UNIVERSITY OF NAIROBI

College of Biological and Physical Sciences

School of Mathematics

MODELLING TELEPHONE CALL CENTER ARRIVALS AS A NONHOMOGENOUS POISSON PROCESS WITH CYCLIC RATE

by

Andrew W. Nyamu

A PROJECT SUBMITTED TO THE SCHOOL OF MATHEMATICS IN PARTIAL FULFILLMENT FOR THE DEGREE OF Master of Science in Social Statistics

June 2014

Declaration

This project is my original work and has not been presented for the award of a degree in any other University.

Signature: .... Date: ....

ANDREW W. NYAMU
(I56/80666/2012)

This project has been submitted with my approval as University supervisor.

Signature: .... Date: ....

PROF. MOSES M. MANENE
SCHOOL OF MATHEMATICS

Abstract

This project demonstrates that telephone call center call-arrival data can be effectively modelled as a nonhomogenous Poisson process with cyclic rate. The data was from an Israeli bank telephone call center collected through the year 1999 and made freely accessible online. The theory underlying the assumptions of a Poisson process has been presented where it has been shown that time-sampling a Poisson process results in a nonhomogenous Poisson process and the method of maximum likelihood presented and applied to estimate the parameters of an exponential-polynomial-trigonometric rate function (EPTF). An analysis of this data confirmed that a nonhomogenous Poisson process with an EPTF-type rate was a good fit with a confidence of 90%.

Key Words: *Nonhomogenous Poisson process (NHPP); Cumulative rate function; Mean Value function; Intensity function; Exponential-Polynomial-Trigonometric rate function (EPTF)*

Acknowledgments

Though this project is the author's personal work, it would not have been completed without the academic, spiritual, or moral support from many individuals.

First and foremost, I want to thank God the creator; Jesus Christ the saviour; and Holy Spirit the companion & comforter, for allowing me to live in this world, blessing me with knowledge, and guiding me throughout my MSc candidature.

Next, I would like to express my heartfelt thanks to my supervisor Prof. Moses Manene, who has been a critical reviewer of my work, very meticulous, diligent, the first filter for all my forays into this project. I am motivated by your passion for research and academic excellence. Thanks for your keen interest in seeing me through this project.

I also want to thank Prof. Peter Waiganjo of the School of Computing and Informatics, University of Nairobi. Thank you for sowing the seeds for this project by initially interesting me with Traffic modelling of a matatu stage. Were it not for the lack of data and the short period allowed for the project work, I would have happily proceeded with it.

I would also like to thank Prof. Avishai Mandelbaum and the Industrial Engineering and Management team at the Technion University, Haifa, Israel, for making the Anonymous Bank Telephone Call center data used in this project publicly accessible. In the same line I would like to thank Prof. J. R. Wilson, a distinguished faculty member of the Department of Industrial and Systems Engineering, North Carolina State University for publicly availing software used in solving the system of m+4 equations in Chapter 3 and for silently challenging me to write the software for simulating an NHPP with cyclic rate by not providing the complementary simulation package.

The support of two very important people in my life, my father, David N. Nyamu, and mother, Alice W. Nyamu, cannot be overstated. Thank you for raising me up until now. Without your love and sacrifices so far, I would not have been able to be what I am today.

I want to thank my dear siblings, Janette Wandia, Joan Njeri , Elizabeth Wanjiku, James Njoka & Grace Nyakio and brothers in law Edward Kipsang & the late James Maina Ichagichu for every way you have shaped me - it has not been a bed of roses but I know you love me dearly. Thank you!

Many thanks to my friends Wallace Muchiri, Ego Obi, Richard Ngamita, Hongli Lv, Elliott Friedman, and Rodrigo Ribeiro for all the support you have given me at one time or another. I look forward to meeting with you some day in future if God allows.

Finally, and very importantly, I would like to thank my mentors at Google SRE-Dub & especially Mr. John Looney for rousing my academic interest so that I decided to advance my studies.

Dedication

I wish dedicate this project to my parents,

Mr & Mrs. David N. Nyamu,

&

small siz,

Grace N. Nyamu.

Contents

List of Figures

THIS PAGE INTENTIONALLY LEFT BLANK

List of Tables

Chapter 1

Introduction

Modelling is the process of generating conceptual representations which can be more easily manipulated for the purpose of understanding underlying relationships of economic or other interest. The resulting models which come closest to observed reality are then applied in areas such as forecasting/prediction, training/education, planning and decision making to name a few.

Businesses which attract large customer numbers automate and decentralise service delivery to their customers in a bid to provide scalable service. One way of doing this is to provide telephone call centers where a customer need only make a telephone call and receive service without being physically present to transact. Having a call center has the following challenges:

a) Authentication: Is the caller the true owner of the identity?;

b) Accounting: Can a complete history of all transactions with the system be reconstructed for each caller?;

c) Authorisation: What transactions can be completed over the phone?; and

d) Data management: Where and how to store the vast amounts of data collected.

The opportunities that a call centre creates are:

a) Automation: Thousands of computer software agents can be easily employed round the clock to serve callers; and

b) Advertising & Entertainment: Customers in queue can receive adverts and be entertained as they wait to be served.

Brown et al. (2005)[1] described telephone call centers as technology-intensive operations where the largest percentage of cost of operations goes to human resource thus necessitating adherence to a sharply defined balance between agent efficiency and service quality for well-run call centers.

In this project, we shall focus on the customers' interaction with the system, specifically in the arrival of calls to the call center - the business wants to learn most about the customer! According to Gans et al. (2003), large call centers generate vast amounts of data and make use of special purpose computers to mediate call flow and records generated at each stage that can in theory for each caller be used to reconstruct a detailed call history. They also cited some reasons why call centers do not store or analyse such records, namely:

a) Historically high cost of maintaining large databases (data management problem);

b) Software for managing such databases only performs rudimentary (summary) statistics; and

c) Ignorance on how and why more detailed analysis should be carried.

A telephone call center arrival can be described as any call made to the call center over the telephone network. Historically, arrival counts have been shown to conform to a Poisson process whose parameter is the rate. In this project, arrival counts will be extracted from the call accounting records of an Anonymous bank[1] in Israel.

Funaki and Matoba (1999)[3] noted that the Poisson process was a model that had seen successful usage in modelling many actual systems such as patient arrivals

[1]The bank is called 'Anonymous' for confidentiality reasons.

at a hospital, database transactions and occurrences of typhoons. However, the restrictions in the properties of the Poisson process made it inadequate for many other real world systems which possessed the characteristics of the Poisson process but with its parameters dependent on a time-point t and the length of the interval τ - such systems they described as nonhomogenous (or nonstationary) Poisson processes (NHPP).

1.1 Statement of the Problem

Gans et al. (2003)[7] noted that accurate estimation and forecasting of parameters were prerequisites for consistent service levels and efficient operation and that though a lot had been done in statistical inference and forecasting, comparatively little had been devoted to stochastic processes, especially in the area of queueing and call centers particularly.

The purpose of this project will be to investigate how well the nonhomogenous Poisson process explains call center data but with a rate that exhibits cyclic and seasonal trends. It is hoped that this will add to our understanding of the stochastic processes applicable to call center data.

1.2 Objectives

The main objective of this project will be to generate a nonhomogenous Poisson process model with a rate function that fits the observed trend in the Anonymous Bank call center data. This will be accomplished by achieving the following specific objectives:

- Justify arrival count data as Poisson processes;
- Test call center arrival data for the properties of a Poisson process;
- Estimate the rate parameter to arrive at a model;

- Conduct a goodness-of-fit test for the model.

1.3 General Outline of the Project

Chapter 2 will largely be a review of important papers that prepare the motivation for this project. A summary of research gaps will be provided at the end. Chapter 3 will discuss the call center data and justify arrival count data as a Poisson process. The characteristics of Poisson processes and nonhomogenous Poisson processes will also be discussed in this section. The method of Maximum Likelihood Estimation (MLE) for the rate function of the nonhomogenous Poisson process will be presented. A few tests whether call center arrival data comes from a nonhomogenous Poisson process and MLE (numerically using software) will be conducted leading to the estimation of the rate parameter. A goodnes-of-fit test will afterwards be performed. Chapter 4 will discuss the results of the goodness-of-fit test and compare a few others and finally, Chapter 5 will conclude the project and give further direction for future work.

Chapter 2

Literature Review

A review of existing papers that motivate and inform this project are presented in this chapter. The previous works discussed here are in no way exhaustive but sufficient for conducting this project. A framework bringing together all the efforts of this literature review is added in the last section.

2.1 Review of Past Work

Gans et al.,(2003)[7] offered an overview of the state of research on telephone call centers by first discussing how call centers functioned, then by surveying the research devoted to their operations. Large call centers generated vast amounts of data and make use of special purpose computers to mediate call flow - records were generated at each stage and thus in theory, a detailed call history for each caller could be reconstructed. They also cited some reasons why call centers do not store or analyse such records, namely:

a) Historically high cost of maintaining large databases (data management problem);

b) Software for managing such databases only performs rudimentary (summary) statistics; and

c) Ignorance on how and why more detailed analysis should be carried.

Further, they stated that the performance of call centers in peak hours could be extremely sensitive to changes in calling rate and service time, thus accurate estimation and forecasting of parameters would be prerequisites for consistent service level & efficient operation. They identified four categories of call center data:

a) Operational: reflects physical process by which calls are handled;

b) Marketing/Business data: record transactions that took place over a customers entire history with the company;

c) Human resource: record the history and profile of agents - skill of the agent data is fed into the computer assigning calls to the agent to enable the software to autonomously route request needing specialised skills to human agents with the skills; and

d) Psychological: collected from customer, agent or manager surveys, they record subjective perceptions of the service level & work environment.

In their discussion on data models, they made three sets of distinctions:

a) Descriptive models: these organise and summarise the data being analysed, eg. histograms;

b) Theoretical models: objectively test whether or not the phenomenon being observed conforms to known mathematical and statistical theories; and

c) Explanatory models: these fall in between the two above and are often created in the context of regression and time-series analysis. These models identify and capture relationships in terms of explanatory variables but do not develop or test mathematical theory to explain the relationships.

They also gave a clear distinction of estimation and prediction by stating that estimation is the use of existing data to make inferences about parameter values of a model whereas prediction is the use of estimated parameters to forecast the behaviour of a sample outside the original dataset.

Referring to future developments, the most pressing practical needs, in their opinion, was for:

a) Improvements in the forecasting of arrival rates: further development of models for estimation & prediction will depend, in part, on access to rich data sets. They held that the randomness of Poisson arrival rates would be explained by uncaptured covariates so the 'richer' the data the better;

b) Procedures for predicting waiting time;

c) Parallel, descriptive studies to validate or refute the robustness of initial findings;

d) Studies of the abandonment of queues to explain impatience;

e) Opportunity to further develop and extend the scope of explanatory models; and

f) Analysis of integrated operational, marketing, human resources and psychological data.

Brown et al.,(2005)[1] described a call center as a service network in which agents provided telephone services and customers seeking these services got delayed in tele-queues. Queueing theoretic models were exploited to ensure balance between agent efficiency and service quality where the inputs are system primitives such as number of agents working, rate of call arrivals, service time, waiting time and the outputs are performance measures such as distribution of time that customers wait "on hold" and the fraction of customers that abandon the queue before being served - the number of agents is a control parameter that can be manipulated to attain desired efficiency-quality trade-off.

They noted that common call center models & practice assumed that the arrival process was Poisson with a rate that remained constant for blocks of time (eg. half-hours), and with a separate queueing model fitted for each block of time. They argued that the time-inhomogenous Poisson process was a more natural model for capturing changes in arrival rate where the arrival rate function could be approximated as being piecewise constant. To test whether the process was homogenous within blocks, a subset of blocks was chosen and which covered the same time interval on various days were sampled. By postulating the null hypothesis that arrival rate is constant within the intervals, and using ordered arrival times, they gave a formula to construct independent standard exponential variables which were used throughout their discussion.

The test for the null hypothesis that arrivals of given types of calls form an inhomogenous Poisson process with piecewise constant rates was done as follows:

1. The duration of a day was broken up into relatively short blocks of time, short enough so arrival rate doesn't change significantly within a block;

2. Arrivals are then considered within a subset of blocks

The Kolmogorov-Smirnov test statistic was used in combination with the exponential Q-Q plots to ascertain goodness-of-fit to the exponential distribution.

They further discussed the statistical prediction of the system load based on a combination of observed arrival times to the system and service times as of great importance to any operations manager of a call center.

Brown et al.,(2004)[4] showed that assuming arrivals to a queue follows a Poisson process was not always valid in practice. Further, they developed statistical procedures to test that a stochastic process is an inhomogenous Poisson process and showed that call arrivals to a real-life call center followed a Poisson process with an inhomogenous arrival rate over time. The procedure they outlined is presented below:

1. Test that arrivals do not depend on exact time clock: they chose specific time intervals over the days of 5 months and performed a χ^2 test of uniformity.

2. Test that arrivals have no serial dependence: proposed independence by looking at correlation of various transformation of inter-arrival times.

3. Test of the exponentiality of inter-arrival times: for a homogenous Poisson process, inter-arrival times have an exponential distribution with scale $\frac{1}{\lambda}$. A χ^2 test was performed to show this.

4. Test to show that Poisson arrival rates are not easily "predictable": by hypothesising that the Poisson rate was a function of certain covariates eg. time of day, call type and day of the week and was ultimately rejected.

They derived statistical models that could be used to construct predictions of the inhomogenous arrival rate, and provided parameter estimates in the models. Realistic calculations or simulations of the performance of a queueing system could thus be constructed. They also constructed two tests of call center arrival processes designed to determine whether the process was inhomogenous Poisson.

Lee et al. (1991)[5] described techniques for identification, estimation and simulation of a nonhomogenous Poisson process (NHPP) whose rate function contained a cyclic component as well as a long-term evolutionary trend. The techniques there discussed could be applied whether the oscillation frequency was known or not. They modelled the instantaneous arrival rate using an exponential function whose exponent is the sum of polynomial and trigonometric components, ie., an exponential-polynomial-trigonometric function (EPTF). Maximum likelihood estimates of the unknown continous parameters of this function were obtained numerically, and the degree of the polynomial component determined by a likelihood ratio test. Spectral analysis of event series was used to gain an initial estimate of the oscillation frequency. A piecewise linear majorizing function (a piecewise linear function which provides a tight upper bound - thus a thinning algorithm) was used to approximate the fitted rate function as clearly as possible.

Their original motivation for the study was a simulation study conducted earlier on the effects of sea conditions and supply ship availability on petroleum exploration operations at an off-shore drilling site.

This paper will be the foundation on which an EPTF type rate function will be derived for the call center data.

Massey et al., (1996)[6] also investigated ways to estimate the parameters of a nonhomogenous Poisson process but with linear rate over a finite interval based on the number of counts in measurement sub-intervals. They compared 3 parameter estimation methods:

a) Ordinary least squares (OLS);

b) Iterative weighted least squares (IWLS); and

c) Maximum likelihood (ML);

all constrained to yield a non-negative rate function.

The theoretically optimal weighted least squares (TWLS) was used as the reference point. Overall, ML performed as well as IWLS and was found to be significantly more effective than OLS. Explicit formulas for OLS variances and the asymptotic TWLS variances (with increasing measurement intervals) revealed the statistical precision of the estimators and the influence of the parameters and method: in their words, "knowing how variance depends on interval length helps determine how to approximate general arrival-rate functions by piecewise linear ones."

They noted that the major difficulty with the nonhomogenous Poisson process was that it had infinitely many parameters, ie., it was parameterised by it's arrival rate $\lambda(t)$. They considered the linear Poisson process model because they wanted to:

a) Understand how different estimation procedures perform when the model is approximately valid.

b) Determine whether a linear model is appropriate, ie., a model is invalid if the arrival process is not poisson or if arrival-rate function over the designated sub-interval is non-linear.

Funaki & Matoba (1999)[3] posited that the parameters of the mean rate function $\Lambda(t)$ should be estimated in order to model a system with NHPP but it's parametric representation was oft unknown. They mentioned that Law & Kelton proposed a nonparametric estimation with piecewise constant intensity function requiring many sample datasets for statistically accurate values and contrasted it with their method which would be a piecewise estimation method covering the case of only a few sample data sets.

Their method gave a pseudo-estimator by using sample data of the adjacent intervals but could theoretically gain more statistical accuracy compared to the Law & Kelton nonparametric method and with fewer sample data sets.

They further defined the nonhomogenous (nonstationary) Poisson process as one where the Poisson distribution of the number of events had a mean dependent on both the time point t ant eh length of the interval τ, making it applicable to many real-world cases. The NHPP would be fully specified by the mean function, $\Lambda(t)$, thus the estimation for NHPP is equivalent to identifying the parameters of the mean rate function, $\Lambda(t)$. The method of maximum likelihood estimation for $\Lambda(t)$ was used when it's parametric representation was known before hand using some collected sample data; but in practice, its parametric representation is unknown for lack of knowledge about the determinants eg. cyclic factors.

They also discussed various smoothing algorithms ie. *a*) Local averaging; and *b*) Piecewise polynomial smoothing in the attempt at nonparametric regression without the Poisson process basis which their proposed estimator was completely drawn from but with similarities to local averaging algorithms.

Frenkel et al.,(2003)[2] considered a nonhomogenous Poisson process with an intensity function $\lambda(t)$ parameterized in two forms: log-linear with $\lambda(t) = \mathrm{e}^{\alpha+\beta t}$ and Weibull-type $\lambda(t) = \alpha^{\beta}\beta t^{\beta-1}$. The Method of Maximum Likelihood was applied for parameter estimation. They showed that there was one and only one solution to maximum likelihood equation of the log-linear form of $\lambda(t)$.

They considered the test of the hypothesis for two cases of the intensity function:

1. Case of the known intensity function: testing the hypothesis that the given sample path is a realization of NHPP could be carried out on the basis of: Let $t_1, t_2, ..., t_n$ be the instants of event occurences in NHPP. Consider the following time-transformed process: $T_1, T_2, ..., T_n$ where $T_i = \int_0^{t_i} \lambda(v)\,\mathrm{d}v$. This process is a standard Poisson process with $\lambda(t) = 1$ if the original process is NHPP. Therefore, the intervals between events in the transformed process formed a sample of i.i.d. standard exponential random variables.

2. Case of unknown intensity function: whose parameters are estimated from the sample path, it turned out that the intervals between adjacent times $\hat{W}_i = \int_0^{t_i} \hat{\lambda}(v)\,\mathrm{d}v$ were not i.i.d. exponential random variables. They demonstrated that their sum, i.e. the value of $\hat{W}_n$, was always equal to n.

They also suggested the following computer intensive procedure for testing the hypothesis that the sample path belongs to NHPP with intesity function equal to the estimated $\hat{\lambda}(t)$:

1. Simulate N trajectories of the NHPP with intensity function $\hat{\lambda}(t)$. Subject each trajectory to the time-transformation described above.

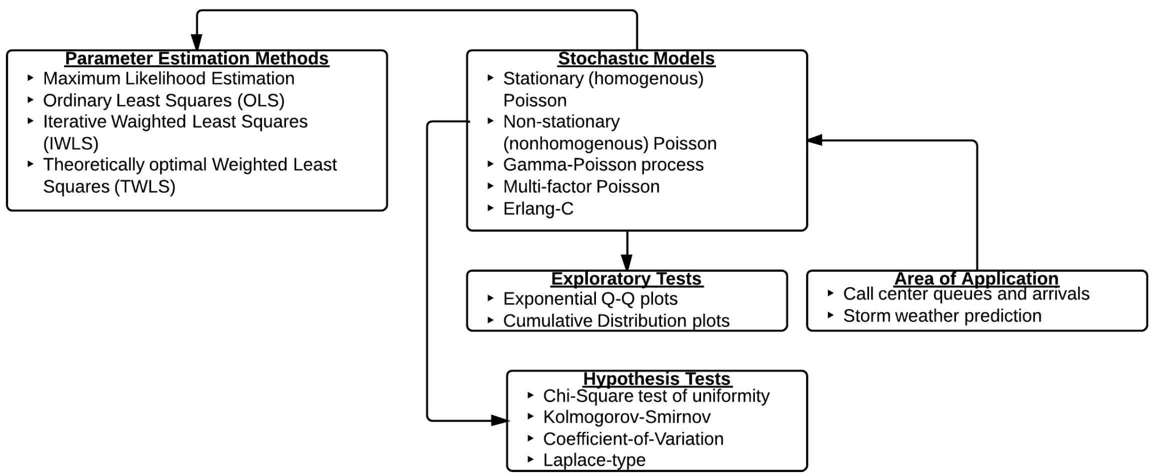
2. Compute for each simulation run the actual values of test statistics S_1, S_2, S_3 for the intervals between the adjacent transformed time instants. S_1 denotes Laplace-type statistic, S_2 denotes Kolmogorov-Smirnov statistic, and S_3 denotes the coefficient of variation of the transformed time intervals.

3. Based on 2, compute the upper and lower α-critical values for these statistics: $S_i(\alpha), S_i(1-\alpha)$.

4. For the given sample path, compute the actual transformed time instants and the actual values of the above statistics, S_1, S_2, S_3. Reject the hypothesis if at least one of these values fall outside the inverval $[S_i(\alpha), S_i(1-\alpha)]$, for i = 1, 2, 3.

2.2 Framework of Important Topics

Figure 2-1: Framework of important topics



Chapter 3

Methodology

In this chapter, we start with a description of data from the Israeli Anonymous Bank Call center data[1] made publicly accessible on the Technion website link given on the footnote. We also provide a quick overview of how this data was collected in the call center. Secondly, we present a justification to the use of Poisson related processes in modelling arrival data. The nonhomogenous Poisson process is then shown to be a generalised realisation of the homogenous Poisson process. The properties of both the homogenous (or stationary) Poisson and nonhomogenous (or nonstationary) Poisson process will be briefly discussed. Finally, the method of Maximum Likelihood for the parameters of a exponential-polynomial-trigonometric function (EPTF) will be presented.

[1]Faculty of Industrial Engineering and Management, Technion, Haifa. (2000, Feb). *12 month Telephone Call-Center for 'Anonymous Bank' in Israel, (Jan-1999)-(Dec-1999).* Retrieved February 12, 2014, from the Faculty of Industrial Engineering and Management, The Technion - Israel Institute of Technology, website: *http://ie.technion.ac.il/serveng/*

3.1 Description of Data

Sakov et. al.,(2001)[9] defined a call center as a service network which consisted of callers (customers), servers (or agents - these provided telephone-based services) and queues. Brown et. al.,(2005) discussed the mediation of calls through this call center in the following sequence:

a) A customer called one of several telephone numbers associated with the call center.

b) The customer would then be connected to a special purpose computer called a VRU/IVR[2].

c) A customer would be immediately forwarded to an available agent if the VRU didn't provide the required service, else, placed in a tele-queue to wait for an agent to become available.

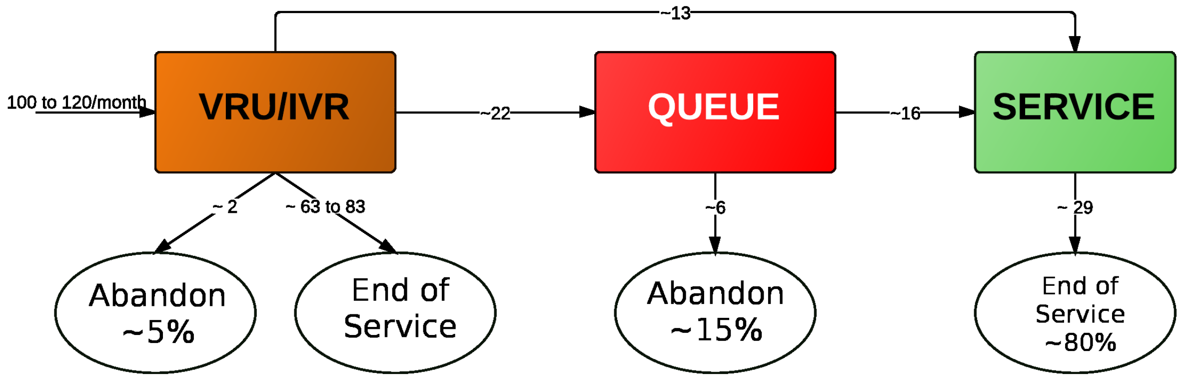
d) Customers in the tele-queue were served on a FCFS[3] basis and queue position distinguished by time of arrival to the queue.

Approximately 450,000 calls were made to the Anonymous Bank call center in 1999 and each call proceeding past the VRU conceptually passed through three stages which generated distinct data: *a*) **Arrival stage:** triggered by call's exit from VRU generating arrival time record; *b*) **Queueing stage:** call enters this stage if no appropriate server is available - data generated here was queue entry time, queue exit time and manner of queue exit; and *c*) **Service stage:** the caller was serviced - data generated was start and end time of service. Figure 3-1 below visually summarises the flow of calls through the call center system.

[2]VRU/IVR: Voice Response Unit/Interactive Voice Response: 65% of callers completed transactions at the VRU wheareas the rest indicated they wanted to speak to an agent

[3]FCFS: First Come First Served

Figure 3-1: Event history of an incoming call (Sakov (2001)[9]). Number of callers in thousands ('000).



3.2 Call Center Arrivals as a Poisson Process

I will present two approaches that justify the call center arrivals as a Poisson process:

(a) Arrivals as a collection of three random variables; and

(b) Arrivals as a Binomial process with large n and small p

3.2.1 Arrivals as a collection of three Random Variables

The three types of random variables that can be observed from the arrivals are:

(a) Sequence of interarrival times: $\{X\} = (X_1, X_2, \ldots)$ where X_1 is the time of the first arrival and X_i the time between the $(i-1)^{st}$ and i^{th} arrival for $i \in \{2, 3, \ldots\}$.

Assuming that customers do not generally influence one another to make calls to a call center, the interarrival times can be assumed to form an independent-and-identically-distributed (iid) sequence. This establishes that the sequence of iterarrival times is a renewal process thus allowing us to completely specify the probabilistic behaviour of the process upto a single, positive parameter, λ.

The renewal assumption also implies the 'memoryless property':

$$\mathrm{P}\{X > t + s \,|\, X > t\} = P\{X > s\} \quad s, t \in [0, \infty)$$

The exponential distribution is the only continous distribution possessing this property:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

Equation 3.1 is the density function for the sequence of interarrival times with rate parameter λ.

(b) Sequence of arrival times: $\{T\} = (T_0, T_1, \ldots)$ where T_n is the time of the n^{th} arrival, $n \in \mathbb{N}_+$ and $T_0 = 0$ is not an arrival. Thus, the n^{th} arrival time is the sum of the first $n-$interarrival times:

$$T_n = \sum_{i=1}^{n} X_i \,, \quad n \in \mathbb{N}$$

thus the sequence of arrival times, $\{T_n\}$, is the partial sum process associated with the sequence of interarrival times, $\{X_n\}$. The distribution function of T_n is the n^{th}-fold convolution of the exponential distribution and thus is a gamma (or Erlang) distribution with parameters n and λ ie., $f_{n+1} = f_n * f_1$

$$f_n(t) = \begin{cases} \lambda e^{-\lambda x} \frac{\lambda t^{n-1}}{(n-1)!} & , x \geq 0 \\ 0 & , x < 0 \end{cases}$$

(c) The counting process, $\{N\} = N(t) : t \geq 0$ where $N(t)$ is the number of arrivals in $(0, t]$, $t \in [0, \infty)$. The counting process distribution can be formed via an inverse relation between $\{N\}$ and $\{T\}$, ie., Since $N(t) \geq n \iff T_n \leq t$, for $t \in [0, \infty)$, $n \in \mathbb{N}_+$

$$\begin{aligned}
\therefore\ P\{N(t) \geq n\} &= P\{T_n \leq t\} \\
1 - e^{-\lambda t} \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} &= \int_0^t f_n(s) \mathrm{d}s \\
\text{but} \quad P\{N(t) = n\} &= P\{N(t) \geq n\} - P\{N(t) \geq n+1\} \\
&= e^{-\lambda t} \left[\sum_{k=0}^{n} \frac{(\lambda t)^k}{k!} - \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} \right] \\
&= e^{-\lambda t} \frac{(\lambda t)^n}{n!}
\end{aligned}$$

Ross (2009)[8] offers two other properties of the increments of a counting process:

(a) Independent increments: A counting process possesses *independent increments* if the numbers of events that occur in disjoint time intervals are independent.

(b) Stationary increments: A counting process possesses *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.

3.2.2 Arrivals as a Binomial process with large n and small p

The arrivals can also be thought of as a sequence of independent Bernoulli variables, where any customer,(x), either calls ($x = 1$) with a probability p or doesn't call ($x = 0$) with a corresponding probability of $1 - p$. The mass function associated with Bernoulli variables is given by:

$$f_X(x) = p^x (1-p)^{1-x} = \begin{cases} p & , x = 1 \\ (1-p) & , x = 0 \end{cases}$$

If we have more than one caller, say k callers, and a total customer base n, then this becomes a 'Bernoulli n-trials' experiment. An instance with k callers from a customer base of n potential callers has a probability of $p^k(1 - p)^{n-k}$ for each single arrangement. All such arrangements combined give the total probability of k callers from n potential callers given below:

$$f_X(x) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Now, the population, n, of potential callers is very large (ie., $n \to \infty$) and from previous experience, a very small percentage of that population actually makes calls to the call center, thus $p \to 0$. In this case the binomial distribution approximates to the Poisson distribution: (Let $\mu = np$:)

$$f_X(x) = e^{-np} \frac{(np)^k}{k!}$$

3.2.3 The Poisson Process

The counting process $N(t), t \geq 0$ is said to be a *Poisson process* having rate λ, $\lambda > 0$ if:

1. $N(0) = 0$.
2. The process has stationary and independent increments.
3. The number of events in an interval of length t is Poisson distributed with mean λt. ie., for all $s, t \geq 0$

$$P\{N(t+s) - N(s) = n\} = e^{-\lambda t} \frac{(\lambda t)^n}{n!}, \quad n = 0, 1, \ldots \tag{3.2}$$

The Poisson process thus has stationary increments and that $E[N(t)] = \lambda t$.

An arbitrary counting process is a Poisson process if it satisfies the above three conditions.(Ross 2009[8])

3.2.4 The Nonhomogenous Poisson Process

The counting process $\{N(t),\ t \geq 0\}$ is a nonhomogenous (or nonstationary) Poisson process with intensity function $\lambda(t)$, $t \geq 0$, if:

1. $N(0) = 0$.
2. $\{N(t), t \geq 0\}$ has independent increments.
3. $P\{N(t+h) - N(t) \geq 2\} = o(h)$.
4. $P\{N(t+h) - N(t) = 1\} = \lambda(t)h + o(h)$.

Definition 3.1 A function $g(\cdot)$ is said to be $o(h)$ if:

$$\lim_{h \to 0} \frac{g(h)}{h} = 0$$

Proposition 3.2.1. *Let $\{N(t), t \geq 0\}$ be a Poisson process with rate λ. An event that occurs at time t, independently of what has occurred before, has a probability of occurence of $p(t)$. Let $N_c(t)$ denote the number of counted events by time t, the counting process $\{N_c(t), t \geq 0\}$ is a nonhomogenous Poisson process with intensity function $\lambda(t) = \lambda \cdot p(t)$.*

The following statements are true of the above proposition

1. $N_c(0) = 0$
2. The number of counted events in $(s, s+t)$ depends only on the number of events of the Poisson process within that interval, which is independent of what

occurred before s. As a consequence, the number of counted events in the interval is independent of the process of counted events prior to s, thus the independent increment property.

3. Let $N_c(t, t+h) = N_c(t+h) - N_c(t)$, with $N(t, t+h)$ similarly defined

$$P\{N_c(t, t+h) \geq 2\} \leq P\{N(t, t+h) \geq 2\} = o(h)$$

4. We compute $P\{N_c(t, t+h) = 1\}$ by conditioning on $N(t, t+h)$.

$$\begin{aligned} P\{N_c(t, t+h) = 1\} &= P\{N_c(t, t+h) = 1 | N(t, t+h) = 1\} P\{N(t, t+h) = 1\} \\ &+ P\{N_c(t, t+h) = 1 | N(t, t+h) \geq 2\} P\{N(t, t+h) \geq 2\} \\ &= P\{N_c(t, t+h) = 1 | N(t, t+h) = 1\} \lambda h + o(h) \\ &= p(t) \lambda h + o(h) \end{aligned}$$

The above shows that time sampling a Poisson process results in a nonhomogenous Poisson process.

Proposition 3.2.2. *If $\{N_i(t), i = 1, \ldots, k\}$, represents the number of type i events occurring by time t then $\{N_i(t), i = 1, \ldots, k\}$, are independent Poisson random variables with expectation:*

$$E[N_i(t)] = \lambda \int_0^t P_i(s) ds \tag{3.3}$$

Suppose that $\{N(t),\ t \geq 0\}$ is a nonhomogenous Poisson process with bounded intensity function $\lambda(t) \leq \lambda$, for all $t \geq 0$. Let $\{M(t), t \geq 0\}$ be a nonhomogenous Poisson process with intensity function $\mu(t) = \lambda - \lambda(t)$, $t \geq 0$, independent from $\{N(t), t \geq 0\}$, then $\{N(t), t \geq 0\}$ can be regarded as the process of time-sampled events of the Poisson process $\{N(t) + M(t), t \geq 0\}$, where an event of the Poisson process that occurs at time t is counted with probability $p(t) = {}^{\lambda(t)}/_{\lambda}$. Consequently from proposition 3.2.2

$N(t)$ is a Poisson random variable with mean

$$E[N(t)] = \lambda \int_0^t \frac{\lambda(x)}{\lambda} dx = \int_0^t \lambda(x) dx$$

Thus, a Poisson process starting at time s, $N(t+s) - N(t)$ is the number of events over t time units and the mean would be $\int_0^t \lambda(x)dx$ giving us a definition for the function $m(t)$: (Ross (2009)[8])

$$m(t) = \int_0^t \lambda(x) dx \tag{3.4}$$

The above equation is called the **mean value function** or the **cumulative rate function** of the nonhomogenous Poisson process. This function is of particular interest throughout this project since it fully describes the nonhomogenous Poisson process (see Funaki and Matoba (1999)[3]).

3.3 Method of Maximum Likelihood for the parameters of the cyclic rate NHPP

In this project, a nonhomogenous Poisson process displaying cyclic rate is assumed to be an exponential-polynomial-trigonometric function (EPTF) of degree m and with the form:

$$\lambda(t) = exp\{h_\Theta(m,t)\} \tag{3.5}$$
$$h_\Theta(m,t) = \sum_{i=0}^{m} \alpha_i t^i + \gamma sin(\omega t + \phi)$$

where:

$\Theta = [\alpha_o, \alpha_1, \ldots, \alpha_m, \gamma, \phi, \omega]$: vector of unknown parameters

$\sum_{i=0}^{m} \alpha_i t^i$: general trend over time

$\gamma sin(\omega t + \phi)$: cyclic effect of the process

Numerical methods have to be used to obtain the maximum likelihood estimates of the parameters. Consider a sequence of n events occurring at the epochs $t_1, t_2, \ldots, t_n$ in a fixed time interval $(0, s]$ according to a NHPP with an intensity function of the form of Equation (3.5), then the log-likelihood function of Θ, given $N(s) = n$ and $t = (t_1, t_2, \ldots, t_n)$, is:

$$\mathcal{L}(\Theta|n,t) = \sum_{i=0}^{m} \alpha_i T_i + \gamma \sum_{j=1}^{n} sin(\omega t_j + \phi) - \int_0^s exp\{h_\Theta(m,z)\}dz \tag{3.6}$$

where

$$T_i = \sum_{j=1}^{n} t_j^i \quad \text{for } i = 0, 1, \ldots, m;$$

The value of m is an unknown parameter and the appropriate value is usually determined by a likelihood ratio test[10].

$$2\left[\mathcal{L}_{m+1}(\hat{\Theta}_{m+1}|n,t) - \mathcal{L}_m(\hat{\Theta}_m|n,t)\right]$$

For a given value of $m, m \geq 0$, we obtain $m + 4$ likelihood equations:

$$\begin{aligned}
\frac{\partial \mathcal{L}(\Theta|n,t)}{\partial \alpha_i} &= T_i - \int_0^s z^i exp\{h_\Theta(m,z)\}dz = 0, \quad i = 0, 1, \ldots, m, \\
\frac{\partial \mathcal{L}(\Theta|n,t)}{\partial \omega} &= \sum_{j=1}^{n} t_j cos(\omega t_j + \phi) - \int_0^s z \cdot cos(\omega z + \phi) exp\{h_\Theta(m,z)\}dz = 0, \\
\frac{\partial \mathcal{L}(\Theta|n,t)}{\partial \gamma} &= \sum_{j=1}^{n} t_j sin(\omega t_j + \phi) - \int_0^s sin(\omega z + \phi) exp\{h_\Theta(m,z)\}dz = 0, \\
\frac{\partial \mathcal{L}(\Theta|n,t)}{\partial \phi} &= \sum_{j=1}^{n} t_j cos(\omega t_j + \phi) - \int_0^s cos(\omega z + \phi) exp\{h_\Theta(m,z)\}dz = 0, \qquad (3.7)
\end{aligned}$$

The above system of nonlinear equations is solved numerically using special-purpose software, yielding the maximum likelihood estimates of the parameters (Lee et al. 1991[5]).

3.4 Simulation of Cyclic Rate NHPP

Once the parameters have been estimated, a trajectory of arrival times of the fitted exponential-polynomial-trigonometric rate function (EPTF) will be generated by the method of inversion: for $j \in (1, 2, \ldots, n)$

a) Generate $\{U\} = \{u_1, u_2, \ldots, u_n\}$ for $u_j \sim unif(a = 0, b = 1)$, (see Appendix B.1, page 47)

b) Generate $\{X\} = \{x_1, x_2, \ldots, x_n\}$ where $x_j = F_X^{-1}(u_j)$, (see Appendix B.2, page 47)

An NHPP with a rate function $\lambda(z)$, $z\in[0,S]$, the cummulative distribution function of the next event time τ_{i+1} conditioned on the observed value $\tau_i = t_i$ is

$$F_{\tau_i+1|\tau_i}(t_{i+1}|t_i)\equiv P\{\tau_{i+1}\leq t|\tau_i = t_i\} = \begin{cases} 1 - exp\left[-\int_{t_i}^{t}\lambda(z)dz\right] & \text{, if } t\geq t_i \\ 0 & , otherwise. \end{cases}$$

We then generate a random number u_{i+1} from the uniform distribution as discussed in the method of inversion above and compute τ_{i+1}:

$$\tau_{i+1} = t_{i+1} = F^{-1}_{\tau_{i+1}|\tau_i}(u_{i+1}|t_i).$$

which is equivalent to the solution of t_{i+1} in the equation:

$$\int_{t_i}^{t_{i+1}} \lambda(z)dz = -ln(1 - u_{i+1}).$$

and the value of t_{i+1} determined by a binary (bisection) search over the interval (t_i, S)([10]).

The results of the simulation run will then be used together with the observed values to ascertain the goodness of fit of the model.

Chapter 4

Data Analysis

In this chapter, I will proceed to empirically show that the telephone call center arrival counts data comes from a Poisson process - the theory was presented in the previous chapter. I will then proceed to estimate the parameters of the process using a special software package created for that purpose and check how well the resulting model fits the actual data by means of a nonparametric goodness-of-fit test and the cumulative distribution plot.

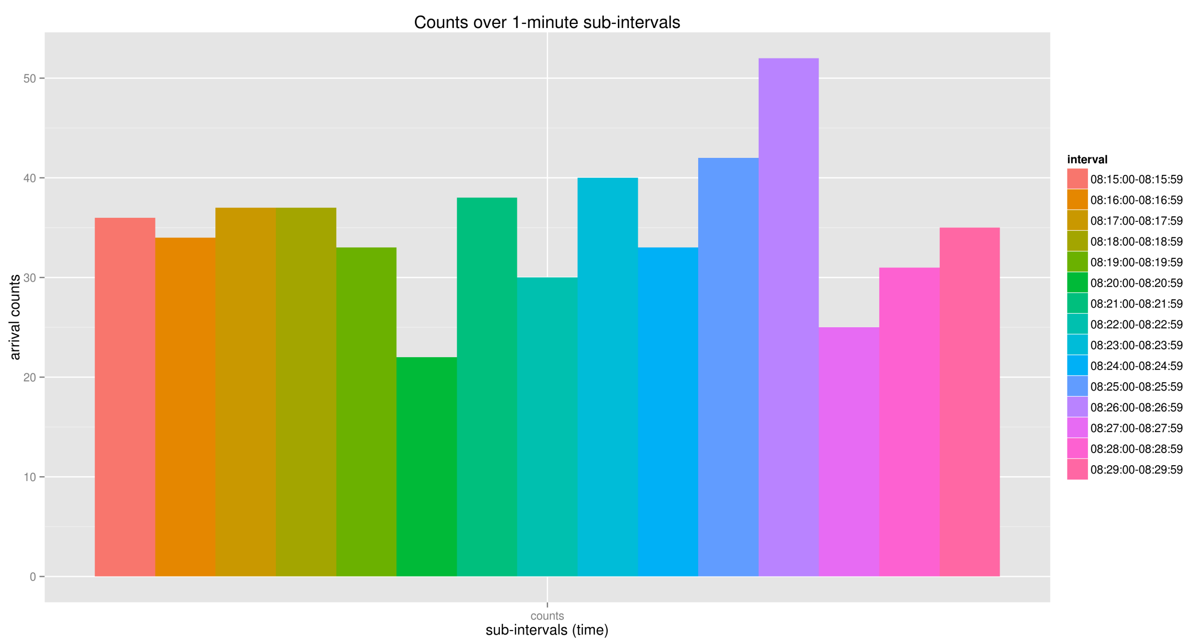
4.1 Arrival Counts Data as a Poisson Process

Brown et. al., (2004)[4] described the test of arrival counts in three parts: *a*) Time dependence of arrivals: the arrivals do not depend on exact time clock; *b*) Independencence of arrivals; and *c*) Interarrival times are exponential iid. The theory behind this test has been explained in the previous chapter and important derivations shown.

4.1.1 Test of Time Dependence

According to Brown et. al., (2004)[4], we should choose a short time interval over which $\lambda(t)$ can be presumed nearly constant on any given day. If $\lambda(t) = \lambda_{date}$ is a constant over this interval on each day, then the counts within this time interval over many days will be approximately uniformly distributed as a function of the clock time - the null hypthesis. Figure 4-1 below is a barplot of arrival counts for regular weekday calls in the month of November 1999.

Figure 4-1: Barplot of arrival counts for regular weekday calls in the month of November 1999 between 08:15am and 08:30am.



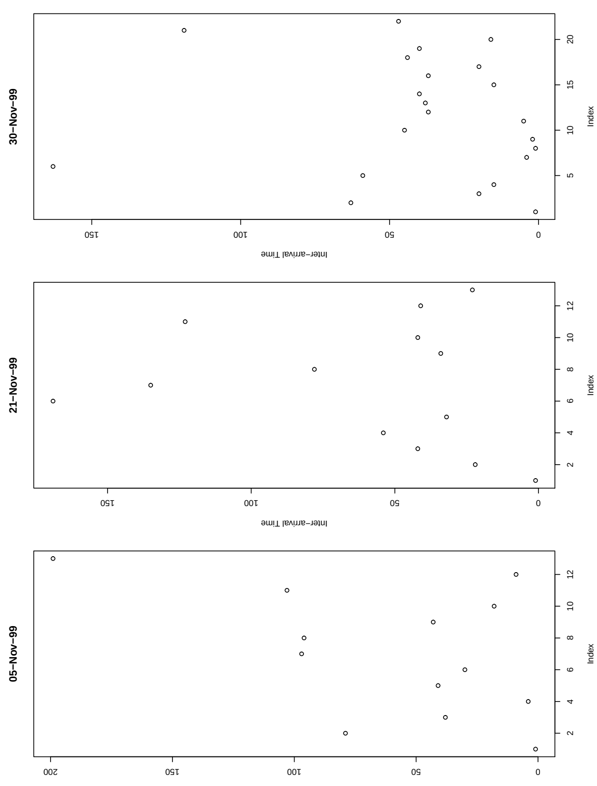
The test statistic has 14 df under the null hypothesis and a $\chi^2 = 20$, P-value $= 0.1301$. We have therefore established uniformity. Inspection of other 15-minute intervals with 1-minute sub-intervals gives similar results.

4.1.2 Test of the Independence of Arrivals

Brown et. al., (2004)[4] established the independence by looking at correlations of transformations of inter-arrival times. As such, we shall determine the correlations of the raw inter-arrival times, and the corresponding power-transforms:

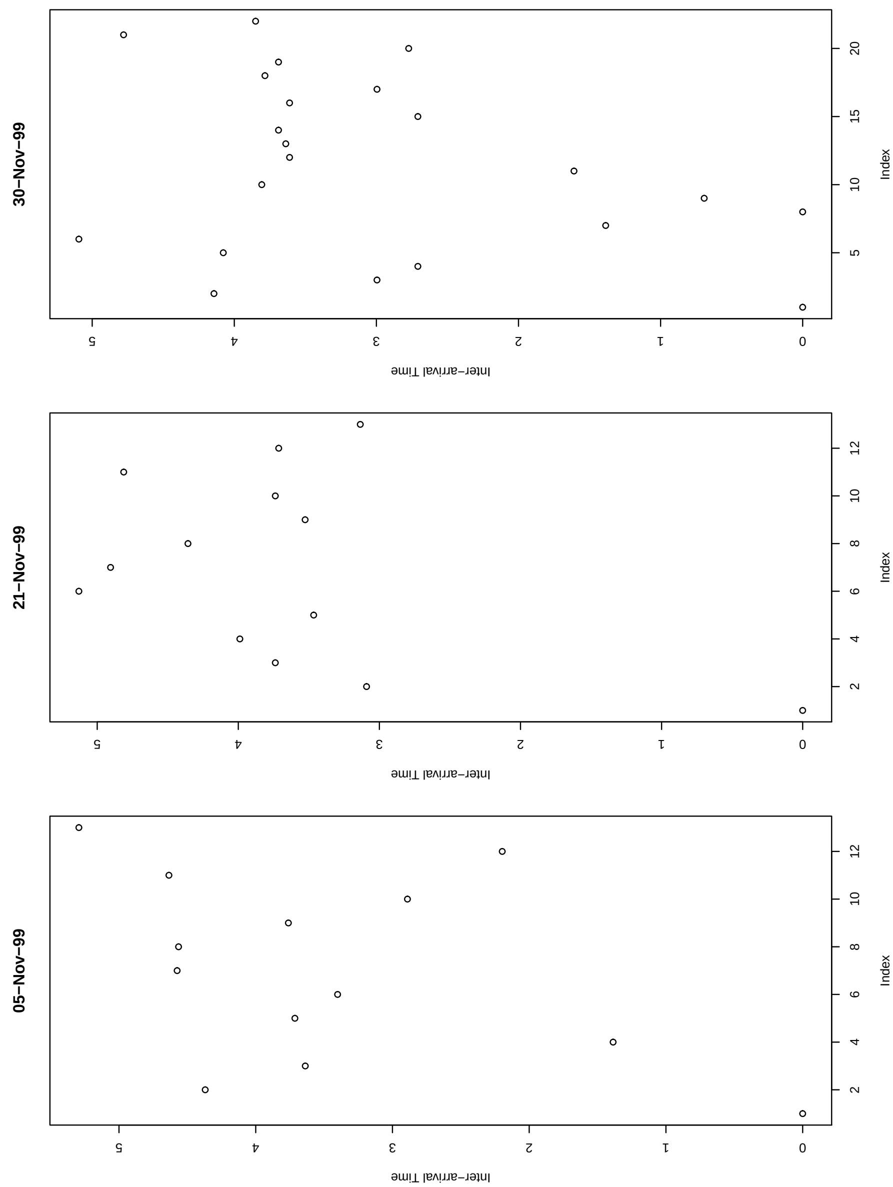
$$f(x) = \begin{cases} x^{\lambda} & ,\lambda \neq 0 \\ ln(x) & ,\lambda = 0 \end{cases}$$

Figure 4-2: Scatterplot of raw inter-arrival times for November 1999 between 08:15am and 08:30am.



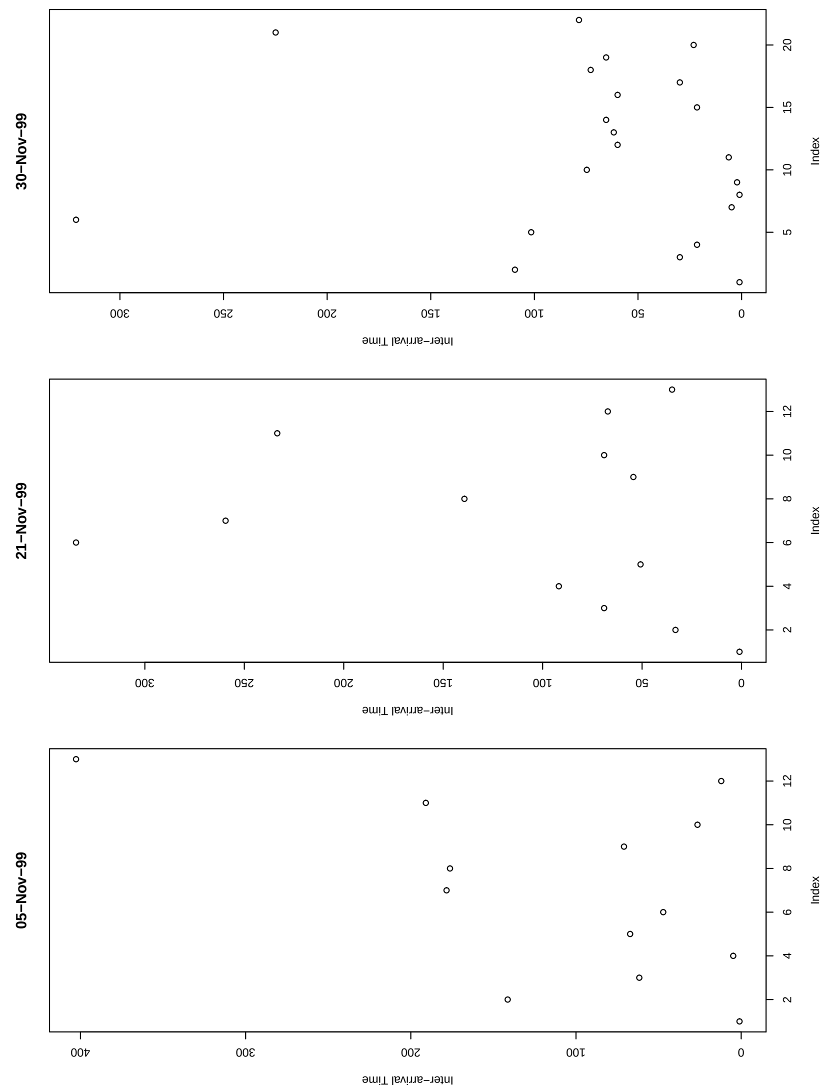
The coefficients of determination, R^2, between inter-arrival times and their ordering within that interval are *a*) 0.2163969 *b*) 0.03291567 *c*) 0.01201348 whereby *a*) 21.64% *b*) 3.29%; and *c*) 1.2% of the variation is explained for the days shown in figure 4-2 respectively. These are lower than the threshold needed to conclude that inter-arrival times are dependent on ordering.

Figure 4-3: Scatterplot of log inter-arrival times for November 1999 between 08:15am and 08:30am.



The coefficients of determination, R^2, between inter-arrival times and their ordering within that interval are *a*) 0.1957118 *b*) 0.171809 *c*) 0.1173514 whereby *a*) 19.57% *b*) 17.18%; and *c*) 11.74% of the variation is explained for the days shown in figure 4-3 respectively. These are lower than the threshold needed to conclude that the transformed inter arrival times are dependent on ordering.

Figure 4-4: Scatterplot of power inter-arrival times for November 1999 between 08:15am and 08:30am.



The coefficients of determination, R^2, between inter-arrival times and their ordering within that interval are *a*) 0.2238258 *b*) 0.02650988 *c*) 0.00713507 whereby *a*) 22.38% *b*) 2.65%; and *c*) .714% of the variation is explained for the days shown in figure 4-4 respectively. These are lower than the threshold needed to conclude that the transformed inter arrival times are dependent on ordering.

All the shown coefficients of determination are very close to 0 thus indicating a strong lack of linear correlation of the inter-arrival times with ordering hence establishing the independece of inter-arrival times.

4.1.3 Test that Interarrival times are exponential i.i.d

A homogenous Poisson process with a rate, λ, has interarrival times from an exponential distribution with scale $\frac{1}{\lambda}$. The data from our interval will be normalised so as to be from an exponential distrubtion with $\frac{1}{\lambda} = 1$. Let

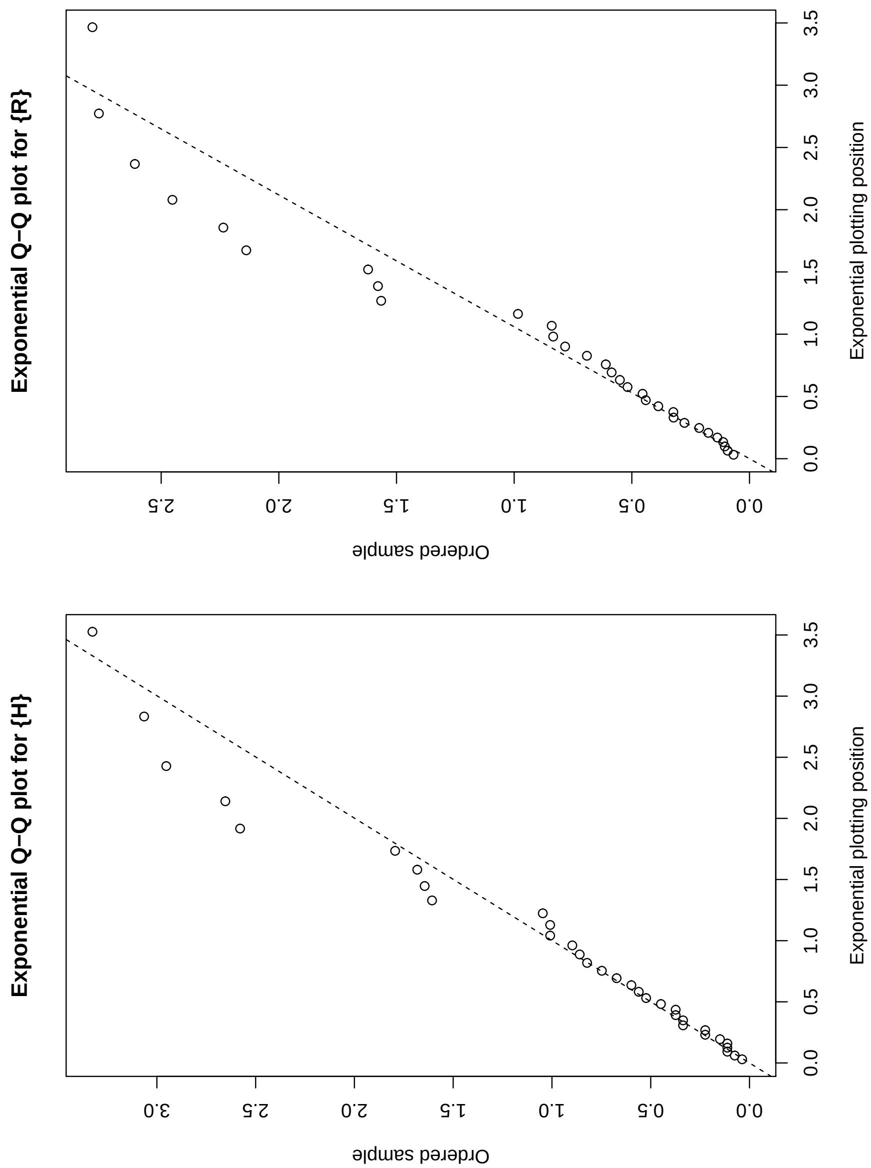
$$\begin{aligned} \lambda_{day} &: \text{denotes constant Poisson rate over a specified interval for the day} \\ T_{day,j} &: \text{denote time of the} j^{th} \text{arrival on the indicated day} \\ G_{day,j} &: T_{day,j} - T_{day,j-1} \\ J_{day} &: \text{denotes the number of calls during the interval for that day} \end{aligned}$$

we define the below inter-arrival time transformations for $j = 1, 2, \ldots, J_{day}$:

$$\begin{aligned} H_{day,j} &= \frac{J_{day} G_{day,j}}{b - a} \\ R_{ij} &= (J(i) + 1 - j) \left(-log \left(\frac{L - T_{i,j}}{L - T_{i,j-1}} \right) \right) \end{aligned}$$

We test the $\{H\}$ and $\{R\}$ under the null hypothesis that they are approximately exponentialy distributed with rate = 1. (see Brown et. al.,(2004)[4] and Brown et. al.,(2005)[1]).

Figure 4-5: Exponential Q-Q plots for 01 November 1999 between 08:15am and 08:30am.



The plots in figure 4-5 indicate that $\{H\}$ and $\{R\}$ are exponentially distributed with λ=1 because of the relatively good alignment along and about the 45° line. We conclude that the interarrivals are thus independent and identically distributed variables from an exponential distribution with $\lambda = 1$.

4.2 Maximum Likelihood Estimation of Parameters

Special purpose software (*MP3MLE*[1]), was used to solve the $m + 4$ system of nonlinear equations 3.7 in Chapter 3 and the output is shown below:(see Appendix C.1 page 51 for input to the software and Appendix C.2 page 51 for the full output)

```
   FINAL PARAMETER ESTIMATES
Polynomial coefficients
alpha( 0) = -0.99045
alpha( 1) = 0.19064E-02
alpha( 2) = -0.89508E-06
alpha( 3) = 0.12335E-09
Trigonometric parameters
amplitude (gamma( 1)) = 1.0716
phase (phi( 1)), in radians = -1.7428
phase (phi( 1)), in time units = -0.27738
frequency (omega( 1)), in radians = 0.41985E-02
frequency (omega( 1)), in time units = 0.66821E-03
```

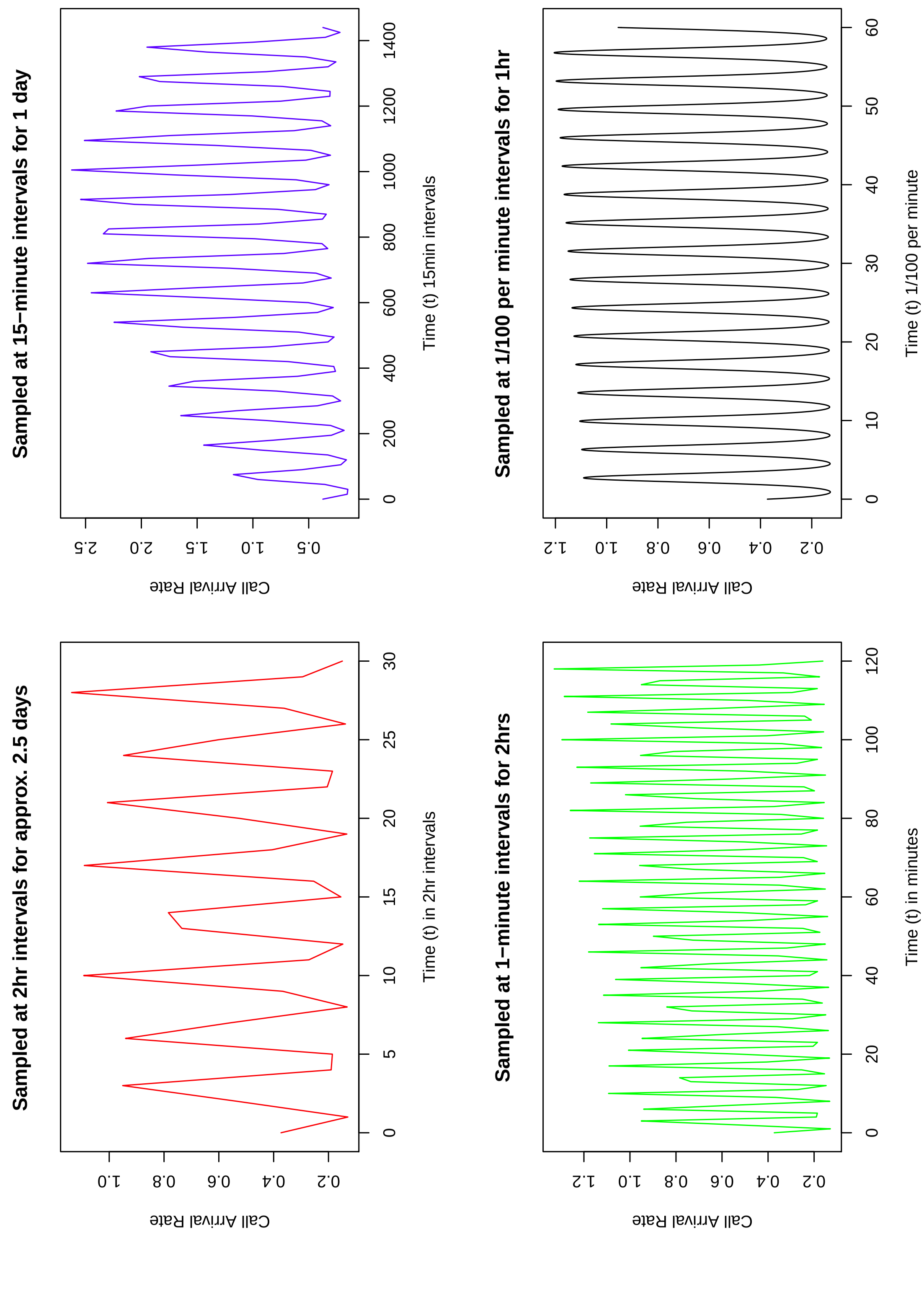
[1]Edward P. Fitts Department of Industrial & Systems Engineering, North Carolina State University, Raleigh, NC 27695-7906. (2005, Jul). *MP3MLE software for fitting to sample arrival-time data a nonhomogeneous Poisson process whose rate function may exhibit multiply periodic effects or a long-term trend (or both).*. Retrieved April 04, 2014, from the Edward P. Fitts Department of Industrial & Systems Engineering, North Carolina State University website: *http://www.ise.ncsu.edu/*

The estimated rate function is thus:

$$\begin{aligned}\hat{\lambda}(t) &= exp(h_{\hat{\Theta}}(4,t)) \\ &= exp(-0.99045 + 0.0019064t - 0.00000089508t^2 - 0.00000000012335t^3 \\ &\quad + 1.0716 \cdot sin(-1.7428t + 0.0041985))\end{aligned}$$

Figure 4-6 below shows graphs of the intensity function over various sub-intervals for the first two and a half days of November 1999.

Figure 4-6: Graphs of the Estimated Intensity function

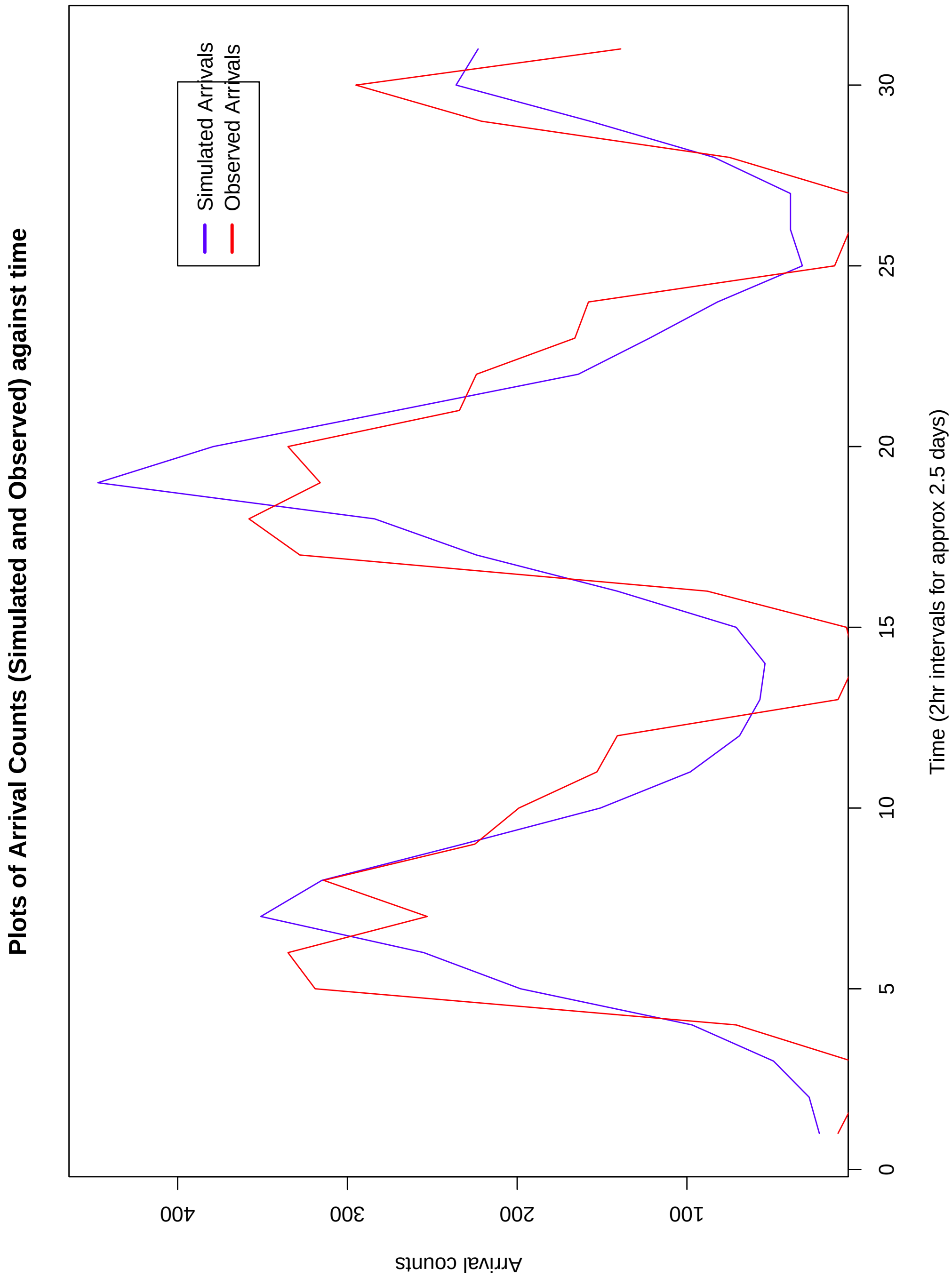


The estimated intensity function was used to run simulations of arrival times (as described in Chapter 3) enabling us to determine goodness-of-fit of the model to the actual realisations via graphical means (see Figures 4-7, and 4-9). A summarised trajectory of the arrival times for both the Simulated and Observed arrivals is shown below:

Table 4.1: A sample of the arrival time trajectories of the Simulated and Observed process.

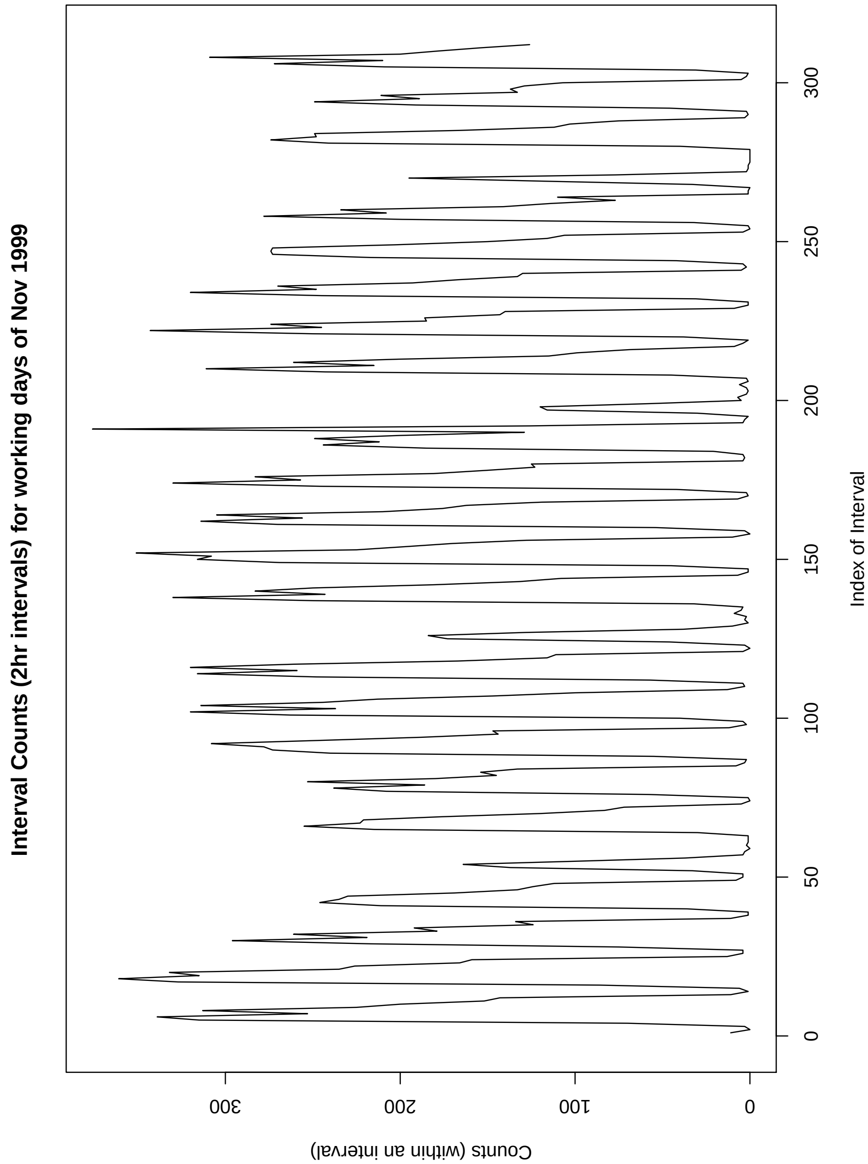
Simulated	Observed
16.5650	1.0167
20.8910	1.2500
27.6364	2.5500
31.3094	10.2167
⋮	⋮
3699.9458	3670.8667
3699.9693	3670.9500
3699.9713	3671.9500
3699.9823	3672.0333

Figure 4-7: Graph of Simulated and Observed Arrival counts over 2hr intervals for the first approx 2.5 days



Below is a graph showing the Call arrivals trend for the whole month of November 1999.

Figure 4-8: Graph of Arrival counts over 2hr intervals for the workdays of Nov 1999.



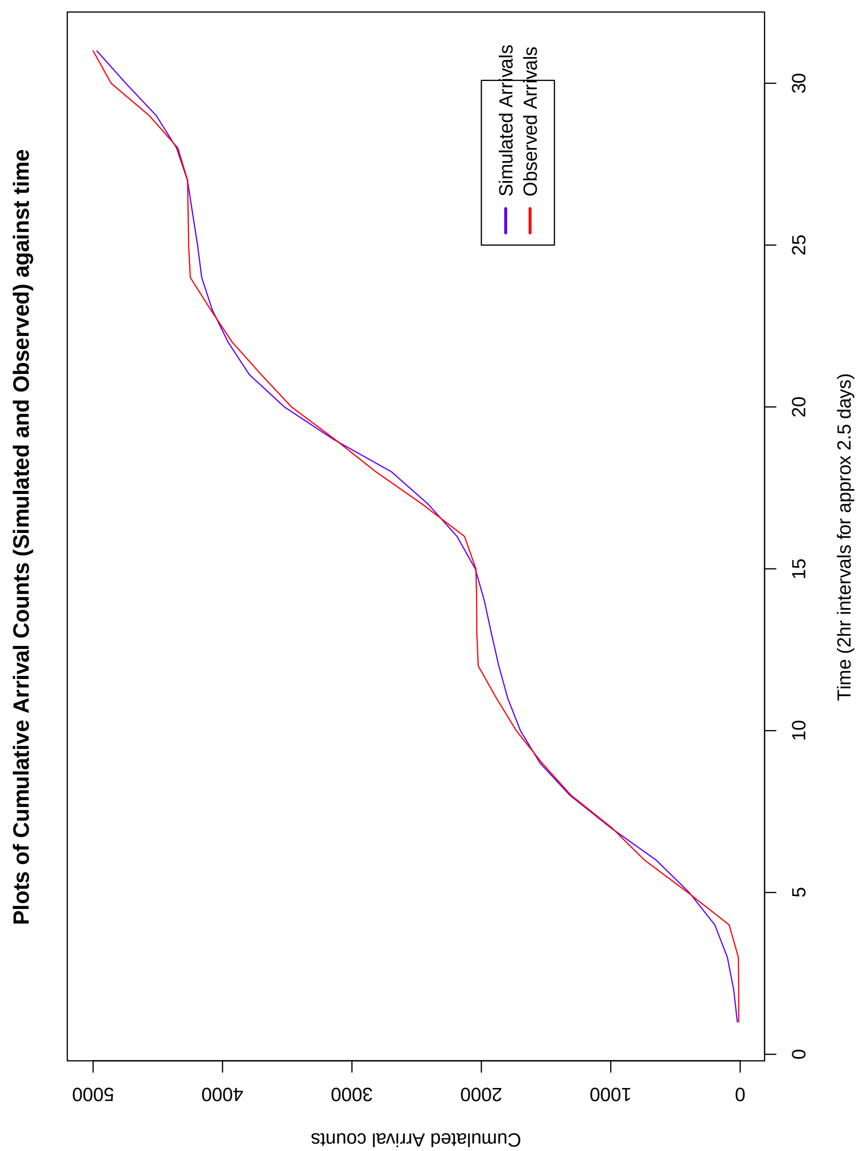
4.3 Goodness-of-Fit Test

The Kolmogorov-Smirnov nonparametric goodness-of-fit test was conducted for the simulated and observed arrival time. We fail to reject the null hypothesis that the data comes from the same distribution at $\alpha = 0.01$. The output is shown below:

```
    Two-sample Kolmogorov-Smirnov test
data:  montharrs$arrivaltime[1:5000] and simonarrs
D = 0.0316, p-value = 0.01362
alternative hypothesis:  two-sided
```

The above is further confirmed by a graph of the cumulated arrival counts (Figure 4-9 below)

Figure 4-9: Graph of Cumulated Arrival counts over 2hr intervals for the first two working days of Nov 1999.



Chapter 5

Conclusions & Further Work

In Chapter 2, Literature Review, Gans et. al.,(2003)[7] opined that one of the most pressing needs was to see improvements in in the forecasting of arrival rates. This project has been an attempt at making this possible. I have shown how call arrivals to a Call Center can be modelled as a nonhomogenous Poisson process. Assumptions have been stated and theory adequately developed towards the empirical realisation of a Stochastic model that fits the observed data with a confidence of 90%.

The contents of previous chapters is summarised below:

- In Chapter 1, I opened with a discussion of Telephone call centers and a few challenges encountered. I then isolated the challenge of modelling call center arrivals; compiled a number of objectives based on this and drew up an outline of what would be contained in the rest of this document.

- In Chapter 2, I did a comprehensive literature review of past work that would guide the project. I also drew up a framework of methods and their interactions that would serve to give a snapshot view of how everything comes together.

- In Chapter 3, I built a theoretical argument for the Telephone call center arrival data as a Poisson process. I also generalised the Poisson process as a nonhomogenous Poisson process by proving that the nonhomogenous Poisson process is a time-sampled Poisson process. I also presented a summary of using the method of Maximum Likelihood from the Lee et. al.,(1991)[5] paper for parameter estimation and a simulation algorithm, the method of inversion, presented to illustrate how the parameter estimates will be used in simulating the arrival process.

- In Chapter 4, I tested the Telephone call center data to confirm whether it fits the assumptions of a Poisson process. I then run the data through special purpose software that computed parameter estimates by solving a set of nonlinear equations. Finally, I conducted a simulation using the parameter estimates and did a nonparametric goodness-of-fit test to gauge how well the model fits the observed process.

5.1 Some Challenges Encountered

The *MP3MLE* special-purpose software could only handle a maximum of 5000 observed arrival times hence limiting the accuracy of estimates to the time span of the 5000 epochs. The software is also old (1994) and written in Fortran - a language not specifically designed for statistical work. I also had to manually iterate estimates of frequency for input to the software.

An 'R' package that can model a nonhomogenous Poisson process with cyclic rate does not seem to exist (before Apr 2014) but it appears this will be in the works soon.

I also found a package, SAPP[1], that could model an EPF[2] and ETF[3] rate but did not include both polynomial and trigonometric components in the same function. I also had to write software that uses the parameter estimates to simulate the nonhomogenous Poisson process with cyclic rate.

I also encountered some difficulty in procuring queueing data from local (Kenyan) Banks - there was reluctance to release this. However, I was able to get anonymised Bank data from an Israeli University site, The Technion University, which has been used for an impressive number of publications.

5.2 Further Work

In addition to the Gans et. al.,(2003)[7] 'list of most pressing needs' given in Chapter 2, I think that:

1. User education on the importance of data-driven decision making should be emphasized especially in the Management Sciences: This would probably see a greater appreciation of Statistical Methods in many aspects of Management. Gans et. al.,(2003)[7] had noted that data in some call centers is not stored because of ignorance on the part of the Management.

2. More modern, and freely available Software packages to model and simulate Stochastic process ought to be developed to aid research and development efforts.

[1]**S***tatistical* **A***nalysis of* **P***oint* **P***rocesses*, The Institute of Statistical Mathematics, Tokyo *http://jasp.ism.ac.jp/ism/sapp/index_e.html*, Available at *http://cran.r-project.org/web/packages/SAPP/*.

[2]EPF: Exponential-Polynomial-Function

[3]ETF: Exponential-Trigonometric-Function

5.3 Concluding Remarks

Soli Deo Gloria!

Appendix A

Proofs

Proof that $\{X_n\}$ are (i.i.d)

Proof. X_2 is the first arrival time after $T_1 = X_1$ thus it is independent of X_1 but with the same distribution. Similarly, X_3 is the first arrival time after $T_2 = X_1 + X_2$, so X_3 is independent of X_1 and X_2. By induction, X_n is independent of any X_{n-1}. □

'Memoryless property' of i.i.d exponential variables

Proof.

$$\begin{aligned}
P(X > s+t | X > t) &= \frac{P(\{X > s+t\} \cap \{X > t\})}{P(X > t)} \\
&= \frac{P(X > s+t)}{P(X > t)} \\
P(X > s+t) &= P(X > s)P(X > t), \qquad \text{by independence} \\
\therefore \quad \frac{P(X > s+t)}{P(X > t)} &= P(X > s) \qquad \text{(A.1)}
\end{aligned}$$

□

Proof that gamma p.d.f is an n^{th}-fold convolution of exponential random variables

Proof.

$$\begin{aligned} f_2(t) &= \int_0^t f_1(s) f_1(t-s)\mathrm{d}s \\ &= \int_0^t \lambda e^{-\lambda(t-s)} \lambda e^{-\lambda s}\mathrm{d}s \\ &= \lambda e^{-\lambda t} \lambda t \end{aligned}$$

by induction on n

$$\begin{aligned} f_{n+1}(t) &= \int_0^t f_n(s) f_1(t-s)\mathrm{d}s \\ &= \int_0^t \lambda e^{-\lambda s} \frac{(\lambda s)^{n-1}}{(n-1)!} \times \lambda e^{-\lambda(t-s)}\mathrm{d}s \\ &= \lambda e^{-\lambda t} \frac{\lambda^n}{(n-1)!} \int_0^t s^{n-1}\mathrm{d}s \\ &= \lambda e^{-\lambda t} \frac{\lambda^n}{(n-1)!} \left[\frac{t^n}{n}\right] \\ &= \lambda e^{-\lambda t} \frac{(\lambda t)^n}{n!}\,. \end{aligned}$$

□

Appendix B

Code Snippets

B.1 Generate {U} for $u_i \sim Unif(0,1)$ rv

```
univalues <- runif(5000,min=0,max=1)
write(univalues, file="unifvalues.txt",sep="\n")
```

B.2 Generate {X} for $x_j = F_X^{-1}(u_j)$

```
def rhs( unifval ):
    return -1 * math.log(1 - unifval)

def rateFunction(x, params):
    coefficients = params[0]
    amplitude    = params[1][0]
    phase        = params[1][1]
    frequency    = params[1][2]
```

```
    value = float()
    power = 0
    for c in coefficients:
        value  += (c * pow(x, power))
        power  += 1
    value += amplitude * math.sin(frequency * x + phase)

    return math.exp(value)

def simulate(unifile, paramfile, outfile, s):
    sanitise = lambda x: float(x.strip())
    params = getParams(paramfile)
    tmevals = [0]
    tmelen = len(tmevals)

    ufile  = open(unifile, "r")
    unifvalues = ufile.readlines()
    unifvalues = map(sanitise, unifvalues)
    # Pre-compute Right-Hand side values
    rhss = map(rhs, unifvalues)

    # Loop through remaining uniform values
    for lnrhs in rhss:
        # Enter bisect loop to converge on value of t_i+1
        upper = s
        lower = tmevals[-1]
        while True:
            mid = (lower + upper) / 2.0
            area = IntegrateMidrule(tmevals[-1], mid, rateFunction, params)
            # Solution found
```

```
        if abs(area - lnrhs) <= 0.0001:
            tmevals.append(mid)
            break
        # No solution
        if '%.4f'%lower == '%.4f'%upper: break

        if (area > lnrhs):
            upper = mid # Upper bound moves down
        else:
            lower = mid # Lower bound moves up

writeOutfile(outfile, tmevals[1:], "w")
```

THIS PAGE INTENTIONALLY LEFT BLANK

Appendix C

Program Input and Output

C.1 Input to MP3MLE

```
'Call Arrivals at Call Center'
3700.0 5000
3 10 1 0
0.004037388
```

C.2 Output from MP3MLE

```
 *** Estimation Program mp3mle ***
Version 1.0 - July 1994


   Call Arrivals at Call Center
Number of observations = 5000
Length of observation interval = 3700.0
```

```
No.  of trigonometric components to be included in rate function = 1
Maximum allowable order of polynomial in rate function = 3
Initial estimate of frequency( 1) (in radians/unit of time) 0.40374E-02
Ratio-test level of significance = 10
Ratio-test level of significance assigned = 10


    Initial parameter estimates
gamma0(1)= 1.14495
phi0(1)= 1.53764
omega0(1)= 0.403739E-02


   Fit zero order polynomial
alpha(0)= 0.531549E-01
gamma(1)= -1.11191
phi(1)= 1.19262
omega(1)= 0.427422E-02


   Fit polynomial of order 1
alpha(0)= -0.159200E-01
alpha(1)= 0.372615E-04
gamma(1)= -1.11460
phi(1)= 1.23519
omega(1)= 0.424868E-02


   Fit polynomial of order 2
alpha(0)= -0.390376
alpha(1)= 0.585129E-03
```

```
alpha(2)= -0.144837E-06
gamma(1)= -1.09587
phi(1)= 1.22689
omega(1)= 0.429133E-02


   Fit polynomial of order 3
alpha(0)= -1.02245
alpha(1)= 0.197282E-02
alpha(2)= -0.932599E-06
alpha(3)= 0.129626E-09
gamma(1)= -1.07126
phi(1)= 1.41716
omega(1)= 0.418860E-02


   Fit polynomial of order 4
alpha(0)= -2.95893
alpha(1)= 0.774415E-02
alpha(2)= -0.641267E-05
alpha(3)= 0.215499E-08
alpha(4)= -0.252517E-12
gamma(1)= -1.10336
phi(1)= 1.68077
omega(1)= 0.410742E-02


    Final estimates do not pass the likelihood ratio test.


   FINAL PARAMETER ESTIMATES
```

```
Polynomial coefficients
alpha( 0) = -1.0225
alpha( 1) = 0.19728E-02
alpha( 2) = -0.93260E-06
alpha( 3) = 0.12963E-09


    Trigonometric parameters
amplitude (gamma( 1)) = 1.0713
phase (phi( 1)), in radians = -1.7244
phase (phi( 1)), in time units = -0.27445
frequency (omega( 1)), in radians = 0.41886E-02
frequency (omega( 1)), in time units = 0.66664E-03


   Value of log-likelihood function -2150.9
```

Bibliography

[1] Lawrence Brown, Noah Guns, Avishai Mandelbaum, Anat Sakov, Haipeng Shen, Sergey Zeltyn, and Linda Zhao. Statistical analysis of a telephone call center: A queueing-science perspective. *Journal of the American Statistical Association*, 100(469), 2005.

[2] I. B. Frenkel, I. B. Gertsbakh, and L. V. Khvatskin. Parameter estimation and hypothesis testing for nonhomogenous poisson process. *Transport and Telecommunication*, 4(2), 2003.

[3] Kenichi Funaki and Hideaki Matoba. Piecewise pseudo estimation for nhpp using data in the adjacent intervals. *Operations Research Society of Japan*, 42(4), December 1999.

[4] Brown Lawrence, Zhao Linda, Haipeng Shen, and Avishai Mandelbaum. Multi-factor poisson and gamma-poisson models for call center arrival times. Technical report, University of Pennsylvania, Technion University, University of North Carolina at Chapel Hill, 2004.

[5] Sanghoon Lee, James R. Wilson, and Melba M. Crawford. Modelling and Simulation of a Nonhomogenous Poisson Process having Cyclic Behaviour. *Communications in Statistics - Simulation and Computation*, 20:777–809, 1991.

[6] William A Massey, Geraldine A Parker, and Ward Whitt. Estimating the parameters of a nonhomogeneous poisson process with linear rate. *Telecommunication Systems*, 5(2):361–388, 1996.

[7] Gans N., G. Koole, and A. Mandelbaum. Telephone call centers: Tutorial, review and research prospects. *Manufacturing & Service Operations Management*, 5(2):79–141, April 2003.

[8] Sheldon M. Ross. *Introduction to Probability Models*. Academic Press, Elsevier, The Boulevard, Langford Lane, Kidlington, Oxford, OX5 1GB, UK, tenth edition, 2009.

[9] Anat Sakov, Larry Brown, Sergey Zeltyn, Linda Zhao, Haipeng Shen, and Avishai Mandelbaum. Empirical analysis of a telephone call center. Technical report, Tel-Aviv University, 2001.

[10] James R Wilson. Users manual for *mp3mle* and *mp3sim*. Technical report, North Carolina State University.