Analysis of Call Center Data

By – Yen Chu Cheng

Advisor – Lawrence D. Brown

Wharton Undergraduate Research Scholars

WH-299-301

April 2004

ABSTRACT

A call center is a place where a group of agents service customers remotely via the telephone. Queueing theory is used extensively in the study of call centers. One of the most important statistical models used to predict call arrivals in queueing theory is the Poisson process. This paper summarizes the results of the application of this model in the study of call arrivals to a particular bank. Specifically, the call arrivals to the VRU (Voice Response Unit) and to the service queue are examined. These calls are put through different techniques that test whether the actual data conform to the hypothetical model for a Poisson process. An explanation of each of these tests and the results are described in this paper.

1 Introduction

This research focuses on the issue of efficiency within a call center. A call center is a place where calls are answered by service agents and it can handle a considerable volume of calls at the same time. This service is typically operated with the help of automated voice response systems which screen and forward the calls to the service agents. The central problem in such a service is the concept of queueing. Operations manager often have to resolve the question of how to minimize the queueing time of customers, but at the same time minimize the operating costs of the service center. This involves determining the optimal number of service agents that offers the best trade-off between queueing times and operating costs.

To perform such an analysis, there is a need to understand the queueing system as a whole. This will require information relating to the arrival process, the service mechanism and the queue characteristics. This research will focus on the arrival process. This includes how call arrival rates and distributions vary across different times of the day. The actual data that was studied in this research was obtained from the call center of a large bank. This bank handles up to 100 calls per minute during the peak period. The Statistics Department of the Wharton School has been obtaining these data since April 2001.

Section 2 of the paper will provide a brief discussion of some of the literature relevant to this study of the call center, in particular papers within the field of queueing theory and statistical techniques employed in the tests for a Poisson process. Section 3 will give an in depth description of the process flow of an incoming call to a call center and describe the specific issue that is addressed in the research. Section 4 will discuss some of the statistical approaches and research findings obtained from our study of the call center data. Section 5 is the conclusion and it will give a summary as well as provide directions for future research in the field of queueing theory. Section 6 records the list of references used. Section 7 is an appendix that contains the computer program written to analyze the data from the call center and also contains some additional graphs generated as part of the research findings.

2. Literature Survey

Our research is contained within the general area of queueing theory. Queueing theory is the theoretical discipline used to mathematically model a variety of real world processes that involve waiting times (queues) and consequent service times. Our investigation has been conducted in relation to data collected at one moderately large financial call center. We study a particular component within this theory; namely customer arrival times.

The conventional queueing theory models assume that these arrival times follow what is called a Poisson process whose rate function is approximately constant over short time intervals (of at least a few minutes in length). Otherwise the rate function may vary over longer time horizons. For this reason, the assumed arrival process is mathematically called a time-inhomogeneous Poisson process. Such processes have a stochastic regularity that is particularly amenable to theoretical calculations related to properties of the consequent queueing system. We will make use of a few simple properties of such a Poisson process, and these will be mentioned later. The reader should consult standard textbooks on probability theory or stochastic processes for further discussion on these properties, for example *A First Course in Probability* by Ross, S. (2002).

The literature that gives the best overview for empirical research into queueing systems is the paper by Brown, Gans, et. al. (2002). That paper summarizes the analysis of the complete operational history of a banking call center. The data used in that research came from a bank call center that is considerably smaller than the one we use in our present analysis. Using the basic ideas behind the queueing theory, analysis was performed in that paper on three different components of the system: arrivals, customer abandonment behavior and service duration. For each component, different statistical techniques were developed and used to carry out the analysis. Gans, Koole and Mandelbaum (2002) contains an extensive bibliography of call center literature.

Of particular interest and relevance to our study is the statistical technique that was employed in that paper to test for a Poisson process. This test is based on the fact that the intervals between arrivals in a Poisson process follow an exponential distribution. In that paper, the use of this test indicated that the arrival times of calls requesting service by an agent are well modeled as an inhomogeneous Poisson process with a smoothly varying rate function, λ. This test is replicated in my current study of the call center, but surprisingly the results are different from those found previously.

There is another paper that focuses on a new and different statistical technique for testing whether a sample of observations comes from a single Poisson distribution. That paper is by Brown and Zhao (2001) and proposes an additional test for the Poisson distribution based on Anscombe's (1948) variance stabilizing transformation. This test was found to be advantageous due to its breadth of application and ease of use and it is also employed in the analysis of call center data in this paper.

3 Framing of Specific Problem

This research focuses on just one particular aspect of the queueing theory – the arrival of calls to a call center. The data source for this research comes from the call center of a large U.S. bank. Some features of this general example of a queueing system are described below.

3.1 Process Flow of an Incoming Call

Before we begin to study the statistics related to the call arrivals to the bank, we need to understand the process flow of an incoming call to the call center. This process flow can be summarized in Figure 1.

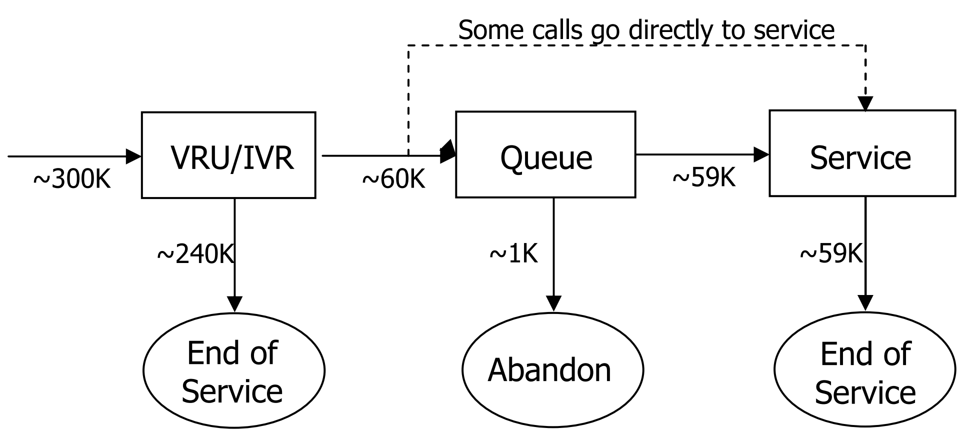


Figure 1: Process flow of an incoming call to the call center.

On a typical weekday, the call center of the bank receives approximately 300,000 calls. When a customer makes a call to the bank, the call is first routed to a Voice Response Unit (VRU) (sometimes also known as Interactive Voice Response (IVR)). During this phase of the call, the response unit will prompt the customer to identify himself and find out the type of services the customer requires. It might be possible for the customer to perform some self-service transactions at this point, such as receiving bank account information or finding out general information about the bank. In this bank, approximately 80% (~240,000) of the customers complete their service within the VRU. The remaining approximately 20% (~60,000) of the customers request the assistance of a service agent.

These customers are said to enter the service queue. The time they "enter the service queue" is recorded (in seconds) as the first time at which they could have been transferred to an agent, if there had been a free agent available to serve them. In actuality some of the customers entering the service queue ***do*** get served immediately. Others are forced to wait in a queue until there is either an agent available to serve them or until the customer decides to abandon the system by hanging up the phone.

This is a very simplified description of the process flow of an incoming call. In reality, the process can be much more complicated if we take into consideration actual flows of calls within the VRU and the difference in services rendered to high-priority and low-priority customers. However, for the purpose of this research, these complications need not be fully examined.

3.2 Call Center Optimization

The focus of operational research on queueing systems such as call centers is on optimizing efficiency within the system. This involves minimizing the time a customer spends in the service queue as well as minimizing the number of service agents. Logically, the most relevant aspect of the process flow model that has to be understood is the rate at which customers exit the VRU and enter the service queue (or go directly into service) because this will have a direct impact on the optimal number of service agents. The focus of this research will thus be on the 20% of the calls that require the service of an agent.

The approach that has to be taken involves modeling the arrival times of the calls that enter the service queue. The common call center practice is to assume that the times of calls entering the VRU form a time inhomogeneous Poisson process. There is then reason to further believe that the times of calls exiting the VRU and entering the service queue also follow a time inhomogeneous Poisson process. To determine whether this is true, a test of a null hypothesis that arrivals of calls to the service queue form an inhomogeneous Poisson process has to be constructed. This will be the focus of Section 4 of this paper.

In the test for an inhomogeneous Poisson process, a key consideration is that the arrival rate may fluctuate across different times of the day. To resolve this, the interval of the day has to be broken down into relatively short blocks of time and the arrival rate is then assumed to be constant within each block of time. A test can then be performed on each block of time to determine whether the call arrivals form an inhomogeneous Poisson process.

3.2 Data Description

In order to perform the tests on the call arrivals, there is a need to have a program that can manipulate the data that comes in from the call center. The data that was received from the call center was for all the call arrivals in the entire year of 2002. As described previously, the bank receives approximately 300,000 calls each day and the data set that has to be managed is really huge.

In order to manage the data on a scalable basis, programs have to be written in C++ on the Unix platform to allow the user to specify certain parameters (for example the date of the call, the time which the call enters the service queue and the service type required) and retrieve the data of calls that fall within these parameters. This will allow the user to conveniently narrow down the data set and perform statistical analysis on the required data.

3.3 Statistical Tests

Once programs have been written to conveniently narrow down the data set, statistical tests have to be employed to determine whether call arrivals to the service queue follow an inhomogeneous Poisson process. In the course of our research, three different types of tests were employed. Each of these tests will be briefly described in Section 4 where the results from these tests will also be discussed.

4 Research Findings

The three tests on the call arrivals to the service queue are based on very different principles. The first test works on the basic premise that the inter-arrival rates of a Poisson process are independent and follow the exponential distribution. This test is taken from the paper by Brown, Gans, et. al. (2002). The second technique is a direct test of the property that the number of arrivals $N(t)$ in a finite interval of length t in a Poisson process obeys the Poisson distribution. Finally the third technique is taken from Brown and Zhao (2001). This technique is similar to the second technique except that it is an indirect test on whether call arrivals in a finite interval is Poisson. The use of these tests and the results are discussed and compared in the sections that follow.

4.1 First Test for the Poisson Process

The first test on whether the call arrivals form an inhomogeneous Poisson process is carried out as follows. We first construct a null hypothesis that the call arrivals to the service queue form an inhomogeneous Poisson process with arrival rates varying so slowly that they are effectively constant within very short time intervals. The characteristics of the system might be different at different times of day, or on different days, when the intensity of activity in the system is different. In other words, call arrivals at 12 noon on all the Mondays (assuming that all Mondays are a normal working day) might be well-behaved (i.e., Poisson) but arrivals at 07:00:00am or those at 12:00:00pm on any Tuesdays might not be well behaved.

In the construction of the test we considered the call arrivals within a targeted time interval (usually a few minutes long) on a certain day of the week for the month of June. We then broke up the targeted interval of a day into relatively short blocks of time. For convenience, we used blocks of equal time-length, L = 360 seconds, resulting in a total of I blocks. Let T_{ij} denote the j-th ordered arrival time in the i-th block, I = 1,…, I. thus $T_{i1} \leq \ldots \leq T_{iJ(i)}$, where J(i) denotes the total number of arrivals in the i-th block. Then define $T_{i0} = 0$ and

$$R_{ij} = (J(i)+1-j)(-\log(\frac{L-T_{ij}}{L-T_{i,j-1}})),\ j = 1,\ldots,J(i)\ .$$

Under the hypothesis that the arrival rate is constant within each given time interval, $\{R_{ij}\}$ will be independent standard exponential variables. The justification for this is discussed further in the paper by Brown, Gans, et. al. (2002).

Once the $\{R_{ij}\}$ for the set of calls have been computed, the null hypothesis was tested by applying the Kolmogorov-Smirnov goodness-of-fit test for exponential distribution on $\{R_{ij}\}$ and also by plotting the exponential quantile plot for $\{R_{ij}\}$. One problem that arose in the calculations of $\{R_{ij}\}$ was that on many occasions, due to the high call volume, many of the calls within a single block fell on the same second. This resulted in T_{ij} being equal to $T_{i,j-1}$ on many occasions, causing R_{ij} to be 0. Such a situation is not consistent with the theory in the cited paper that supports the use of this test. But there is a simple and mathematically correct solution to this dilemma. It is overcome by spreading the time of the calls uniformly over -0.5000 and 0.5000 seconds. In this manner R_{ij} is no longer zero.

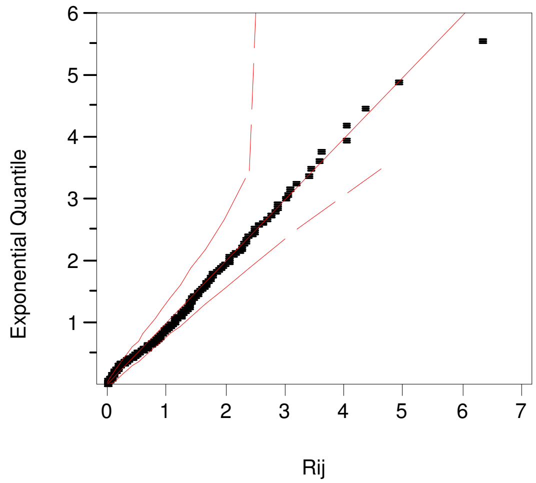


Figure 2: Exponential quantile plot for $\{R_{ij}\}$ on call arrivals at the service queue between 07:00:00am and 07:02:59am on all Wednesdays in June 2002.

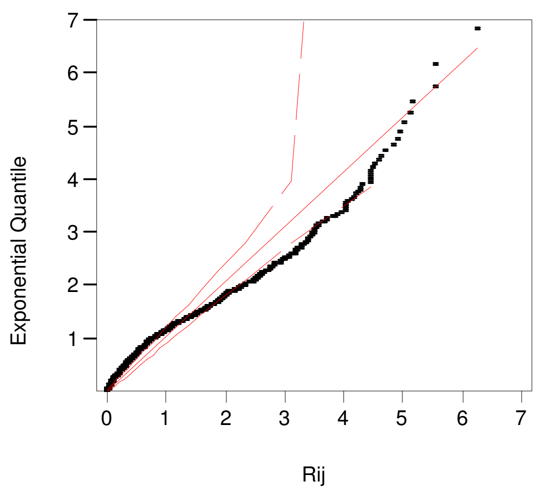


Figure 3: Exponential quantile plot for $\{R_{ij}\}$ on call arrivals at the service queue between 12:00:00am and 12:02:59am on all Wednesdays in June 2002.

Figures 2 and 3 display an application of this test. Figure 2 shows an exponential quantile plot for $\{R_{ij}\}$ computed on the call arrivals at the service queue between 07:00:00am and 07:02:59am on all the Wednesdays (total of four days) within the month of June 2002. Figure 3 shows a similar plot for $\{R_{ij}\}$ computed from call arrivals between 12:00:00pm and 12:02:59pm on all the Wednesdays (total of four days) within the month of June 2002.

As demonstrated in Figure 2, for the blocks from 07:00:00am to 07:02:59am, the fidelity of the exponential quantile plot to a straight line provides evidence that the null hypothesis is not rejected. Thus we can conclude that the assumption of an inhomogeneous Poisson process between 07:00:00am and 07:02:59am holds. In this case, the Kolomogorov-Smirnov statistic has a value K = 1.2807 (P-value = 0.0752 with n = 262) and we do not reject the null hypothesis. However, as seen in Figure 3, for the blocks from 12:00:00pm to 12:02:59pm on Wednesdays, it seems that the call arrivals substantially deviate from an inhomogeneous Poisson process. In this case the Kolomogorov-Smirnov statistic has value K = 3.5835 (P-value < 0.0001 with n = 942) and the null hypothesis is rejected.

Using the same form of analysis, the test was carried out for all the three minute periods starting at 07:00:00am, 10:00:00am, 12:00:00pm, 03:00:00pm, 06:00:00pm and 10:00:00pm for all the weekdays in June. The Kolomogorov-Smirnov statistics are summarized in Table 1 below. Please refer to Appendix I for the actual exponential quantile plots of the $\{R_{ij}\}$ computed for each of these times on each of the weekdays in June.

A summary for the overall results is as follows. The Kolomogorov-Smirnov statistics in Table 1 indicate that for the call arrivals at 07:00:00am, 06:00:00pm and 10:00:00pm (when the call volumes are lower - less than 50 calls per minute), there is no significant evidence from this test that the arrivals to service queue deviate from an inhomogeneous Poisson process. However, for the call arrivals starting at 10:00:00am, 12:00:00pm and 03:00:00pm (when the call volumes are much higher, up to 100 calls per minute), the call arrivals clearly deviated away from an inhomogeneous Poisson process.

	Monday			Tuesday			Wednesday			Thursday			Friday		
	K-value	P-value	n	K-value	P-value	n	K-value	P-value	n	K-value	P-value	n	K-value	P-value	n
07:00:00am - 07:02:59am	1.5699	0.0145	289	1.6476	0.0088	297	1.2807	0.0752	262	2.2263	0.0001	286	2.5779	0.0000	338
10:00:00am - 10:02:59am	4.8634	0.0000	1,255	4.4185	0.0000	1,080	3.5790	0.0000	1,037	3.8371	0.0000	976	5.0680	0.0000	1,038
12:00:00pm - 12:02:59pm	5.0767	0.0000	1,125	4.9690	0.0000	1,003	3.5835	0.0000	942	5.3432	0.0000	981	5.1352	0.0000	953
03:00:00pm - 03:02:59pm	4.4558	0.0000	1,023	3.5637	0.0000	919	2.6221	0.0000	882	4.1463	0.0000	957	3.9612	0.0000	873
06:00:00pm - 06:02:59pm	2.1127	0.0003	489	1.9137	0.0013	429	1.8694	0.0018	404	2.5003	0.0000	436	2.3732	0.0000	407
10:00:00pm - 10:02:59pm	0.8465	0.4707	224	0.9063	0.3841	181	0.7066	0.7003	155	1.5654	0.0149	167	0.9930	0.2776	151

Table 1: Kolomogorov-Smirnov statistics for the exponential distribution of $\{R_{ij}\}$ computed on the call arrivals to the service queue using different times of the days and on different days of the week in June.

Thus, the conclusion that can be drawn from this test is that during the non-peak periods, call arrivals to the service queue may well follow an inhomogeneous Poisson process but during the peak periods, the call arrivals do not. The abnormality of the results presents several questions that can be explored. Primarily, do the call arrivals into the VRU follow a Poisson process in the first place? If so, it must be that the inherent technical nature of the VRU affecting the length of time that a call spends in the VRU and changing the statistics of the calls that exit the VRU and enter the service queue. This primary question will later be explored in the subsequent two tests that were employed in the tests for an inhomogeneous Poisson process.

4.2 Second Test for the Poisson Process

The above test indicates that there seems to be different behaviors for call arrivals to the service queue during the peak and non-peak period. This second test will thus focus specifically on two time intervals, one during the peak period and one during the non-peak period, and examine whether the results from the first test is conclusive. This test is based on the idea that the number of arrivals N(t) in a finite interval of length t obeys the Poisson distribution (with parameter λt where λ is the mean arrival rate).

$$P\{N(t)=n\}=\frac{(\lambda t)^n}{n!}e^{-\lambda t}$$

Specifically, the non-peak period selected for testing is between 07:00:00am and 07:59:59am, and the peak period is between 12:00:00pm and 12:59:59pm.

The procedure for the test is as follows. First, we consider the time interval of 07:00:00am and 07:59:59am for all the weekdays in June 2002. These intervals are broken down into relatively short blocks of time of 6 minutes each and it is assumed that the arrival rates remains unchanged within each block. Since there are a total of 10 blocks in each 07:00:00am and 07:59:59am interval and a total of 20 weekdays in June 2002, there are altogether 200 six-minute blocks between 07:00:00am and 07:59:59am of all the weekdays in June 2002. For each six-minute block, the number of arrivals per second should approximate to a Poisson distribution if call arrivals follow a Poisson process.

To determine whether the number of arrivals per second in each block approximates to a Poisson distribution, the traditional chi-square goodness-of-fit statistic for each block is calculated in the following manner. Suppose the actual arrivals per second in an observed six-minute block are as summarized in Table 2:

Arrivals (m_i)	Actual Freq ($f_{i,observed}$)	$m_i f_i$
0	238	0
1	76	76
2	34	68
3	8	24
4	4	16
Total	360	184

Table 2: Summary of the frequency of call arrivals in an observed six-minute block.

The estimated average number of arrivals per second is calculated to be:

$$\overline{X} = \frac{\sum_i m_i f_i}{\sum_i f_i} = \frac{184}{360} = 0.5111$$

Subsequently, the data is broken down into 5 different bins as shown in the Table 3. The average number of arrivals per second is then used for the purposes of finding the probabilities of the Poisson distribution of a single call arriving in that bin. The theoretical frequency for each bin is obtained by multiplying the appropriate Poisson probability by the sample size of 360 seconds. This gives the results as shown in Table 3.

Arrivals (m_i)	Actual Freq ($f_{i,observed}$)	Probability (p_i) for Poisson Distribution with λ=0.5111	Theoretical frequency $f_{i,theoretical}$
0	238	0.59983	215.93834
1	76	0.30658	110.36849
2	34	0.07835	28.20528
3	8	0.01335	4.80534
>3	4	0.00190	0.68255

Table 3: Calculations of the theoretical call arrival frequencies in an observed six-minute block.

The chi-square statistics is computed as follows:

$$\chi^2_{k-p-1} = \sum_i \frac{(f_{i,observed} - f_{i,theoretical})}{f_{i,theoretical}}$$

where

k = number of bins with different call arrival frequencies
p = number of parameters estimated from the data
k-p-1 = degrees of freedom of the chi-square statistics

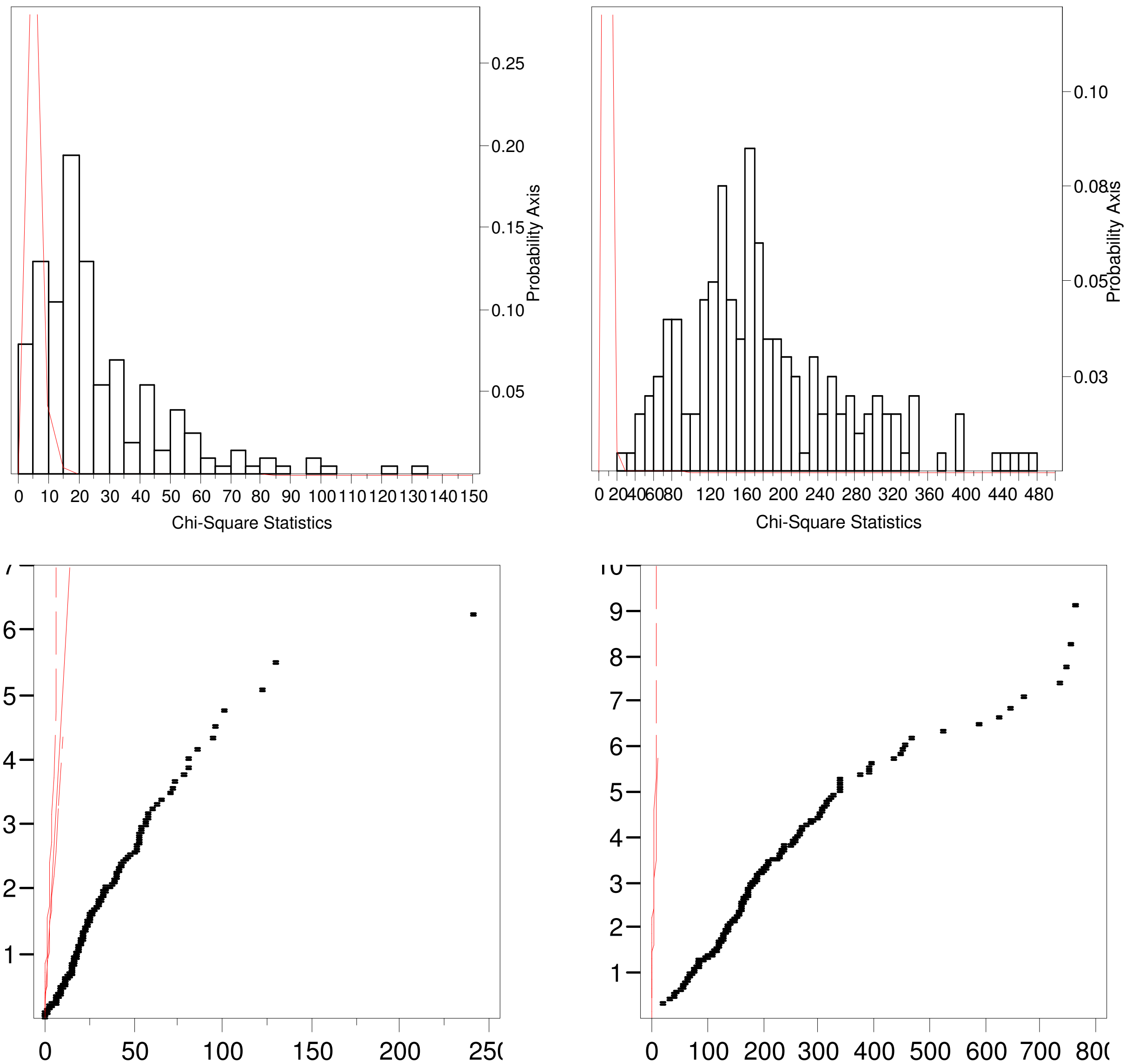
Here the chi-square statistics is found to be 35.011 with 3 degrees of freedom (k = 5 and p = 1 since the mean of the Poisson distribution is estimated from the data).

The chi-square statistics are calculated in a similar fashion for all the 200 six-minute blocks that falls within 07:00:00am and 07:59:59am for all weekdays in June 2002. Under the null hypothesis that call arrivals follow an inhomogeneous Poisson process (which implies that call arrival frequency per second within each relatively short block has a Poisson distribution), the distribution of the values in this set, $\{\chi^2\}$, should approximate to a chi-square distribution with 3 degrees of freedom.

All the calculations for the chi-square statistics are automated using a visual basic program written for Microsoft Excel. This program (Appendix II) takes in a list of all calls which enter the service queue between 07:00:00am and 07:59:59am and calculates the chi-square statistics for each of the six-minute blocks. These chi-square statistics are then fed into JMP and the histogram displaying the distribution is fitted to a chi-square distribution with 3 degrees of freedom. The quantile plot for the distribution of the statistics is also plotted.

The output from JMP is as shown in Figure 4(a). In the diagram, the red line is the probability density of the chi-square distribution and it is the shape that the histogram of the calculated chi-square statistics should take if the calls are Poisson. Clearly from this histogram and the quantile plot, $\{\chi^2\}$ computed does not conform to a chi-square distribution and we reject the null hypothesis that the call arrivals between 07:00:00am and 07:59:59am follow an inhomogeneous Poisson process.

A similar test for an inhomogeneous Poisson process is conducted for all calls between 12:00:00pm and 12:59:59pm on all weekdays in June 2002. The result is shown in Figure 4(b). Once again, the distribution of the chi-square statistics clearly does not fit a chi-square distribution and this test rejects the null hypothesis that the call arrivals to the service queue during the peak period follow an inhomogeneous Poisson process. Note that the deviation from chi-squared here is even considerably more obvious than that in Fig 4(a). This – at least – is consistent with the results from the first test that found the evidence against Poissonicity at 12pm to be very much stronger than that against Poissonicity at 7am.



(a) 07:00:00am and 07:59:59am

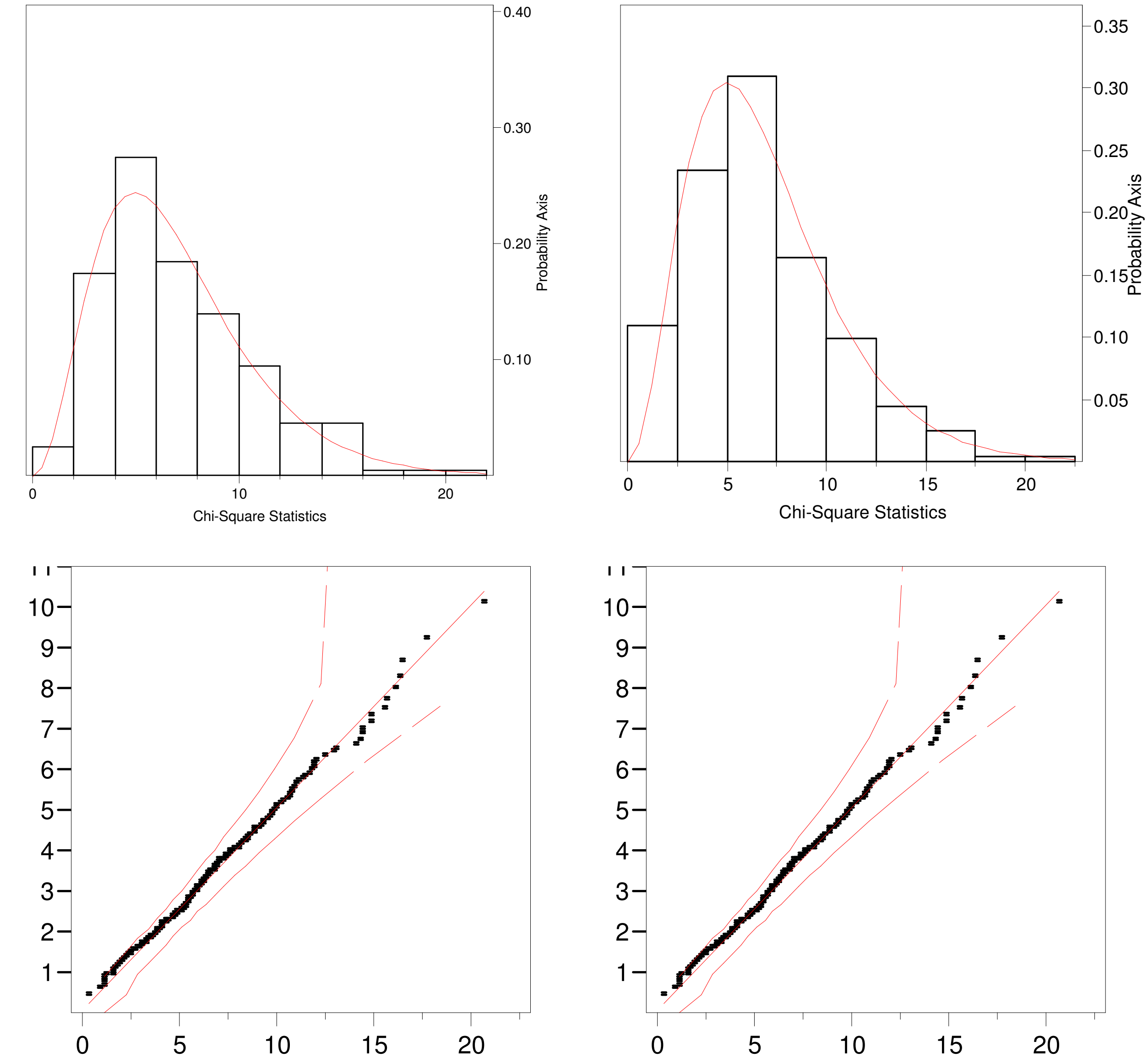
(b) 12:00:00pm and 12:59:50pm

Figure 4: Histogram of the $\{\chi^2\}$ of the call arrivals to the service queue in six-minutes blocks between 07:00:00am and 07:59:59am and between 12:00:00pm and 12:59:50pm fitted to a chi-square distribution.

4.3 Calls to the VRU

As noted earlier, it is important to also consider the arrivals to the VRU, especially since those to the service queue are not Poisson. For comparison, the test is repeated for call arrivals to the VRU. Figure 5(a) and 5(b) show the results of the test. From the graphs, it is evident that the chi-square statistics comes very close to a chi-square distribution, indicating the test does not reject the null hypothesis that the call arrivals to the VRU follow an inhomogeneous Poisson process.

From this test, we can conclude that the call arrivals to the VRU do indeed appear to follow an inhomogeneous Poisson process, but the calls that enter the service queue do not. This leads us to infer that there may be some mechanisms within the VRU that affect the statistics of the calls that exit the VRU.



(a) 07:00:00am and 07:59:59am

(b) 12:00:00pm and 12:59:50pm

Figure 5: Histogram of the chi-square statistics of the call arrivals to the VRU in six-minutes blocks between 07:00:00am and 07:59:59am and between 12:00:00pm and 12:59:50pm fitted to a chi-square distribution.

4.4 Third Test for the Poisson Process

This test is constructed in the hope of gaining a better understanding of the process that goes on within the VRU. We believe that the VRU may be causing small delays of a few seconds in some of the calls, causing the statistics of the calls that enter the service queue to deviate from an inhomogeneous Poisson process, but that over longer time horizons the process may appear indistinguishable from an inhomogeneous Poisson process. To eliminate some of the effects of these hypothesized, very short delays, the test considers the frequency of call arrivals in bins of 5 seconds, 10 seconds and 20 seconds. These larger bin sizes (compared to the previous test where the bin size is only 1 second) should eliminate the effects of small delays that the VRU may introduce to the calls.

The details of this test are described in the paper by Brown and Zhao (2001). This test is being performed on the calls that enter the service queue during the same time interval as before, namely between 07:00:00am and 07:59:59am and between 12:00:00pm and 12:59:59pm on all the weekdays in June 2002. Consider first the interval between 07:00:00am and 07:59:59am. We break the time interval into two-minute blocks, giving a total of 600 blocks for the 20 weekdays in June 2002. Each block is then analyzed in bins of 5 seconds (Number of bins per block, m_i = 24). Let N_{ij} denotes the number of arrivals to the service queue in the j^{th} 5-second bin in block i. We then define

$$Y_{ij} = \sqrt{N_{ij} + \frac{3}{8}}$$

and

$$T_i = 4\sum_{i}(Y_{ij} - \overline{Y})^2$$

Under the null hypothesis that the calls in the interval between 07:00:00am and 07:59:59am follow an inhomogeneous Poisson process, $\{T_i\}$ will follow a chi-square distribution with m_i-1 = 23 degrees of freedom. Figure 6(a) shows the result when $\{T_i\}$ is fitted to a chi-square distribution using JMP. Figure 6(b) shows the result when $\{T_i\}$ is computed for the interval between 12:00:00pm and 12:59:59pm under the same parameters. Note that JMP automatically scales the histogram and inserts the red line which is the probability density of the desired chi-square distribution.

This test is repeated for two-minute blocks with bins of 10 seconds duration (giving the number of bins per block, m_i = 12) as shown in Figure 6(c) and 6(d). It is also carried out for four-minute blocks with bins of 20 seconds duration (giving the number of bins per block, m_i = 12) as shown in Figure 6(e) and 6(f).

For comparison, the same set of tests is performed on the call arrivals to the VRU in the same time frame. The results are shown in Figure 7.

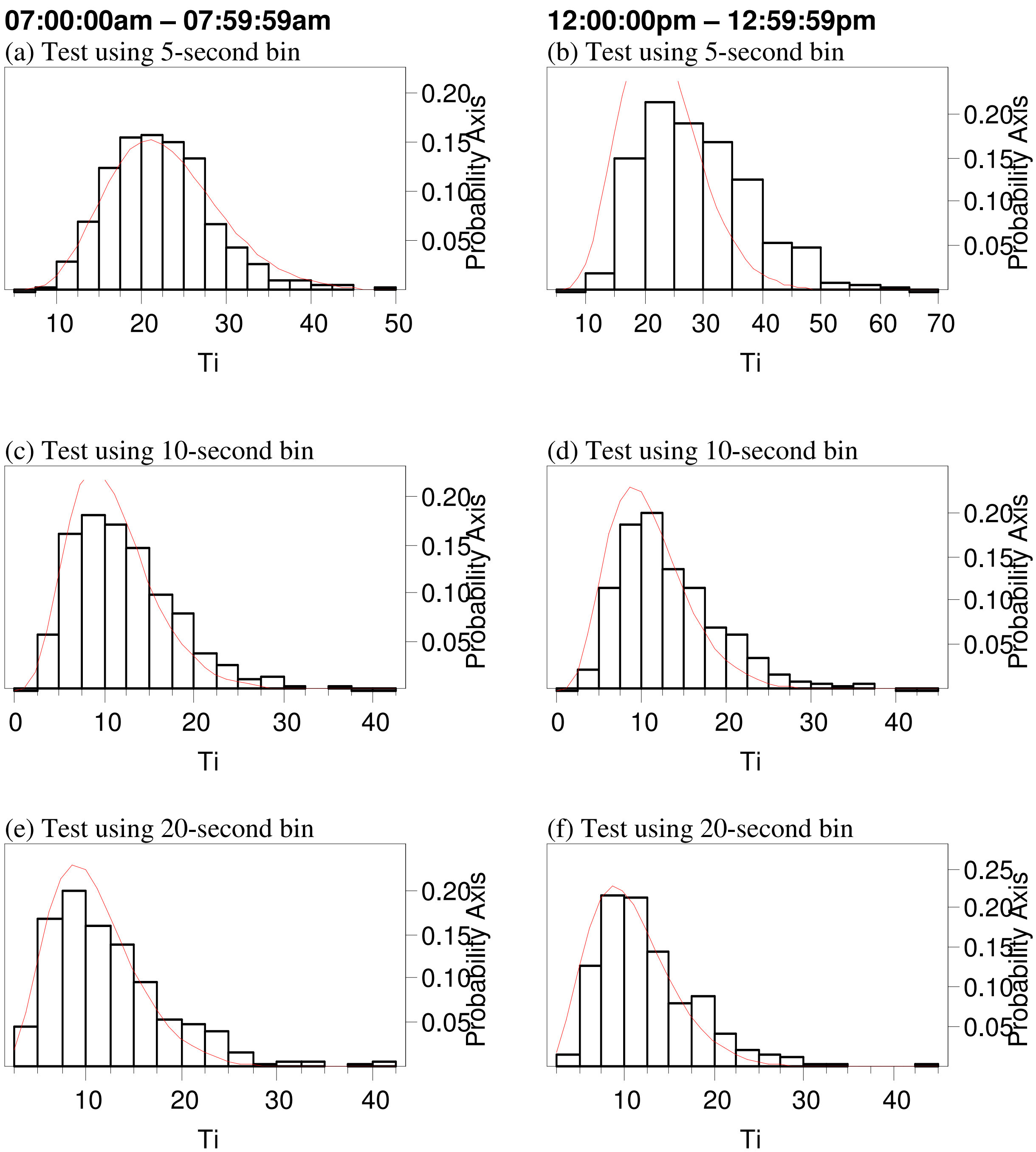


Figure 6: $\{T_i\}$ computed for the call arrivals to the service queue between 07:00:00am and 07:59:59am and between 12:00:00pm and 12:59:59pm on all the weekdays in June 2002 using various bin sizes.

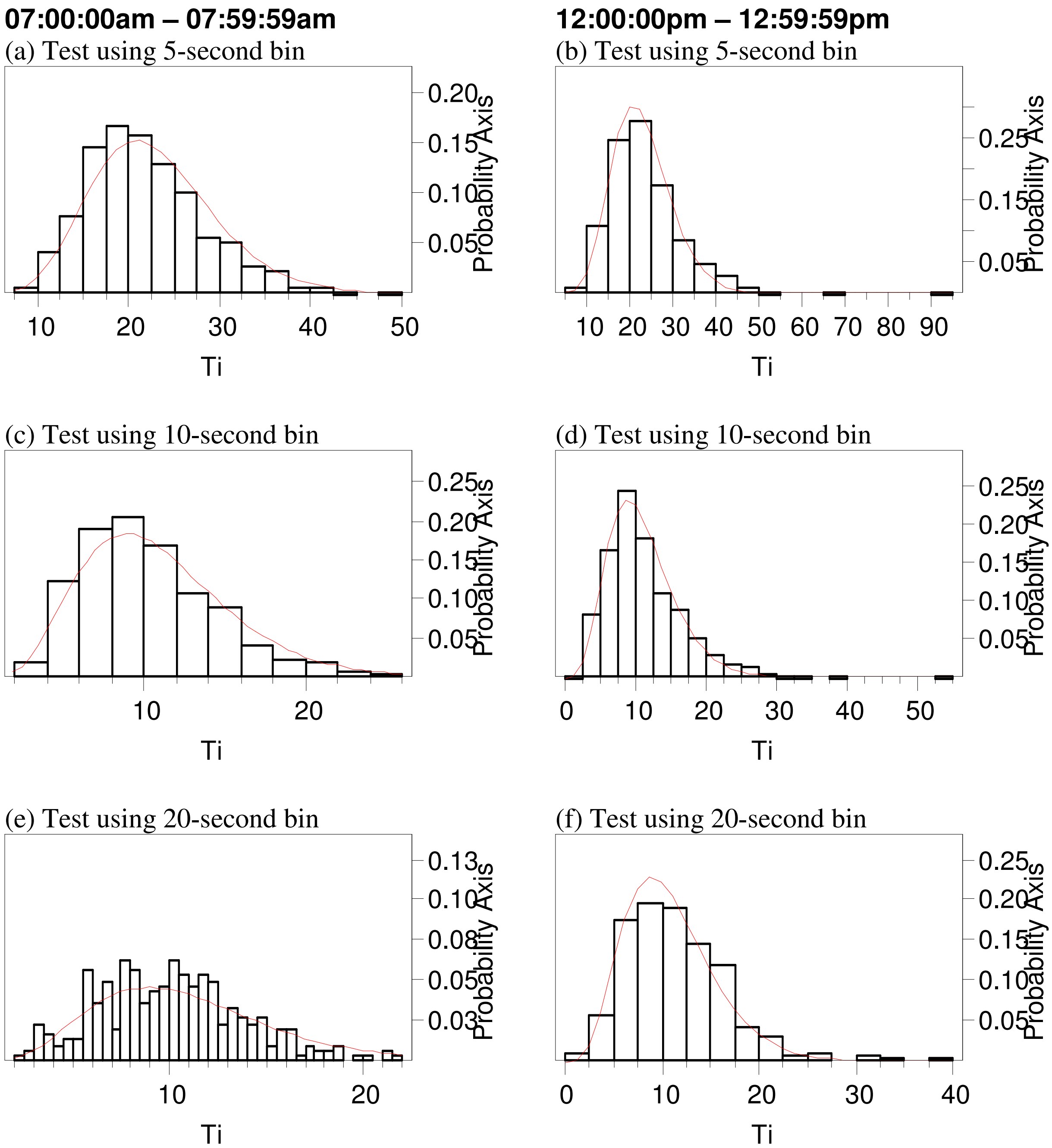


Figure 7: $\{T_i\}$ computed for the call arrivals to the VRU between 07:00:00am and 07:59:59am and between 12:00:00pm and 12:59:59pm on all the weekdays in June 2002 using various bin sizes.

We also applied the goodness-of-fit test on the calculated $\{T_i\}$ to have a quantitative idea as to how well these statistics fit to a chi-square distribution. The results are shown in Table 4.

	Entry to service queue						Entry to VRU					
	07:00:00am - 07:59:59am			12:00:00pm - 12:59:59pm			07:00:00am - 07:59:59am			12:00:00pm - 12:59:59pm		
	K-value	P-value	n	K-value	P-value	n	K-value	P-value	n	K-value	P-value	n
5-second bin 2-minute block	1.7051	0.0060	600	7.0002	0.0000	600	2.2589	0.0001	600	1.0569	0.2139	600
10-second bin 2-minute block	2.5353	0.0000	600	3.9856	0.0000	600	2.0140	0.0006	600	0.8588	0.4521	600
20-second bin 4-minute block	1.7858	0.0034	300	2.1367	0.0002	300	1.5987	0.0120	300	1.0139	0.2554	300

Table 4: Results of the goodness-of-fit test to a chi-square distribution on {Ti}.

From the results, it is evident that call arrivals to the service queue follow more closely to an inhomogeneous Poisson process during the period of 07:00:00am to 07:59:59am than those during the period of 12:00:00pm to 12:59:59pm. This is consistent with the results found in the earlier tests.

The call arrivals to the VRU during the period of 12:00:00pm to 12:59:59pm clearly demonstrate fidelity to a Poisson process, a result similar to that from the second test. However, for the arrivals during the period of 07:00:00am to 07:59:59am, the P-values are lower than expected, indicting that the call arrivals during that time frame might not quite follow an inhomogeneous Poisson process. This might be somehow related to the inherent nature of the test itself and is an area that can be further explored.

Another observation from the table is that the tests that make use of larger bin sizes seem to result in larger P-values, although this is not particularly significant. This indicates that the larger bin sizes may have eliminated some of the minor irregularities that the VRU have introduced into the system, but not completely.

5 Conclusion

A conclusion that can be drawn is that the call arrivals to a call center (into the VRU) follow with reasonable fidelity an inhomogeneous Poisson process. However, this does not seem to be the case for calls that exit the VRU and enter the service queue. This suggests that there is something going on within the VRU that alters the distribution of the calls requiring service. The effect of the VRU is particularly pronounced during peak periods. This may be a result of some kind of mechanism or bottleneck within the VRU during peak periods that is causing the call arrivals to the service queue to deviate away from an inhomogeneous Poisson process.

The results from this study expose some of the irregularities in the VRU that cause the call arrivals rates to the service queue to exhibit non-Poisson behavior. Further studies will be needed give a better understanding of the nature of the distribution of calls entering the service queue at different times of the day. At the same time, it will certainly be useful to conduct further studies that examine the different call arrivals distribution and the impact on the staffing requirements and efficiency of call centers. Although these studies may only be a single component of the queueing theory, it will also have implications on studies relating to other aspects of the queueing theory.

Within the field of queueing theory, the study of call arrivals is closely related to studies on queue characteristics and service mechanisms. In the design of a particular queueing system, such as deciding on the queue discipline (whether the queue should be FIFO, LIFO or random) and service preemption (whether a service agent should stop processing a customer to handle emergency situations), many interrelated factors will have to be considered, the call arrivals being one of them. It is thus critical to combine our understanding of the call arrivals from this research with studies on queue characteristics and service mechanisms in order for us to have a better understanding of the system as a whole.

From a managerial perspective, this research can help to answer some of the practical questions dealing with resource allocation: What should managers do in order to minimize the operating costs of a call center and yet minimize queueing time? What is the optimal number of service agents that a call center should have at various times of the day? What are the consequences of replacing some of the service agents with automated voice response systems?

Finally, from a mathematical standpoint, the analysis of the call center data involves the use of different statistical tools for a Poisson process. These tests gave rise to somewhat contradictory results in some cases. A study of the assumptions and theories behind these tests may give a better idea of the circumstances where these statistical tools can be most suitably employed.

6 References

(1) Brown, L., Gans, N., Mandelbaum, A. Sakov, A., Shen, H., Zeltyn, S., and Zhao, L. (2002), "Statistical analysis of a telephone call center: A queueing-science perspective", *Technical Report,* University of Pennsylvania.

(2) Brown, L., and Zhao, L. (2002) "A new test for the Poisson distribution", *Sankhya: The Indian Journal of Statistics*, Series A, **64**, 611-625.

Gans, N., Koole, G., and Mandelbaum, A. (2002), "Telephone calls centers: a tutorial and literature review", *Technical Report.*
Downloadable from http://iew3.technion.ac.il/serveng/References/references.html.

7 Appendices

A-1 Exponential Quantile Plots for Test 1

A-2 Macros to Calculate Chi-Square Statistics for Test 2