selects the true model with probability 1 when the sample size n goes to infinity. That is because of the large penalty term in the criterion:

$$\textbf{BIC} = -2\log(\mathcal{L}(\hat{\theta}|data)) + p\log(n),$$

where p is the dimension of the parameters, and $\mathcal{L}$ is the likelihood function of the data. The procedure selects the model with the smallest BIC value.

In Schwarz (1978), BIC was originally established for the case of independent, identically distributed observations, and linear models, with the assumption that the observations come from a Koopman-Darmois family, i.e., relative to some fixed measure on the sample space they possess a density of the form

$$f(x,\theta) = \exp(\theta \cdot y(x) - b(\theta)),$$

where y is the sufficient p-dimensional statistic. However, the density function does not necessarily need to be in such form. We give a review of the general derivation of BIC based on Schwarz (1978) and Burnham and Anderson (2002) Chapter 6. We refer to Berger and Pericchi (2001) for more discussion on the objective Bayesian methods for model selection.

It was evident that the assumptions and interpretations about prior probabilities are irrelevant in deriving the basic BIC result. It suffices to assume that the prior probability of the jth model being the true one is α_j, and $\pi_j(\theta)$ is the conditional prior distribution of θ given the jth model. Via Bayes' formula the posterior probability of the jth model being the true one is

$$\frac{g_j(x,\pi_j)\alpha_j}{\sum_r g_r(x,\pi_r)\alpha_r},$$

where $g_j(x,\pi_j)$ is the marginal likelihood of the jth model with $g_j(x,\pi_j) = \int g_j(x|\theta)\pi_j(\theta)d\theta$. The prior probability α_j will be bounded in n and thus being dropped from the penalty term in

BIC, and it is sufficient to assume equal prior probability for each model. We refer to Schwarz (1978) for more details. The Bayes solution consists of selecting the model that is a posteriori most probable, that is, by dropping the index j, we select the model which maximizes the following quantity:

$$\int g(x|\theta)\pi(\theta)d\theta. \tag{3.9}$$

The Laplace method is used to approximate the above integral (3.9). Then $\log(g(x|\theta))$ is substituted by the following approximation:

$$\log(g(x|\theta)) \approx \log(g(x|\hat{\theta})) - \frac{1}{2}[\theta-\hat{\theta}]'\Big[-\frac{\partial^2\log(g(y|\hat{\theta}))}{\partial\theta^2}\Big][\theta-\hat{\theta}].$$

So the needed integral (3.9) is approximately

$$\begin{aligned}
&\int \exp\{\log(g(x|\hat{\theta})) - \frac{1}{2}[\theta-\hat{\theta}]'\Big[-\frac{\partial^2\log(g(y|\hat{\theta}))}{\partial\theta^2}\Big][\theta-\hat{\theta}]\}d\theta \\
=&\exp\{\log(g(x|\hat{\theta}))\}\int \exp\{-\frac{1}{2}[\theta-\hat{\theta}]'\Big[-\frac{\partial^2\log(g(y|\hat{\theta}))}{\partial\theta^2}\Big][\theta-\hat{\theta}]\}d\theta \\
=&\exp\{\log(g(x|\hat{\theta}))\}(2\pi)^{p/2}\parallel -\frac{\partial^2\log(g(y|\hat{\theta}))}{\partial\theta^2}\parallel^{-1/2}.
\end{aligned} \tag{3.10}$$

where $\parallel\cdot\parallel$ denotes the determinant of a matrix and the last line in (3.10) follows by the property of multivariate normal density function

$$\int (2\pi)^{-p/2}\parallel -\frac{\partial^2\log(g(y|\hat{\theta}))}{\partial\theta^2}\parallel^{1/2}\exp\{-\frac{1}{2}[\theta-\hat{\theta}]'\Big[-\frac{\partial^2\log(g(y|\hat{\theta}))}{\partial\theta^2}\Big][\theta-\hat{\theta}]\}d\theta = 1.$$

Taking -2 times the log of the last equation in (3.10), we have

$$-2\log(g(x|\hat{\theta})) + \log\Big(\parallel -\frac{\partial^2\log(g(y|\hat{\theta}))}{\partial\theta^2}\parallel\Big) - p\log(2\pi). \tag{3.11}$$

When there are n independent random observations, and under general regularity conditions, $-\frac{\partial^2\log(g(y|\hat{\theta}))}{\partial\theta^2} \approx n\Sigma$ where Σ^{-1} is the asymptotic variance-covariance matrix of the maximum likelihood estimate $\hat{\theta}$ and its determinant is of size of $O(1)$. So by the property of determinant,

$$\parallel -\frac{\partial^2\log(g(y|\hat{\theta}))}{\partial\theta^2}\parallel \approx n^p\parallel\Sigma\parallel, \tag{3.12}$$

expression (3.11) is approximately

$$-2\log(g(x|\hat{\theta})) + p\log(n) + \log(\| \Sigma \|) - p\log(2\pi).$$

Schwarz (1978) drops the last two terms of the expression above presumably because, asymptotically, they are dominated by the term of order $\log(n)$ as well as by the log-likelihood term. Now we finish the derivation of BIC. Note that there is no mathematical requirement in the derivation of BIC that the model g be true, but the proof of the consistency property of BIC requires the assumption of a true model in the candidate models.

3.3 AIC and BIC in the almost periodic Poisson processes context

We derive the AIC and BIC in the almost periodic Poisson processes in this section, starting from the likelihood function and the maximum likelihood estimates of the processes under consideration.

3.3.1 Likelihood function of a Poisson process

Denote the intensity function of a Poisson process $\lambda(t;\theta)$ as $\lambda(t)$. For $0 \leq t_1 < t_2 < \cdots < t_n$, the likelihood of a Poisson process is

$$\mathcal{L}(\theta) = \Pr(T_1 = t_1, T_2 = t_2, \ldots, T_n = t_n, t_n \leq T) = \lambda(t_1)\lambda(t_2)\cdots\lambda(t_n)e^{\int_0^T \lambda(s)ds},$$

where T is the observation length.

So the log-likelihood function is

$$\ell(\theta) = \sum_{j=1}^{N(T)} \log \lambda(t_j) - \int_0^T \lambda(t)dt.$$

If conditioning on the random sample size n, then

$$\Pr(T_1 = t_1, T_2 = t_2, \ldots, T_n = t_n | N(T) = n) = \lambda(t_1)\lambda(t_2)\cdots\lambda(t_n)e^{\int_0^{t_n} \lambda(s)ds}.$$

The above results can be derived from the property of Poisson processes that the number of points in the time interval $(0, t)$ follows a Poisson distribution with mean $\int_0^t \lambda(s)ds$. We refer the details to Durrett (1999) page 137.

3.3.2 MLE of the almost periodic Poisson processes

In the derivation of AIC and BIC, we need the asymptotic normality of the maximum likelihood estimates. We shall follow the usual approach, based on the asymptotic expansion of the first derivative of the log-likelihood function about the true value $\theta = \theta_0$ of the parameters,

$$0 = \frac{\partial \ell}{\partial \theta}\Big|_{\theta=\hat{\theta}} = \frac{\partial \ell}{\partial \theta}\Big|_{\theta=\theta_0} + \frac{\partial^2 \ell}{\partial \theta^2}\Big|_{\theta=\bar{\theta}}(\hat{\theta} - \theta_0), \tag{3.13}$$

where $\bar{\theta}$ satisfies $\| \bar{\theta} - \theta_0 \| \leq \| \hat{\theta} - \theta_0 \|$.

To justify this approach, we should first establish the asymptotic consistency and order of convergence of the estimates, then check that the Taylor expansion can be sustained, and finally determine the covariance matrix of the estimates.

We shall begin with the consistency of the estimates.

From the likelihood function of the Poisson processes in section 3.1.3, it is easy to see that the observations $\{t_1, t_2, \ldots, t_n\}$ are weakly dependent.

$$\Pr(T_{j+1} = t | T_j = t_j, \ldots, T_1 = t_1) = \lambda(t) e^{-\int_{t_j}^t \lambda(s)ds} = \lambda(t) e^{-\Lambda(t)+\Lambda(t_j)}, \quad t > t_j.$$

Bar-Shalom (1971) and Crowder (1976) among others provided the regularity conditions to ensure the consistency and asymptotic normality of the maximum likelihood estimate from dependent observations. The key idea is that if the dependency is weak, then the accumulated information on the parameters from the sample increases indefinitely, and thus the MLE possess desired statistical property such as consistency and asymptotic normality.

The model we consider is a non-homogeneous Poisson process, with the following almost periodic intensity function (this was previously introduced in Chapter 2):

$$\lambda(t;\theta) = \sum_{k=1}^{K} A_k \cos(\omega_k t + \phi_k) + B, \tag{3.14}$$

where θ is the set of parameters, and A_k, B, ω_k, ϕ_k are unknown parameters with $A_1 > A_2 > \cdots > A_K > 0$, $-\pi/2 \leq \phi_k < 3\pi/2$ and $\omega_k > 0$, $k = 1, \ldots, K$. The baseline B is a constant such that $\lambda(t)$ is positive for any $t > 0$. A sufficient condition to guarantee the positivity of $\lambda(t)$ is $B > \sum_{k=1}^{K} A_k$. For simplicity, we assume $\lambda(t) \geq \varepsilon > 0$ for any $t > 0$, since in the likelihood function $\lambda(t)$ needs to be positive. This assumption is easy to be justified in practice, for example, by removing the overnight waiting time in the stock transaction process, we can take out the part that $\lambda(t) = 0$. In addition, we assume that the intensity function is upper-bounded in a suitable range and thus the parameter space of our model is compact. We state the assumptions in proposition 15.

Following the discussion of the consistency of MLE in Severini (2000), we can easily prove the consistency of the MLE in our model. Moreover, as shown in Chapter 2, the convergence rate of the periodogram estimate of the frequencies is $o(T^{-1})$, and we should expect a similar result for the MLE. As for the MLE of other parameters in the model, the convergence rate is $o(1)$. We state the results here and refer the proof to Appendix B.

Proposition 15 *For a non-homogeneous Poisson process with an almost intensity function* (3.14)*, assume that* $\lambda(t) \in [\varepsilon, M]$ *for any* $t > 0$ *where* $M > \varepsilon > 0$ *and* $M = O(1)$*, and the frequencies* $\{\omega_k, k = 1, \ldots, K\}$ *satisfy the assumption 6, then the maximum likelihood estimate* $\hat{\omega_k}$ *is a consistent estimate of* ω_k *with*

$$(\hat{\omega_k} - \omega_k) = o(T^{-1}),$$

and the maximum likelihood estimates of other parameters, $\hat{B}$*,* $\hat{A}_k$ *and* $\hat{\phi}_k$*, are also consistent, where* $k = 1, \ldots, K$*.*

One remark is needed that not all non-homogeneous Poisson processes have consistent estimates for its parameters. Nayak et al. (2008) proved that for the type of non-homogeneous Poisson process model that the expected number of events, such as software failures, is finite in infinite observing time, the parameters of these models cannot be estimated consistently. In other words, if the integral of the intensity function from 0 to ∞ is finite, then the number of events in the long run is finite on average, and the information from the process is limited and cannot accrue steadily as observing time approaches infinity, and thus cannot provide consistent estimates of the model parameters. In our model, this is not the case.

Let $\theta = (A_1, \ldots, A_K, B, \phi_1, \ldots, \phi_K, \omega_1, \ldots, \omega_K)'$. Define the diagonal matrix D_T as

$$D_T = \begin{pmatrix} T^{\frac{1}{2}} I_{(2K+1)\times(2K+1)} & 0 \\ 0 & T^{\frac{3}{2}} I_{K\times K} \end{pmatrix},$$

where I is the identity matrix.

The Taylor expansion (3.13) is easy to be justified; $0 \leq \| \bar{\theta} - \theta_0 \| \leq \| \hat{\theta} - \theta_0 \|$ is needed in evaluating the limit value of the second derivative term and we refer the details to Appendix B. In fact, we have the results:

$$\begin{aligned} &- \lim_{T\to\infty} D_T^{-1} \frac{\partial^2 \ell}{\partial \theta^2}\Big|_{\theta=\bar{\theta}} D_T^{-1} = - \lim_{T\to\infty} D_T^{-1} \frac{\partial^2 \ell}{\partial \theta^2}\Big|_{\theta=\theta_0} D_T^{-1} \\ &= \lim_{T\to\infty} D_T^{-1} \mathrm{var}\Big(\frac{\partial \ell}{\partial \theta}\Big|_{\theta=\theta_0}\Big) D_T^{-1} = \lim_{T\to\infty} D_T^{-1} \int_0^T \frac{1}{\lambda(t)} \Big(\frac{\partial \lambda(t)}{\partial \theta}\Big)^2\Big|_{\theta=\theta_0} dt D_T^{-1}. \end{aligned} \tag{3.15}$$

Denote the limiting value as Σ. We find the asymptotic normality of the parameter estimates as follows.

Proposition 16 *In the non-homogeneous Poisson process with intensity function* (3.14)*, and under the assumptions stated in proposition 15, $D_T(\hat{\theta} - \theta)$ is asymptotically normally distributed with mean* $\mathbf{0}$*, and variance-covariance matrix Σ^{-1}.*

It is difficult to obtain the explicit form of Σ, but Σ can be calculated numerically. Here Σ serves as the Fisher's information matrix. The main purpose of deriving the asymptotic results

of MLE is to show that MLEs of parameters in the almost periodic Poisson process model are well behaved as i.i.d. random variables case, with a different normalizing factor D_T^{-1} rather than $n^{-1/2}$ where n is the sample size. The value of Σ is of less importance in the derivation of AIC and BIC.

3.3.3 AIC in the almost periodic Poisson processes

To carry over the general derivation of AIC in section 3.2.1 to the almost periodic Poisson processes, there are two steps that need to be justified: first, in the Taylor series expansion of the log-likelihood function, the truncation at the quadratic term entails an unknown degree of approximation, but it is an error of approximation that goes to zero as $T \to \infty$. We already justified this in the discussion of the MLE of almost periodic Poisson processes in section 3.3.2. The second step is that $[\hat{\theta} - \theta_0]' \Big[- \frac{\partial^2 \log(g(y|\hat{\theta}))}{\partial \theta^2} \Big] [\hat{\theta} - \theta_0]$ and $[\hat{\theta} - \theta_0]' \mathcal{I}(\theta_0) [\hat{\theta} - \theta_0]$ both asymptotically follow χ^2 distribution with degrees of freedom p which is the dimension of θ. Here $\log(g(\cdot))$ is the log-likelihood function denoted as $\ell(\cdot)$ in section 3.3.2, and $\mathcal{I}(\theta_0)$ is the Fisher's information matrix for n observations. In the almost periodic Poisson processes, we have

$$[\hat{\theta} - \theta_0]' \mathcal{I}(\theta_0) [\hat{\theta} - \theta_0] = [D_T^{-1}(\hat{\theta} - \theta_0)]' \Big[D_T^{-1} \frac{\partial^2 \ell}{\partial \theta^2} \Big|_{\theta = \theta_0} D_T^{-1} \Big] [D_T^{-1}(\hat{\theta} - \theta_0)].$$

It follows from proposition 16 that $D_T^{-1}(\hat{\theta} - \theta)$ and is asymptotically normally distributed with mean $\mathbf{0}$, and variance-covariance matrix Σ^{-1}. In addition, equation (3.15) shows that $D_T^{-1} \frac{\partial^2 \ell}{\partial \theta^2} \Big|_{\theta = \theta_0} D_T^{-1} \to \Sigma$ as $T \to \infty$. So by the property of multivariate normal distribution, $[\hat{\theta} - \theta_0]' \mathcal{I}(\theta_0) [\hat{\theta} - \theta_0]$ in the almost periodic Poisson processes is asymptotically χ^2 distributed with degrees of freedom $3K + 1$. Similarly, we can also proof that $[\hat{\theta} - \theta_0]' \Big[- \frac{\partial^2 \log(g(y|\hat{\theta}))}{\partial \theta^2} \Big] [\hat{\theta} - \theta_0]$ behaves the same as $[\hat{\theta} - \theta_0]' \mathcal{I}(\theta_0) [\hat{\theta} - \theta_0]$ in the almost periodic Poisson processes, so AIC

in the almost periodic Poisson processes is

$$\textbf{AIC} = -2\ell(\hat{\theta}) + 2(3K+1).$$

3.3.4 BIC in the almost periodic Poisson processes

Consider the derivation of BIC in the almost periodic Poisson processes, we only need to modify the determinant of the second derivative of the log-likelihood function in expression (3.11) in section 3.2.2. The results (3.15) yields that

$$\begin{aligned} \parallel -\frac{\partial^2 \ell}{\partial \theta^2}\Big|_{\theta=\hat{\theta}} \parallel &\approx \parallel D_T \Sigma D_T \parallel \\ &= \parallel D_T^2 \parallel \parallel \Sigma \parallel = T^{2\{\frac{1}{2}(2K+1)+\frac{3}{2}K\}} \parallel \Sigma \parallel, \end{aligned}$$

where Σ is defined after equation (3.15). In this case, the penalty term in BIC changes to $\log(T^{(2K+1)+3K}) = (5K+1)\log(T)$. So BIC in the almost periodic Poisson processes is

$$\textbf{BIC} = -2\ell(\hat{\theta}) + (5K+1)\log(T).$$

The coefficient of $\log(T)$ is greater than the number of parameters $3K+1$ because of the super-efficiency of the frequency estimates.

3.4 Summary and implementation of the proposed methodology

In this section, we summarize the procedures to implement the proposed methodology.

The use of AIC and BIC requires MLE, and the likelihood surface of the almost periodic Poisson process has many local maxima and local minima and thus the usual optimization algorithms such as Newton-Raphson method can be easily trapped by the local maxima or minima unless good initial values are provided. So in the search of MLE of our model, we suggest to use the periodogram estimates proposed in Chapter 2 as good initial values. The procedures to implement the proposed methodology are described as follows.

For a given set of data, consider $K = 1, 2, \ldots, \tilde{K}$, where $\tilde{K}$ is the largest K value under consideration.

1. Compute the periodogram estimates of the parameters when $K = \tilde{K}$.

2. For each K value, set the corresponding periodogram estimates as initial values, and search for MLE.

3. Calculate AIC and BIC values for each K. Select K based on these two criteria.

4. If two criteria select different K, conduct the out-of-sample prediction and select the model with smaller (averaged) squared prediction error.

3.5 Simulation

In this section, we consider the same almost periodic intensity functions in Chapter 2, section 2.6:

$$\text{Case 1:} \quad \lambda(t) = 1.6 + \cos\Big(\frac{\pi}{4\sqrt{3}}t\Big) + 0.5\cos\Big(\frac{\pi}{3\sqrt{2}}t + \frac{\pi}{4}\Big),$$

$$\text{Case 2:} \quad \lambda(t) = \sqrt{3.1 + 3\cos\Big(\frac{\pi}{3\sqrt{2}}t\Big)},$$

$$\text{Case 3:} \quad \lambda(t) = 0.1 + 0.5\text{Mod}[t, 2\pi],$$

$$\text{Case 4:} \quad \lambda(t) = 1.3\exp\Big\{\cos\Big(\frac{\pi}{3\sqrt{2}}t + \frac{\pi}{4}\Big)\Big\}.$$

We use the same simulated data set as described in section 2.6, that is, for each case, simulate 100 independent replicates, and cut off the process at $T = 500$ for model fitting, and conduct out-of-sample prediction for 50 data points ($901st$ to $950th$ data points). We perform the analysis using the procedures described in section 3.4, and show that our model (AP) outperforms the existing models, Vere-Jones (1982) and the homogeneous Poisson process model (HPP), in terms of the out-of-sample predictability.

In Case 1, we know that the true K is 2. Among 100 replicates, AIC and BIC select the same model in 19 of them, 37 are in favor of AIC, that is, the models selected by AIC have

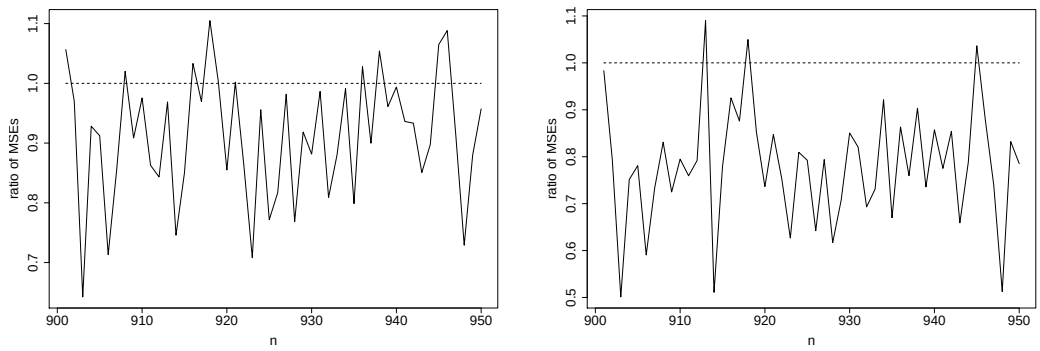
smaller averaged squared prediction error, and 44 are in favor of BIC. Table 3.1 shows the selection frequencies for different K values in 100 replicates. In this table, the final selection is done by comparing the averaged squared prediction errors from models selected by AIC and BIC. To clarify, there is no comparison between the last row and the first two rows. Take the last number in the last row, 25, for example. It means that in 100 replicates, we choose $K = 4$ 25 times after comparing the models selected by AIC and BIC in terms of smaller averaged squared prediction error, and since AIC selects $K = 4$ 52 times in 100 replicates and BIC does not select any $K = 4$, then among those 52 replicates, only 25 are in favor of AIC.

Table 3.1: Selection frequencies for different K values in 100 replicates using AIC and BIC, and the final selection by comparing the averaged squared prediction errors from models selected by AIC and BIC.

K	1	**2**	3	4
AIC	0	19	29	52
BIC	3	**96**	1	0
Final selection	0	62	13	25

Since BIC is consistent, so it selects the true K ($K = 2$) most of the time. Note that for a given set of data with finite sample size, the best K for this data set may not be the true K, and using the information distance as the criterion to measure the distance between model candidates, the K values selected by AIC have larger variability.

Figure 3.1 shows the the ratios of MSE under our model and MSE under the existing models. If the ratio is below 1, then our model is better. MSE is calculated as $MSE = \frac{1}{100}\sum_{i=1}^{100}(t_{n+1}^{i} - \hat{t}_{n+1}^{i})^2$. The averaged reductions in MSE compare to Vere-Jones's model



(a) AP vs Vere-Jones(1982)

(b) AP vs HPP

Figure 3.1: Case 1: MSE ratios. Ratio less than 1 indicates that our model is better.

and homogeneous Poisson model are 8.9% and 21.8% , respectively.

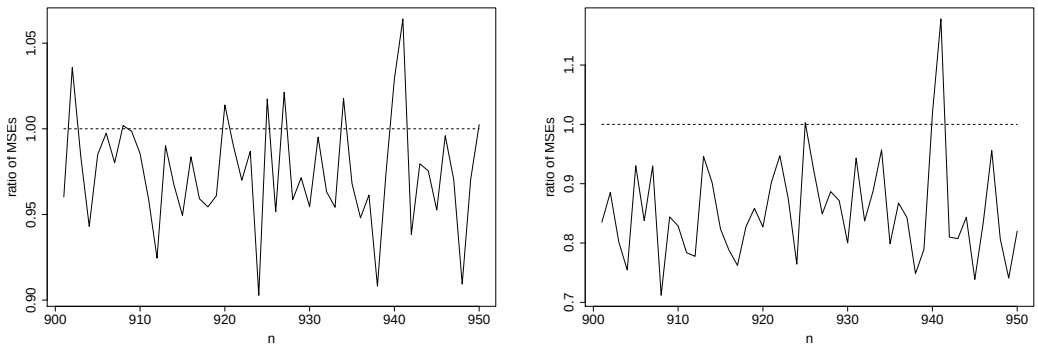
In Case 2, 3, and 4, we do not know the true value of K or the 'true' K is infinity since it takes infinite terms to express the intensity function in the form of our model by Fourier expansion. With a finite sample size, we can only approximate the true intensity function within certain precision. It is expected that with increased sample size, the data should be able to capture more and more sinusoidal terms to better approximate the true intensity function.

We do not compare AIC and BIC in the simulation. The main purpose of the simulation is to implement the procedure in section 3.4 and show that our model outperforms the existing models. In Table 3.2, we show the selection frequencies for different K values in 100 replicates for Case 2, 3 and 4, respectively. And the ratios of MSE under our model and existing models are shown in Figure 3.2, 3.3 and 3.4 for Case 2, 3 and 4, respectively. In Case 2, the averaged reductions in MSE compare to Vere-Jones's model and homogeneous Poisson model are 2.53% and 14.6%. In Case 3, the averaged reductions are 3.73% and 12.6%, and in Case 4, the averaged reductions are 1.4% and 23.9%. One interesting result here is that even Case 4 is in the same function form of Vere-Jones (1982), and thus the MSE under Vere-Jones model is expected to

be smaller than the MSE under our model and the corresponding MSE ratios are expected to be greater than 1, but in this simulation our model does even better than Vere-Jones model by reducing the MSE by 1.4% on average. Although this result may be by coincidence, it is an evident that our model is much more flexible than the existing models.

Table 3.2: Selection frequencies for different K values in 100 replicates

	Case 2				Case 3					Case 4			
K	1	2	3	4	1	2	3	4	5	1	2	3	4
AIC	1	3	30	66	0	3	30	28	39	5	22	33	40
BIC	99	1	0	0	10	70	19	1	0	81	19	0	0
Final selection	63	3	11	23	5	35	31	13	16	47	19	18	16

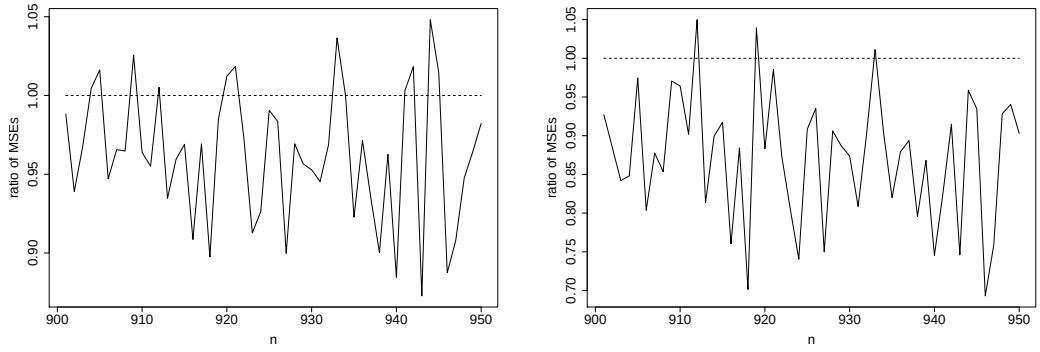


(a) AP vs Vere-Jones(1982)

(b) AP vs HPP

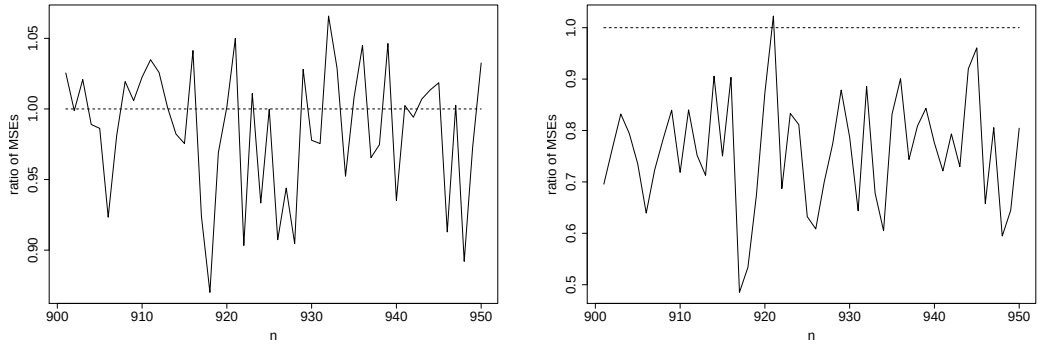
Figure 3.2: Case 2: MSE ratios. Ratio less than 1 indicates that our model is better.

To complete the discussion of MLE in section 3.3.2, we calculate the sample means and standard errors of the MLE from the 100 replicates in Case 1 and compare them to the peri-



(a) AP vs Vere-Jones(1982) (b) AP vs HPP

Figure 3.3: Case 3: MSE ratios. Ratio less than 1 indicates that our model is better.



(a) AP vs Vere-Jones(1982) (b) AP vs HPP

Figure 3.4: Case 4: MSE ratios. Ratio less than 1 indicates that our model is better.

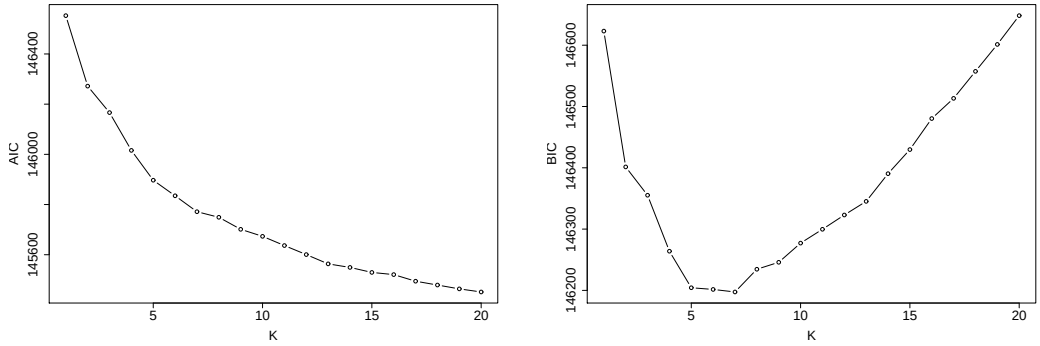
odogram estimates in Chapter 2. The results are shown in Table 3.3. We see that MLE only slightly reduces the standard errors compared to the standard errors of the periodogram estimates, but the computational complexity of MLE is much higher. So the periodogram estimates may be enough to provide accurate information about the true parameters.

Table 3.3: The means and standard errors of the MLE from the 100 replicates in Case 1 with comparison to the periodogram estimates

	ω_1	ω_2	A_1	A_2	ϕ_1	ϕ_2	B
true value	0.45345	0.74048	1	0.5	0	0.78540	1.6
MLE sample mean	0.45342	0.74051	1.00854	0.50322	-0.00420	0.77388	1.59834
MLE sample sd	0.00047	0.00097	0.07537	0.07709	0.16396	0.29610	0.05772
peri est sample mean	0.45374	0.74116	1.01223	0.50671	-0.07806	0.59168	1.60616
peri est sample sd	0.00052	0.00112	0.07507	0.07838	0.16499	0.33445	0.05613

3.6 IBM example revisited

We use the same IBM data set as in Chapter 2, section 2.7. The first four weeks data are used for model fitting and the day after the 4-week period is left out for out-of-sample prediction. We first exam AIC and BIC for $K = 1, 2, \cdots, 20$. The likelihood values differ a lot for different K values, and the order of AIC is dominated by the likelihood function. As expected, the likelihood value increases as we include more and more sinusoidal terms in the model, and the smallest AIC occurs at $K = 20$. The penalty term in BIC is much larger than AIC, and the smallest BIC occurs at $K = 7$. Figure 3.5 shows the AIC and BIC at different K values.



(a) AIC (b) BIC

Figure 3.5: IBM data analysis: AIC and BIC

Since BIC at $K = 5, 6$ and 7 are very close to each other with difference less than 3, we include all of them as well as $K = 20$ into the model candidates and then select the best K value by comparing the averaged prediction errors. More sinusoidal terms may result in a better approximation to the 'true' intensity function, but they also bring more uncertainty, that is, larger variation, into the model fitting and thus result in larger out-of-sample prediction errors. It turns out that after comparing the averaged prediction errors under $K = 5, 6, 7$ and 20, the model with $K = 5$ produces the smallest averaged prediction error. We report the estimation results here:

$$\boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3, \omega_4, \omega_5) = 10^{-4}(2.6762, 0.5922, 8.0812, 1.2067, 5.3786),$$

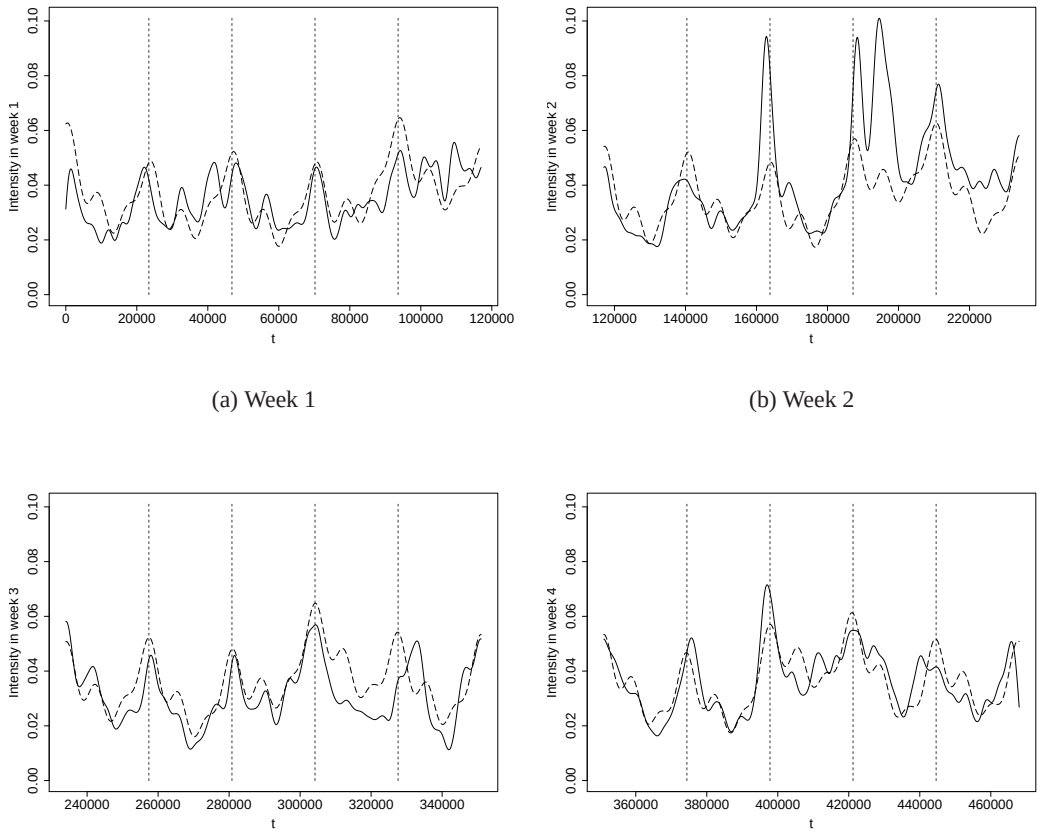
$$\boldsymbol{A} = (A_1, A_2, A_3, A_4, A_5) = 10^{-2}(1.070, 0.640, 0.462, 0.469, 0.411),$$

$$\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3, \phi_4, \phi_5) = (0.0305, 0.3577, -1.0143, 0.8374, 0.4087),$$

and $B = 0.036334$.

Note that these parameter estimates are maximum likelihood estimates and they are slightly different from the periodogram estimates in section 2.7. The estimated intensity function is very similar to the one shown in Figure 2.6 since the maximum likelihood estimates do not differ a

lot from the periodogram estimates. Figure 3.6 displays the results.



(a) Week 1

(b) Week 2

(c) Week 3

(d) Week 4

Figure 3.6: Solid line: the non-parametric estimate of the intensity function as described in section 2.7. Long dashed line (– – –): the estimated intensity function using MLE. The vertical light dashed lines separate days in each week.

Chapter 4

Summary and Future Work

4.1 Summary

In this thesis, a very general model for the intensity function of a non-homogeneous Poisson process is proposed and investigated. The processes under consideration usually show a pattern of periodicity, such as the timing of stock transactions with higher activity during the opening and closing than the middle of the day. And such processes can be modeled as almost periodic point processes. The proposed model is for the almost periodic Poisson process which includes the purely periodic Poisson process as a special case, and it is much more general and useful than the existing models. We demonstrate this by both simulations and real life data analysis.

The basic concepts of point processes are introduced in Chapter 1. Readers with knowledge in the general Poisson processes (both homogeneous and non-homogeneous Poisson processes) may skip section 1.2 and section 1.3. We suggest reading section 1.4 since the periodogram introduced there has a close relationship with the intensity function we proposed to use.

In Chapter 2, the model for the almost periodic Poisson processes is introduced. We give a very detailed literature review on the study of purely periodic Poisson processes. We then

show that almost periodic Poisson processes are much more general and they should be used for processes with any pattern of periodicity. The concept of almost periodicity is introduced in section 2.2. After that, we study the estimation of the model parameters, and propose a class of simple, consistent, and asymptotically normal parameter estimates mainly based on the periodogram. The computational issues in estimating the frequencies are discussed in section 2.4 and the prediction of the next occurrence using our model is studied in section 2.5. Then we use simulations and application to show the usefulness of the proposed model.

The model we studied is actually a class of models indexed by K which is the number of sinusoidal terms in the intensity function. In Chapter 2, the parameter estimates are defined for each fixed K. In the data analysis, K is usually unknown, and we need to select the best K so that the corresponding model can best approximate the unknown truth. So in Chapter 3, we propose to use model selection criteria to determine K. We first review and give interpretation of two representatives in the model selection: AIC and BIC; and show the general derivation of these two criteria when there are i.i.d. observations. Then we derive AIC and BIC in the frame work of our model. One thing we would like to point out is that we choose to regard the number of points/observations to be random and the observation length to be fixed for mathematical convenience. The implementation of proposed methodology in both Chapter 2 and Chapter 3 are summarized in section 3.4, and then we follow the procedures to re-analyze the simulated data set and IBM data set in Chapter 2.

4.2 Future work

This thesis proposes a parametric model for the first order property (the intensity function) of a point process. There are three directions for future research. The first one is to study the almost periodic intensity function using non-parametric method. Many papers have investigated

the non-parametric approach in the purely periodic Poisson processes, we refer to Chapter 2 for a detailed literature review. Hall and Yin (2003) proposed non-parametric methods for deconvolving multiperiodic functions where the periods are relatively irrational, and we may borrow some ideas from this paper and apply to the point process case. The challenge is that in Hall and Yin (2003) the value of the multiperiodic functions are observable (with errors), but in the point process, the intensity function is unobservable.

The second direction is to consider higher order property of the point processes. Under Poisson assumption, the process has independent increment. The question is whether this assumption can be relaxed to conditions like weak dependency in the increment process. In addition, there is no paper studying almost periodically correlated point processes although many papers have discussed the almost periodically correlated time series. There is still a lot of room on this problem.

The third direction is in the estimation of K in the proposed model. In this thesis, we use model selection criteria and derive AIC and BIC in the framework of our model. How to construct a criterion which is tailored for point process is challenging.

Appendix A

Proofs in Chapter 2

For the proofs, the following results will be useful.

Under Assumption 6, simple calculations yield that for $\omega > 0$ and constant $c \neq 0$, $\gamma > 0$,

$$T^{-1}\int_0^T Be^{-i\omega t}dt = B\frac{\exp(-i\omega T)-1}{-i\omega T} = \begin{cases} B + o(1), & \text{if } \omega = cT^{-1-\gamma} = o(T^{-1}), \\ B\frac{1-e^{-ic}}{ic}, & \text{if } \omega = cT^{-1}, \\ o(1), & \text{if } \omega = cT^{-1+\gamma}, \end{cases}$$

$$\begin{aligned} & T^{-1}\int_0^T A_k\cos(\omega_k t+\phi_k)e^{-i\omega t}dt \\ & = T^{-1}\int_0^T A_k\frac{e^{i(\omega_k t+\phi_k)}+e^{-i(\omega_k t+\phi_k)}}{2}e^{-i\omega t}dt \\ & = o(1)+\frac{A_k}{2}e^{i\phi_k}T^{-1}\int_0^T e^{-i(\omega-\omega_k)t}dt \\ & = \begin{cases} \frac{A_k}{2}\{\cos(\phi_k)+i\sin(\phi_k)\}+o(1), & \text{if } |\omega-\omega_k| = cT^{-1-\gamma} = o(T^{-1}), \\ \frac{A_k}{2}e^{i\phi_k}\frac{1-e^{-ic}}{ic}+o(1), & \text{if } |\omega-\omega_k| = cT^{-1}, \\ o(1), & \text{if } |\omega-\omega_k| = cT^{-1+\gamma}. \end{cases} \end{aligned} \tag{A.1}$$

Moreover,

$$
\begin{aligned}
&T^{-2}\int_0^T Bte^{-i\omega t}dt = o(1), \text{if } \omega = cT^{-1+\gamma},\\
&T^{-2}\int_0^T A_k\cos(\omega_k t+\phi_k)te^{-i\omega t}dt\\
&=\begin{cases}
\frac{A_k}{4}\{\cos(\phi_k)+i\sin(\phi_k)\}+o(1), & \text{if } |\omega-\omega_k| = cT^{-1-\gamma} = o(T^{-1}),\\
\frac{A_k}{2}e^{i\phi_k}\frac{-1+e^{-ic}+ice^{-ic}}{c^2}+o(1), & \text{if } |\omega-\omega_k| = cT^{-1},\\
o(1), & \text{if } |\omega-\omega_k| = cT^{-1+\gamma}.
\end{cases}\\
&T^{-3}\int_0^T Bt^2e^{-i\omega t}dt = o(1), \text{if } \omega = cT^{-1+\gamma},\\
&T^{-3}\int_0^T A_k\cos(\omega_k t+\phi_k)t^2e^{-i\omega t}dt\\
&=\begin{cases}
\frac{A_k}{6}\{\cos(\phi_k)+i\sin(\phi_k)\}+o(1), & \text{if } |\omega-\omega_k| = cT^{-1-\gamma} = o(T^{-1}),\\
\frac{A_k}{2}e^{i\phi_k}\frac{-2+2e^{-ic}+2ice^{-ic}-c^2e^{-ic}}{ic^3}+o(1), & \text{if } |\omega-\omega_k| = cT^{-1},\\
o(1), & \text{if } |\omega-\omega_k| = cT^{-1+\gamma},
\end{cases}
\end{aligned}
\tag{A.2}
$$

where $k = 1, \ldots, K$.

Proof of Lemma 7. The proof was originally provided by Vere-Jones (1982). We correct the typo there and extend the case to $m \geq 1$. First consider $m = 1$.

Let $J > 0$ be the upper bound of $\lambda(t)$, so $\lambda(t) \leq J$.

Define

$$M_T := \sup_{0\leq\omega\leq\Omega_T}\Big|\int_0^T e^{i\omega t}dZ(t)\Big|.$$

Now let L denote a positive integer (to be fixed later), and consider a division of the interval

$[0,T]$ into subintervals each of length $\Delta = T/L$. We shall write

$$\begin{aligned}\int_0^T e^{i\omega t}dZ(t) &= \sum_{l=0}^{L-1}\int_{l\Delta}^{(l+1)\Delta}(e^{i\omega l\Delta} + e^{i\omega l\Delta}(e^{i\omega(t-l\Delta)}-1))dZ(t)\\ &= \sum_{l=0}^{L-1} e^{i\omega l\Delta}Z_l + \sum_{l=0}^{L-1} e^{i\omega l\Delta}\int_0^{\Delta}(e^{i\omega u}-1)dZ(u+l\Delta) \qquad \text{(A.3)}\\ &= S_1(\omega,T) + S_2(\omega,T),\end{aligned}$$

where

$$Z_l = \int_{l\Delta}^{(l+1)\Delta} dZ(t).$$

The first sum S_1 has the same form as in the discrete case and can be handled by similar methods:

$$|S_1(\omega,T)|^2 = \Big|\sum_{r=-L+1}^{L-1} e^{i\omega r\Delta}\sum_{l=0}^{L-|r|} Z_l Z_{l+r}\Big| \le \sum_{r=-L+1}^{L-1}\Big|\sum_{l=0}^{L-|r|} Z_l Z_{l+r}\Big|,$$

so that

$$\mathrm{E}\Big\{\sup_{\omega}|S_1(\omega,T)|^2\Big\} \le \sum_{r=-L+1}^{L-1}\mathrm{E}\Big\{\Big|\sum_{l=0}^{L-|r|} Z_l Z_{l+r}\Big|\Big\}.$$

The expectation on the right side of this expression can be estimated using the Schwarz inequality and the fact that the Z_l are uncorrelated. This leads to the estimate

$$\mathrm{E}\Big\{\sup_{\omega}|S_1(\omega,T)|^2\Big\} \le \sum_{l=0}^{L}\mathrm{E}(Z_l^2) + 2\sum_{r=1}^{L-1}\Big\{\sum_{l=0}^{L-|r|}\mathrm{E}(Z_l^2)\mathrm{E}(Z_{l+r}^2)\Big\}^{\frac{1}{2}}.$$

Since

$$\begin{aligned}\mathrm{E}(Z_l^2) &= \mathrm{E}\Big\{\int_{l\Delta}^{(l+1)\Delta}\int_{l\Delta}^{(l+1)\Delta} dZ(\tau)dZ(t)\Big\}\\ &= \int_{l\Delta}^{(l+1)\Delta}\lambda(t)dt \le J\Delta,\end{aligned}$$

we have

$$\mathrm{E}\Big\{\sup_{\omega}|S_1(\omega,T)|^2\Big\} \le J\Delta\Big\{L + 2\sum_{r=0}^{L-1} r^{\frac{1}{2}}\Big\} \le J_1\Delta L^{\frac{3}{2}}$$

where $J_1 < \infty$ is a further constant.

To estimate the second sum in (A.3), observe that

$$d|Z(t)| \le |dZ(t)| \le \lambda(t)dt + dN(t),$$

so

$$\mathrm{E}\{d|Z(t)|\} \le 2\lambda(t)dt \le 2Jdt.$$

Using this estimate and the inequalities

$$|e^{i\omega u} - 1| \le |\omega u| \le \Omega_T u \quad (u \ge 0)$$

we obtain

$$\mathrm{E}\{\sup_{0\le\omega\le\Omega_T} |S_2(\omega,T)|\} \le \mathrm{E}\Big\{\sup_{0\le\omega\le\Omega_T} \sum_{l=0}^{L-1} \int_0^\Delta |e^{i\omega u} - 1| d|Z(u+l\Delta)|\Big\}$$

$$\le JL\Delta^2 \mathrm{E}(\Omega_T).$$

Combining estimates for both sums, and writing $L = T/\Delta$, we find

$$\mathrm{E}(M_T/T) \le T^{-1}\Big[\mathrm{E}\Big\{\sup_{0\le\omega\le\Omega_T} |S_1(\omega,T)|^2\Big\}\Big]^{\frac{1}{2}} + T^{-1}\mathrm{E}\Big\{\sup_{0\le\omega\le\Omega_T} |S_2(\omega,T)|\Big\} \le J_1^{\frac{1}{2}}(T\Delta)^{-\frac{1}{4}} + J\Delta\mathrm{E}(\Omega_T),$$

where Δ is still at our disposal. Under Assumption 6 that $E(\Omega_T) = O(T^{1-\delta})$ ($\delta > 0$), and taking $\Delta = O(T^{-\nu})$, we find an optimum choice of ν is $1 - 4\delta/5$ by equating the order of $(T\Delta)^{-\frac{1}{4}}$ and $\Delta\mathrm{E}(\Omega_T)$. And this leads to $\mathrm{E}(M_T/T) = O(T^{-\delta/5})$.

It remains to show that from this bound on the expectations we can deduce the almost sure convergence to 0 of the ratios M_T/T. Making use of Chebyshev's inequality and the Borel-Cantelli lemma, we can find a sequence of times T_k, $T_k \to \infty$, and a sequence of positive numbers α_k, $\alpha_k \to 0$, such that $T_k/T_{k+1} \to 1$ and $\Pr(M_{T_k}/T_k \le \alpha_k$ for all but a finite number of k$) = 1$. For example, the choice of $\alpha_k = k^{-\beta}$ and $T_k = k^\gamma$, where $\beta > 0$ and $\gamma > 5(1+\beta)/\delta$, will suffice since by Chebyshev's inequality

$$\Pr(M_{T_k}/T_k \le \alpha_k) \ge 1 - \mathrm{E}(M_{T_k}/T_k)/\alpha_k = 1 - O(T_k^{-\delta/5})/k^{-\beta} = 1 - O(k^{-\gamma\delta/5+\beta}),$$

and $-\gamma\delta/5 + \beta < -1$, then

$$\sum_{k=1}^{\infty} \Pr(M_{T_k}/T_k \le \alpha_k) = \infty,$$

and thus

$$\Pr(M_{T_k}/T_k \leq \alpha_k \text{ for all but a finite number of k}) = 1.$$

With such a choice, it is clear that $M_{T_k}/T_k \to 0$ (a.s.). For intervening values of T, $T_k < T \leq T_{k+1}$, we can write

$$M_T/T \leq M_{T_k}/T_k + (1/T_k)\sup_\omega \Big| \int_{T_k}^T e^{i\omega t}\lambda(t)dt \Big| + (1/T_k)\sup_\omega \Big| \int_{T_k}^T e^{i\omega t} dN(t) \Big|$$

The first term we already know tends to 0 almost surely, while for the second term we have

$$(1/T_k)\sup_\omega \Big| \int_{T_k}^T e^{i\omega t}\lambda(t)dt \Big| \leq (1/T_k)J(T - T_k) \to 0, \quad \text{(a.s.)},$$

and for the last term we have, by the ergodic theorem and the assumption $T_{k+1}/T_k \to 1$, that

$$\begin{aligned}(1/T_k)\sup_\omega \Big| \int_{T_k}^T e^{i\omega t} dN(t) \Big| &\leq \{N(T_{k+1}) - N(T_k)\}/T_k \\ &= \{N(T_{k+1})/T_{k+1}\}\{T_{k+1}/T_k\} - N(T_k)/T_k \to m' - m' = 0 \quad \text{(a.s.)}\end{aligned}$$

where m' is the mean rate of occurrence of points. Hence $M_T/T \to 0$ (almost surely), and this completes the proof of the case when $m = 1$.

Now consider $m > 1$. Since

$$T^{-m} \sup_{0\leq\omega\leq\Omega_T} \Big| \int_0^T t^{m-1} e^{-i\omega t} dZ(t) \Big| = T^{-1} \sup_{0\leq\omega\leq\Omega_T} \Big| \int_0^T (\frac{t}{T})^{m-1} e^{-i\omega t} dZ(t) \Big|,$$

and by adding $(\frac{t}{T})^{m-1}$ in the integral in (A.3) and proceed the proof similarly, we should obtain the result that

$$T^{-m} \sup_{0\leq\omega\leq\Omega_T} \Big| \int_0^T t^{m-1} e^{-i\omega t} dZ(t) \Big| \to 0, \quad \text{(a.s.)}.$$

■

Proof of Proposition 8. In the search range that $O(T^{\delta'-1}) \leq \omega \leq \Omega_T$, Lemma 7 implies that $J_T^{(Z)}(\omega)$ in (2.6) converges to 0 uniformly, so $J_T^{(\lambda)}(\omega)$ is the dominant term in $J_T(\omega)$. Consider $m = 1$ in $J_T(\omega)$. It follows from (A.1) that $J_T^{(\lambda)}(\omega)$ converges to 0 uniformly for

$\omega \in \{\omega : |\omega - \omega_k| \geq cT^{-1+\gamma}, c \neq 0, \forall \gamma > 0$ and $O(T^{\delta'-1}) \leq \omega \leq \Omega_T\}$, so the local maxima of $J_T^{(\lambda)}(\omega)$ as well as the periodogram must occur in the neighborhood of ω_k where the width of the neighborhood is at most $O(T^{-1})$ ($k = 1, \ldots, K$). Under Assumption 5 that the distance of two frequencies should be wider than the order of $O(T^{-1})$, these neighborhoods of ω_k's are disjoint for large T. Note that by the definition of local maxima of the periodogram that their corresponding frequencies have to be well separated with distance $O(T^{-1+\gamma})$ ($\gamma > 0$), only one maximum of the periodogram should be considered in each such neighborhood of ω_k. So the K largest local maxima of the periodogram must occur in these neighborhoods of $\omega_1, \ldots, \omega_K$, with each one capturing one local maximum, namely, the frequency estimate $\hat{\omega}_{k,T}$ is in such neighborhood of ω_k, $k = 1, \ldots, K$. Since these neighborhoods tend to 0 as $T \to \infty$, it follows that $\hat{\boldsymbol{\omega}}_T \to \boldsymbol{\omega}$ (a.s.), as $T \to \infty$, namely, $\hat{\boldsymbol{\omega}}_T$ is a consistent estimate of $\boldsymbol{\omega}$.

Moreover, $|J_T(\omega)|$ obtains its kth maximum at $\hat{\omega}_{k,T}$ while $|J_T^{\lambda}(\omega)|$ has its kth maximum at ω_k ($k = 1, \ldots, K$), so for

$$|J_T(\hat{\omega}_{k,T})| \geq |J_T(\omega_k)| \geq |J_T^{(\lambda)}(\omega_k)| - |J_T^{(Z)}(\omega_k)| = |J_T^{(\lambda)}(\omega_k)| + R_1(T),$$

and

$$|J_T^{(\lambda)}(\omega_k)| \geq |J_T^{(\lambda)}(\hat{\omega}_{k,T})| \geq |J_T^{(\lambda)}(\hat{\omega}_{k,T}) + J_T^{(Z)}(\hat{\omega}_{k,T})| - |J_T^{(Z)}(\hat{\omega}_{k,T})| \geq |J_T(\hat{\omega}_{k,T})| + R_2(T),$$

where $R_1(T) = -|J_T^{(Z)}(\omega_k)|$ and $R_2(T) = -|J_T^{(Z)}(\hat{\omega}_{k,T})|$, and they tend to 0 almost surely. So

$$\lim_{T\to\infty} |J_T(\hat{\omega}_{k,T})| = \lim_{T\to\infty} |J_T^{(\lambda)}(\omega_k)| = A_k/2,$$

where the second equality is determined by (A.1). On the other hand, the direct calculation of $|J_T(\hat{\omega}_{k,T})|$ shows that

$$|J_T(\hat{\omega}_{k,T})| = \frac{A_k}{2}\Big|\frac{\exp[-i(\omega_k - \hat{\omega}_{k,T})T] - 1}{-i(\omega_k - \hat{\omega}_{k,T})T}\Big| + o(1),$$

so we have

$$\Big|\frac{\exp[-i(\hat{\omega}_{k,T}-\omega_k)T]-1}{-i(\hat{\omega}_{k,T}-\omega_k)T}\Big| \to 1, \quad a.s.,$$

and thus

$$(\hat{\omega}_{k,T}-\omega_k) = o(T^{-1}), \quad a.s..$$

■

Proof of Proposition 9. Since $dZ(t)$ is a process with independent and bounded variance increment, the random variables $(U, \mathbf{V}', \mathbf{W}', \mathbf{X}', \mathbf{Y}')'$ satisfy a central limit theorem. Take U for example. Let $U_l = \int_{(l-1)\Delta}^{l\Delta} dZ(t)$, $l = 1, \ldots, [T]$, where $[T]$ is the largest integer which does not exceed T, and $\Delta = T/[T]$. Then U_l's are independent with mean 0, and for any $\varepsilon > 0$,

$$\sum_{l=1}^{[T]} \Pr\{|U_l| \le \varepsilon s_T\} \ge \sum_{l=1}^{[T]} \{1 - \text{var}(U_l)/(\varepsilon^2 s_T^2)\} = [T] - 1/\varepsilon^2 \to \infty, \quad \text{as } T \to \infty,$$

where $s_T^2 = \sum_{l=1}^{[T]} \text{var}(U_l) = T\text{var}(U)$ which tends to ∞ as $T \to \infty$. So

$$\Pr\{|U_l| \le \varepsilon s_T, \text{ for all but a finite number of l}\} = 1.$$

Since $s_T^2 \to \infty$ as $T \to \infty$, the *Lindeberg condition* (Roatgi and Saleh (2001), p.298)

$$\lim_{T\to\infty} \frac{1}{s_T^2} \sum_{l=1}^{[T]} \text{E}(U_l^2 : |U_l| > \varepsilon s_T) = 0$$

is satisfied. So $\sum_{l=1}^{[T]} U_l/s_T = T^{1/2}U/\{T\text{var}(U)\}^{1/2}$ converges to the standard normal distribution as $T \to \infty$. This central limit theorem easily extends to the random vector $(U, \mathbf{V}', \mathbf{W}', \mathbf{X}', \mathbf{Y}')'$. We omit the proof here. The asymptotic variance-covariance matrix of above random vector can be obtained by direct calculation. For example,

$$\begin{aligned}
\text{cov}(V_k, W_{k'}) &= T^{-1} \int_0^T \int_0^T \cos(\omega_k t)\ \sin(\omega_{k'}\tau)\ \text{cov}(dZ(t), dZ(\tau)) \\
&= T^{-1} \int_0^T \cos(\omega_k t)\ \sin(\omega_{k'} t)\ \lambda(t)dt \\
&\to \sum_{j=1}^{K} \frac{A_j}{4} \sin(\phi_j)\ (-\delta_{j,k+k'} + \delta_{j,k-k'} - \delta_{j,k'-k}) = E_3(k,k'), \quad \text{as } T \to \infty.
\end{aligned}$$

■

Proof of Theorem 10. The main steps to obtain (2.12) are shown before Theorem 10.

The expansion of $2\pi T I'_T(\omega)$ at $\omega = \omega_k$ is the summation of the two terms (a) and (b) as follows,

(a) :=

$$2\Big\{\int_0^T \sin(\omega_k t)\ \lambda(t)dt + \int_0^T \sin(\omega_k t)\ dZ(t)\Big\}\Big\{\int_0^T t\cos(\omega_k t)\ \lambda(t)dt + \int_0^T t\cos(\omega_k t)\ dZ(t)\Big\}$$
$$= 2\{-\frac{1}{2}TA_k\sin(\phi_k) + O(1) + T^{\frac{1}{2}}W_k\}\{\frac{1}{4}T^2A_k\cos(\phi_k) + O(T) + T^{\frac{3}{2}}X_k\},$$
$$= -\frac{1}{4}T^3A_k^2\sin(\phi_k)\cos(\phi_k) - T^{\frac{5}{2}}A_kX_k\sin(\phi_k) + \frac{1}{2}T^{\frac{5}{2}}A_kW_k\cos(\phi_k) + O(T^2),$$

and

(b) :=

$$2\Big\{\int_0^T \cos(\omega_k t)\ \lambda(t)dt + \int_0^T \cos(\omega_k t)\ dZ(t)\Big\}\Big\{-\int_0^T t\sin(\omega_k t)\ \lambda(t)dt - \int_0^T t\sin(\omega_k t)\ dZ(t)\Big\}$$
$$= 2\{\frac{1}{2}TA_k\cos(\phi_k) + O(1) + T^{\frac{1}{2}}V_k\}\{\frac{1}{4}T^2A_k\sin(\phi_k) + O(T) - T^{\frac{3}{2}}Y_k\}$$
$$= \frac{1}{4}T^3A_k^2\sin(\phi_k)\cos(\phi_k) - T^{\frac{5}{2}}A_kY_k\cos(\phi_k) + \frac{1}{2}T^{\frac{5}{2}}A_kV_k\sin(\phi_k) + O(T^2).$$

We can see that the terms with order $O(T^3)$ in equation (a) and (b) are canceled out when adding (a) and (b) together. And the terms with order $O(T^{5/2})$ are the products of one λ-integral and one Z-integral in the second equation of the expansion of (a) and (b), and they are the leading terms. So by taking the summation of (a) and (b), we obtain the value of $2\pi T I'_T(\omega)$ at $\omega = \omega_k$:

$$2\pi T I'_T(\omega_k) = T^{\frac{5}{2}}A_k\{\frac{1}{2}V_k\sin(\phi_k) + \frac{1}{2}W_k\cos(\phi_k) - X_k\sin(\phi_k) - Y_k\cos(\phi_k)\} + o(T^{\frac{5}{2}}).$$

Analogously, we can obtain (2.11) by breaking it into λ-integrals and Z-integrals, All terms involving Z-integrals uniformly (in ω) converge to 0 (a.s.) by Lemma 7, and the leading terms are the products of λ-integrals. We omit the tedious calculation and just show the final

calculation of the products of λ-integrals here:

$$
\begin{aligned}
&2\pi T^{-3} I_T''(\bar{\omega}_{k,T}) \\
=&2T^{-4}\Big\{\Big(\int_0^T t\sin(\bar{\omega}_{k,T}t)\ \lambda(t)dt\Big)^2 - \Big(\int_0^T \cos(\bar{\omega}_{k,T}t)\ \lambda(t)dt\Big)\Big(\int_0^T t^2\cos(\bar{\omega}_{k,T}t)\ \lambda(t)dt\Big) \\
&\quad + \Big(\int_0^T t\cos(\bar{\omega}_{k,T}t)\ \lambda(t)dt\Big)^2 - \Big(\int_0^T \sin(\bar{\omega}_{k,T}t)\ \lambda(t)dt\Big)\Big(\int_0^T t^2\sin(\bar{\omega}_{k,T}t)\ \lambda(t)dt\Big)\Big\} \\
&+ o(1).
\end{aligned}
$$

By Proposition 8 we have $|\bar{\omega}_{k,T} - \omega_k| = o(T^{-1})$, and the results in the equations (A.1) and (A.2) imply that in the above equations of the products of λ-integrals, the limiting value will not change if $\bar{\omega}_{k,T}$ is replaced by ω_k. So by direct calculation, we have

$$
\begin{aligned}
&\lim_{T\to\infty} 2\pi T^{-3} I_T''(\bar{\omega}_{k,T}) \\
&\quad = A_k^2\{\frac{1}{8}\sin^2(\phi_k) - \frac{1}{6}\cos^2(\phi_k)\} + A_k^2\{\frac{1}{8}\cos^2(\phi_k) - \frac{1}{6}\sin^2(\phi_k)\} \\
&\quad = -\frac{1}{24}A_k^2.
\end{aligned}
$$

Equation (2.12) follows from (2.9) and Slutsky's theorem (Roatgi and Saleh (2001), page 270). Moreover, $\hat{\boldsymbol{\omega}}_T$ is asymptotically normally distributed because of the asymptotic normality of the random variables $(V_1, \ldots, V_K, W_1, \ldots, W_K, X_1, \ldots, X_K, Y_1, \ldots, Y_K)'$ by Proposition 9. The variance-covariance of $\hat{\boldsymbol{\omega}}_T$ is obtained by direct calculation. ■

Proof of Theorem 11. The discussions of (2.13) and (2.14) and thereafter form the proof of the asymptotic normality of $\hat{\boldsymbol{A}}_{\boldsymbol{T}}$ and $\hat{\boldsymbol{\phi}}_{\boldsymbol{T}}$. ■

Proof of Theorem 12. Since $N(T) = \int_0^T dN(t) = \int_0^T \lambda(t)dt + \int_0^T dZ(t)$, by the definition of U, it follows that $T^{\frac{1}{2}}(N(T)/T - B) = U + o(1) \xrightarrow{d} N(0, B)$.

Now we show that $\hat{\lambda}_T(t)$ is asymptotically non-negative for any $t > 0$. For every $0 < t \leq$

T, we find

$$\hat{\lambda}_T(t) - \lambda(t) = [N(T)/T - B] + [\sum_{k=1}^{K} (\hat{A}_{k,T} - A_k)\cos(\hat{\omega}_{k,T}t + \hat{\phi}_{k,T})] + [\sum_{k=1}^{K} A_k\{\cos(\hat{\omega}_{k,T}t + \hat{\phi}_{k,T}) - \cos(\omega_k + \phi_k)\}].$$

On the right hand side of the above equation, the terms in the first two square brackets are both $O(T^{-\frac{1}{2}})$ following Theorems 11 and 12, and a Taylor expansion of the cosine function in the third term together with Theorems 10 and 11 yield the result that the third term is also $O(T^{-\frac{1}{2}})$. So $\hat{\lambda}_T(t) = \lambda(t) + O(T^{-\frac{1}{2}})$, and thus $\hat{\lambda}_T(t)$ is asymptotically non-negative given that $\lambda(t) \geq 0$.

■

Proof of Theorem 13. The results in Theorem 13 follow by expressing all the estimates as functions of $U, \mathbf{V}, \mathbf{W}, \mathbf{X}, \mathbf{Y}$ and using Proposition 9. ■

Proof of Theorem 14. Under Assumption 4 that the process is a non-homogeneous Poisson process,

$$\Pr(T_{n+1} = t | T_n = t_n, \ldots, T_1 = t_1) = \lambda(t) e^{-\int_{t_n}^{t} \lambda(s) ds} = \lambda(t) e^{-\Lambda(t) + \Lambda(t_n)}, \quad t > t_n. \quad \text{(A.4)}$$

So $\hat{T}_{n+1}$ is easily obtained by (A.4) and taking integration by parts. The mean squared error ν_n is obtained by (A.4) and using the fact that

$$\nu_n = \mathrm{E}_{T_n}[\mathrm{E}\{(T_{n+1} - \hat{T}_{n+1})^2 | T_n\}] = \mathrm{E}_{T_n}\{\mathrm{E}(T_{n+1}^2 | T_n) - \hat{T}_{n+1}^2\}.$$

■

Appendix B

Proofs in Chapter 3

Proof of Proposition 15. As shown in Severini (2000) on page 106, the MLE of the model parameter θ is consistent if the two conditions are satisfied:

1. The parameter space Θ is a compact subset of $\Re^d$ where d is the dimension of Θ.

2. $\sup_{\theta\in\Theta} |n^{-1}\ell(\theta) - \gamma(\theta)| \xrightarrow{P} 0$ as $n \to \infty$, where $\ell(\theta)$ is the log-likelihood function based on n observations and $\gamma(\theta) = n^{-1}\mathrm{E}\{\ell(\theta); \theta_0\}$.

In our model, we assume T to be a fixed quantity for mathematical convenience, so the regularity conditions for the consistency of MLE in our model is similar as above conditions with n replaced with T. Under the assumptions that $\lambda(t) \in [\varepsilon, M]$ for any $t > 0$ where $M > \varepsilon > 0$ and ω_k's are bounded as in assumption 6, the parameter space in our model is compact, and thus condition 1 is satisfied.

We now turn to the second condition. Denote the true intensity function of a Poisson

process as $\lambda(t;\theta_0) = \lambda_0(t)$, then the log-likelihood of a Poisson process at any θ is

$$\begin{aligned}
\ell(\theta) &= \sum_{j=1}^{N(T)} \log \lambda(t_j) - \int_0^T \lambda(t)dt \\
&= \int_0^T \log \lambda(t) dN(t) - \int_0^T \lambda(t)dt \\
&= \int_0^T [\lambda_0(t) \log \lambda(t) - \lambda(t)]dt + \int_0^T \log \lambda(t) dZ(t) \\
&= (i) + (ii).
\end{aligned} \tag{B.1}$$

The second last equation is followed by the decomposition

$$dN(t) = \lambda_0(t)dt + dZ(t), \tag{B.2}$$

where $dZ(t)$ is a process with mean 0, and independent but non-stationary increments. In addition, $dZ(t)$ is mean-squared bounded. So we have $\mathrm{E}\{\ell(\theta)\} = \mathrm{E}\{(i) + (ii)\} = (i)$. Since $\mathrm{var}((ii)) = \int_0^T [\log \lambda(t)]^2 \lambda_0(t)dt \leq MT \max\{(\log \varepsilon)^2, (\log M)^2\} = O(T)$, we have $(ii) \leq O(T^{1/2})$. So condition 2 is satisfied:

$$\sup_\theta \{\frac{1}{T}\ell(\theta) - \frac{1}{T}(i)\} = \sup_\theta \frac{1}{T}(ii) \xrightarrow{P} 0, \quad \text{as } T \to \infty.$$

And thus the MLE in our model is consistent. We now discuss the convergence rate of the frequency estimates.

Notice that (i) obtains its maximum at $\theta = \theta_0$, and $\ell(\theta)$ obtains its maximum at $\theta = \hat{\theta}$. It follows from condition 2 that $\frac{1}{T}\ell(\hat{\theta}) - \frac{1}{T}(i)|_{\theta=\theta_0} \xrightarrow{P} 0$, as $T \to \infty$. And that is, $\frac{1}{T}(i)|_{\theta=\hat{\theta}} + \frac{1}{T}(ii)|_{\theta=\hat{\theta}} - \frac{1}{T}(i)|_{\theta=\theta_0} \xrightarrow{P} 0$, as $T \to \infty$. Since $\frac{1}{T}(ii) \xrightarrow{P} 0$ at any $\theta \in \Theta$, so $\{\frac{1}{T}(i)|_{\theta=\hat{\theta}} - \frac{1}{T}(i)|_{\theta=\theta_0}\} \xrightarrow{P} 0$, that is

$$\begin{aligned}
&\frac{1}{T}(i)|_{\theta=\hat{\theta}} - \frac{1}{T}(i)|_{\theta=\theta_0} \\
=&\frac{1}{T}\int_0^T \lambda_0(t)[\log \hat{\lambda}(t) - \log \lambda_0(t)]dt - \frac{1}{T}\int_0^T [\hat{\lambda}(t) - \lambda_0(t)]dt \\
=&(iii) + (iv) \xrightarrow{P} 0, \quad \text{as } T \to \infty.
\end{aligned} \tag{B.3}$$

Consider (iv). By direct calculation,

$$(iv) = -\frac{1}{T}\int_0^T (\hat{B} - B_0)dt + o(1),$$

and since $\hat{B} - B_0 \xrightarrow{P} 0$ as $T \to \infty$, we have $(iv) \xrightarrow{P} 0$.

Since $(iii) + (iv) \xrightarrow{P} 0$ by equation (B.3), and $(iv) \xrightarrow{P} 0$ by the above discussion, we have $(iii) \xrightarrow{P} 0$. Now we will prove that in order to have $(iii) \xrightarrow{P} 0$, the frequency estimates must converge in the rate of $o(T^{-1})$.

It follows by a Taylor expansion of $\log \hat{\lambda}(t)$ around θ_0 that

$$\begin{aligned}
&(iii)\\
=&\frac{1}{T}\int_0^T \lambda_0(t)\Big[\frac{1}{\lambda_0(t)}\frac{\partial\lambda_0(t)}{\partial\theta}(\hat{\theta}-\theta_0) + \frac{1}{2}(\hat{\theta}-\theta_0)'\frac{\frac{\partial^2\lambda_0(t)}{\partial\theta^2}\lambda_0(t) - \big(\frac{\partial\lambda_0(t)}{\partial\theta}\big)^2}{\lambda_0^2(t)}(\hat{\theta}-\theta_0) + R_T(\theta)\Big]dt\\
&= \frac{1}{T}\int_0^T \frac{\partial\lambda_0(t)}{\partial\theta}(\hat{\theta}-\theta_0)dt + \frac{1}{T}\int_0^T \frac{1}{2}(\hat{\theta}-\theta_0)'\Big[\frac{\partial^2\lambda_0(t)}{\partial\theta^2} - \Big(\frac{\partial\lambda_0(t)}{\partial\theta}\Big)^2\frac{1}{\lambda_0(t)}\Big](\hat{\theta}-\theta_0)dt\\
&\qquad + \frac{1}{T}\int_0^T \lambda_0(t)R_T(\theta)dt,
\end{aligned} \tag{B.4}$$

where $R_T(\theta)$ is of smaller order of the second term in the Taylor expansion of $\log \hat{\lambda}(t)$. Consider $\theta = \omega_k$ in (B.4) while fixing other parameters. By using the results that $\frac{\partial\lambda(t)}{\partial\omega_k} = -tA_k\sin(\omega_k t + \phi_k)$ and $\frac{\partial^2\lambda(t)}{\partial\omega_k^2} = -t^2A_k\cos(\omega_k t+\phi_k)$, and by direct calculation, we have $\frac{1}{T}(\hat{\omega_k}-\omega_{k,0})\int_0^T \frac{\partial\lambda_0(t)}{\partial\omega_k}dt = (\hat{\omega_k}-\omega_{k,0})\times O(1)$, and $\frac{1}{T}(\hat{\omega_k}-\omega_{k,0})^2\int_0^T \frac{\partial^2\lambda_0(t)}{\partial\omega_k^2}dt = (\hat{\omega_k}-\omega_{k,0})^2\times O(T)$. In addition, since $1/\lambda(t) \in [1/M, 1/\varepsilon]$, we have

$$\begin{aligned}
&\frac{1}{T}(\hat{\omega_k}-\omega_{k,0})^2\int_0^T \frac{1}{\lambda_0(t)}\Big(\frac{\partial\lambda_0(t)}{\partial\omega_k}\Big)^2dt = (\hat{\omega_k}-\omega_{k,0})^2\frac{1}{T}\int_0^T \frac{1}{\lambda_0(t)}t^2A_{k,0}^2\cos^2(\omega_{k,0}t+\phi_{k,0})dt\\
&\in [(\hat{\omega_k}-\omega_{k,0})^2\frac{1}{T}\frac{1}{M}\int_0^T t^2A_{k,0}^2\cos^2(\omega_{k,0}t+\phi_{k,0})dt,\\
&\quad (\hat{\omega_k}-\omega_{k,0})^2\frac{1}{T}\frac{1}{\varepsilon}\int_0^T t^2A_{k,0}^2\cos^2(\omega_{k,0}t+\phi_{k,0})dt].
\end{aligned}$$

By simple calculation that $\int_0^T t^2\cos^2(\omega_k t+\phi_k)dt = \int_0^T t^2(\frac{1}{2}+\frac{\cos(2\omega_k t+2\phi_k)}{2})dt = O(T^3)$, we have $\frac{1}{T}(\hat{\omega_k}-\omega_{k,0})^2\int_0^T \frac{1}{\lambda_0(t)}\Big(\frac{\partial\lambda_0(t)}{\partial\omega_k}\Big)^2dt = (\hat{\omega_k}-\omega_{k,0})^2\times O(T^2)$.

To summarize the above discussion on (iii) when $\theta = \omega_k$, we have

$$(iii) = (\hat{\omega_k} - \omega_{k,0}) \times O(1) + (\hat{\omega}_k - \omega_{k,0})^2 \times (O(T) + O(T^2)) + \frac{1}{T}\int_0^T \lambda_0(t) R_T(\omega_k) dt.$$

Since $(\hat{\omega}_k - \omega_{k,0}) = o(1)$, the first term in (iii) is $o(1)$. In order to have $(iii) \to 0$ in probability, we need the sum of the last two terms to converge to 0 in probability. Note that $R_T(\omega_k)$ involves with the product of $o((\hat{\omega}_k - \omega_{k,0})^2)$ and higher order of derivatives of $\lambda(t)$ with respect to ω_k, and the corresponding integral, namely, the third term in (iii), can be calculated similarly as the second term in (iii). By direct calculation, we can show that the third term in (iii), namely $\frac{1}{T}\int_0^T \lambda_0(t) R_T(\omega_k)$, involves with terms such as $(\hat{\omega}_k - \omega_{k,0})^4 \times O(T^4)$. Since the second term in (iii) involves with $(\hat{\omega}_k - \omega_{k,0})^2 \times O(T^2)$, if $(\hat{\omega}_k - \omega_{k,0}) = O(T^{-1})$ or $(\hat{\omega}_k - \omega_{k,0}) = O(T^{-1+\delta})$ $(0 < \delta < 1)$, the sum of the last two terms cannot converge to 0 in probability. And thus we prove that $(\hat{\omega}_k - \omega_{k,0}) = o(T^{-1})$.

■

Proof of Proposition 16. The main technique in showing the asymptotic normality of $\hat{\theta}$ is the Taylor expansion of the first derivative of $\ell(\theta)$ in equation (3.13) and the decomposition of the log-likelihood function (B.1). Note that $\frac{\partial(i)}{\partial\theta}\Big|_{\theta=\theta_0} = 0$, we have

$$\begin{aligned} 0 = \frac{\partial \ell}{\partial \theta}\Big|_{\theta=\hat{\theta}} &= \frac{\partial \ell}{\partial \theta}\Big|_{\theta=\theta_0} + \frac{\partial^2 \ell}{\partial \theta^2}\Big|_{\theta=\bar{\theta}}(\hat{\theta} - \theta_0) \\ &= \frac{\partial (ii)}{\partial \theta}\Big|_{\theta=\theta_0} + \Big(\frac{\partial^2 (i)}{\partial \theta^2} + \frac{\partial^2 (ii)}{\partial \theta^2}\Big)\Big|_{\theta=\bar{\theta}}(\hat{\theta} - \theta_0), \end{aligned} \tag{B.5}$$

where $0 \leq \| \bar{\theta} - \theta_0 \| \leq \| \hat{\theta} - \theta_0 \|$. In addition, the second derivative of (i) is of higher order and thus it is the leading term in $\frac{\partial^2 \ell(\theta)}{\partial \theta^2}\Big|_{\theta=\bar{\theta}}$. This can be verified by direct calculation. For example, consider the derivative with respect to ω_k only, we have

$$\frac{\partial^2 (i)}{\partial \omega_k^2}\Big|_{\theta=\bar{\theta}} \simeq \frac{\partial^2 (i)}{\partial \omega_k^2}\Big|_{\theta=\theta_0} = -\int_0^T \frac{1}{\lambda(t)}\Big(\frac{\partial \lambda(t)}{\partial \omega_k}\Big)^2\Big|_{\theta=\theta_0} dt = O(T^3),$$

and $\frac{\partial^2 (ii)}{\partial \omega_k^2}$ has mean 0 and variance

$$\text{var}(\frac{\partial^2 (ii)}{\partial \omega_k^2}\Big|_{\theta=\bar{\theta}}) \simeq \text{var}(\frac{\partial^2 (ii)}{\partial \omega_k^2}\Big|_{\theta=\theta_0}) = \int_0^T \Big(\frac{\partial^2 \log \lambda_0(t)}{\partial \omega_{k,0}^2}\Big)^2 \lambda_0(t) dt = O(T^5).$$

In above approximation, $0 \leq |\bar{\omega_k} - \omega_{k,0}| \leq |\omega_k - \omega_{k,0}| = o(T^{-1})$ is needed because of the results in (A.1) and (A.2). The procedure to obtain the magnitude of above two equations are similar to the procedures discussed in the proof of Proposition 15. So $\frac{\partial^2(ii)}{\partial\omega_k^2}\Big|_{\theta=\bar{\theta}}$ is of order of $O(T^{\frac{5}{2}})$ and it is smaller than $\frac{\partial^2(i)}{\partial\omega_k^2}\Big|_{\theta=\bar{\theta}} = O(T^3)$. When the derivative is with respect to other parameters, the proof of the results that the second derivative of (i) is the leading term in $\frac{\partial^2\ell(\theta)}{\partial\theta^2}\Big|_{\theta=\bar{\theta}}$ can be obtained similarly. In addition, the value of $\frac{\partial^2(i)}{\partial\theta^2}$ evaluated at $\theta = \bar{\theta}$ is approximately the same as the value evaluated at $\theta = \theta_0$.

In addition, $\frac{\partial(ii)}{\partial\theta}\Big|_{\theta=\theta_0}$ follows a central limit theorem and $D_T^{-1}\frac{\partial(ii)}{\partial\theta}\Big|_{\theta=\theta_0}$ is asymptotically normally distributed with mean 0 and variance-covariance matrix Σ as shown in (3.15). Then we obtain the asymptotic normality of $\hat{\theta} - \theta$ by

$$D_T(\hat{\theta} - \theta_0) = \Big\{D_T^{-1}\frac{\partial^2\ell}{\partial\theta^2}\Big|_{\theta=\bar{\theta}}D_T^{-1}\Big\}^{-1}\Big\{D_T^{-1}\frac{\partial(ii)}{\partial\theta}\Big|_{\theta=\theta_0}\Big\}.$$

■

Bibliography

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Selected Papers of Hirotugu Akaike. Springer-Verlag, New York, pp. 199–213.

Bar-Shalom, Y., 1971. On the asymptotic properties of the maximum-likelihood estimate obtained from dependent observations. Journal of the Royal Statistical Society. Series B 33 (1), 72–77.

Bartlett, M. S., 1957. Discussion on 'symposium on spectral approach to time series'. J. Roy. Statisti. Soc. Ser. B 19, 1–63.

Bartlett, M. S., 1963. The spectral analysis of point processes. Journal of the Royal Statistical Society. Series B (Methodological) 25 (2), 264–296.

Bartlett, M. S., 1967. Inference and stochastic processes. J. Roy. Statisti. Soc. Ser. A 130, 457–477.

Bartlett, M. S., 1978. Correlation or spectral analysis? The Statistician 27, 147–158.

Bebbington, M., Zitikis, R., 2004. A robust heuristic estimator for the period of a Poisson intensity function. Methodology and Computing in Applied Probability 6, 441–462.

Berger, J. O., Pericchi, L. R., 2001. Objective bayesian methods for model selection: Introduction and comparison. IMS Lecture Notes-Monograph Series 38, 135–207.

Bohr, H., 1947. Almost-periodic functions. Chelsea, reprint.

Bozdogan, H., 1987. Model selection and akaike's information criterion (aic): the general theory and its analytical extensions. Psychometrika 52 (3), 345–370.

Brillinger, D. R., 1972. The spectral analysis of stationary interval functions. In: Cam, L. L., Neyman, J., Scott, E. L. (Eds.), Proc. Berkeley Symp. 6th. Vol. 1. University of California Press, Berkeley, California, pp. 483–513.

Brillinger, D. R., 1978. Comparative aspects of ordinary time series and of point processes. In: Krishnaiah, P. R. (Ed.), Developments in Statistics. Vol. 1. Academic Press, New York, pp. 33–133.

Brillinger, D. R., 1982. Asymptotic normality of finite Fourier transforms of stationary generalized processes. Journal of Multivariate Analysis 12, 64–71.

Brillinger, D. R., 1983. The finite Fourier transform of a stationary process. In: Brillinger, D. R., Krishnaiah, P. R. (Eds.), Handbook of statistics. Vol. 3. Elsevier Science Publishers B.V., New York, pp. 21–38.

Brillinger, D. R., 2003. Three environmental probabilistic risk problems. Statistical Science 18 (4), 412–421.

Brillinger, D. R., 2008. Extending the volatility concept to point processes. Journal of Statistical Planning and Inference 138, 2607 C 2614.

Burnham, K. P., Anderson, D. R., 2002. Model Selection and Multimodel Inference: A Practical Information - Theoretic Approach, 2nd Edition. Springer, New York.

Chen, S., September 2006. Statistical analysis of point processes and associated marks. Ph.D. thesis, University of California, Riverside.

Corduneanu, C., 1989. Almost Periodic Functions. Chelsea, New York.

Cox, D. R., Isham, V., 1980. Point Processes. Chapman and Hall.

Cox, D. R., Lewis, P. A. W., 1966. The Statistical Analysis of Series of Events. London:Methuen. New York: Barnes and Noble.

Crowder, M. J., 1976. Maimum likelihood estimation for dependent observations. Journal of the Royal Statistical Society Series B 38 (1), 45–53.

Durrett, R., 1999. Essentials of Stochastic Processes. Springer.

Engle, R. F., January 2000. The econometrics of ultra-high-frequency data. Econometrica 68 (1), 1–22.

Hall, P., Reimann, J., Rice, J., 2000. Nonparametric estimation of a periodic function. Biometrika 87 (3), 545–557.

Hall, P., Yin, J., 2003. Nonparametric methods for deconvolving multiperiodic functions. J. R. Statist. Soc. B 65, 869–886.

Hannan, E. J., 1973. The estimation of frequency. Journal of Applied Probability 10 (3), 510–519.

Hassan, M. Y., Lii, K.-S., 2006. Modeling marked point processes via bivariate mixture transition distribution models. Journal of the American Statistical Association 101 (475), 1241–1252.

Helmers, R., Mangku, I. W., 2003. On estimating the peirod of a cyclic poisson process. In: Moore, M., Froda, S., Lẽger, C. (Eds.), Mathematical Statistics and Applications: Festschrift for Constance van Eeden. Vol. 42 of Lecture Notes-Monograph Series. Institute of Mathematical Statistics, pp. 345–356.

Helmers, R., Mangku, I. W., 2009. Predicting a cyclic poisson process. Report PNA-E0917, Centre for Mathematics and Computer Science (CWI), Amsterdam, The Netherlands.

Helmers, R., Mangku, I. W., Zitikis, R., 2003. Consistent estimation of the intensity function of a cyclic Poisson process. Journal of Multivariate Analysis 84, 19–39.

Helmers, R., Mangku, I. W., Zitikis, R., 2005. Statistical properties of a kernel-type estimator of the intensity function of a cyclic Poisson process. Journal of Multivariate Analysis 92, 1–23.

Helmers, R., Mangku, I. W., Zitikis, R., 2007. A non-parametric estimator for the doubly periodic poisson intensity function. Statistical Methodology 4, 481–492.

Hurvich, C. M., Tsai, C.-L., 1989. Regression and time series model selection in small samples. Biometrika 76 (2), 297–307.

Imoto, M., Maeda, K., Yoshida, A., 1999. Use of statistical models to analyze periodic seismicity observed for clusters in the Kanto Region, central Japan. Pure appl. geophys 155, 609–624.

Kuhl, M. E., 1994. Estimation and simulation of nonhomogeneous Poisson processes having multiple periodicities. Master's thesis, Department of Industrial Engineering and Graduate Probgram in Operations Research, North Carolina State University, Raleigh, North Carolina.

Kuhl, M. E., Wilson, J. R., Johnson, M. A., 1995. Estimation and simulation of nonhomogneous Poisson processes having multiple periodicities. In: C. Alexopoulos, K. Kang, W. L., Goldman, D. (Eds.), Proceedings of the 1995 Winter Simulation Conference. Institute of Electrical and Electronics Engineers, Piscataway, NJ, pp. 374–383.

Kuhl, M. E., Wilson, J. R., Johnson, M. A., 1997. Estimating and simulating Poisson processes having trends or multiple periodicities. IIE Transactions 29, 201–211.

Lee, S., Wilson, J. R., Crawford, M. M., 1991. Modeling and simulation of a nonhomogeneous Poisson process having cyclic behavior. Communications in Statistics - Simulation and Computation 20 (2/3), 77–809.

Lewis, P. A. W., 1970. Remarks on the theory, computation and application of the spectral analysis of series of events. J. Sound Vib. 12 (3), 353–375.

Lewis, P. A. W., 1972. Recent results in the statistical analysis of univariate point processes. In: Lewis, P. A. W. (Ed.), Stochastic point processes: statistical analysis, theory, and applications. Wiley, New York, p. 1:54.

Lewis, P. A. W., Shedler, G. S., 1979. Simulation of nonhomogeneous poisson processes by thinning. Naval Res. Logist. Quart. 26 (3), 403–413.

Lii, K.-S., Masry, E., 1994. Spectral estimation of continuous-time stationary processes from random sampling. Stochastic Processes and Their Applications 52, 39–64.

Lu, Y., Garrido, J., 2005. Doubly periodic non-homogeneous poisson models for hurricane data. Statistical Methodology 2, 17–35.

Mangku, I. W., 2001. Estimating the intensity of a cyclic Poisson process. Ph.D. thesis, University of Amsterdam, Amsterdam.

Nayak, T. K., Bose, S., Kundu, S., 2008. On inconsistency of estimators of parameters of non-homogeneous poisson process models for software reliability. Statistics and Probability Letters 78, 2217–2221.

Priestley, M. B., 1981. Spectral Analysis and Time Series. Academic Press.

Rice, J. A., Rosenblatt, M., 1988. On frequency estimation. Biometrika 75 (3), 477–484.

Roatgi, V. K., Saleh, A. K. M. E., 2001. An introduction to probability and statistics, 2nd Edition. John Wiley and Sons, Inc.

Ross, S. M., 2007. Introduction to Probability Models, 9th Edition. Academic Press.

Schwarz, G., 1978. Estimating the dimension of a model. The Annals of Statistics 6 (2), 461–464.

Severini, T., 2000. Likelihood methods in statistics. Oxford University Press, Oxford; New York.

Shibata, R., 1989. Statistical aspects of model selection. In: Willems, J. (Ed.), From Data to Model. Springer-Verlag, London, pp. 215–240.

Takeuchi, K., 1976. Distribution of informational statistics and a criterion of model fitting. Suri-Kagaku (Mathematic Sciences) 153, 12–18, in Japanese.

Vere-Jones, D., 1982. On the estimation of frequency in point-process data. Journal of Applied Probability 19, 383–394.

Vere-Jones, D., Ozaki, T., 1982. Some examples of statistical estimation applied to earthquake data. Ann. Inst. Statist. Math 34, 189–207.

Walker, A. M., 1971. On the estimation of a harmonic component in a time series with stationary independent residuals. Biometrika 58, 21–36.