USE OF QUEUING THEORY IN DETERMINING OPTIMAL SUPER MARKET CHECK-OUT FACILITIES

By

John Y. Lu
yaotung

A THESIS

Submitted to the School for Advanced Graduate Studies
of Michigan State University of Agriculture and
Applied Science in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

Department of Agricultural Economics

1959

G 15163
6/19/61

ACKNOWLEDGMENTS

The author wishes to acknowledge his indebtedness to the following individuals who have directly or indirectly helped the author in the preparation of this thesis.

The financial assistance given to the author during his stay at Michigan State University by Dr. Lawrence L. Boger and the facilities accorded to him by the Department of Agricultural Economics are greatly appreciated.

Dr. Robert L. Gustafson has given much valuable guidance to the author in formulating the problem. Many of his ideas appear in the following pages, especially in Chapter IV. He has also read the manuscript and his comments have resulted in an improved presentation of the materials contained here.

Thanks are due to Mr. Mike Wood who suggested this problem to the author and contributed much useful information, and also to Mr. Tom McDermott for enabling the author to obtain data in connection with the check-out operation of a super market.

The author also had the benefit of comments by Drs. K. J. Arnold, W. A. Cromarty, and L. V. Manderscheid when the project was first being set up.

Bill Crosswhite and Peter Hildebrand have read parts of the manuscript and offered a number of useful suggestions concerning its style. The mathematical exposition in Appendix B has been considerably improved as a result of comments by Willard Sparks. Needless to say, any error remaining in the thesis is the author's sole responsibility.

The author cannot fully express here the great debt he owes to his many teachers, but he cannot refrain from singling out Dr. Clifford Hildreth to whom he owes his interest in econometrics. During the entire course of the author's graduate study, Dr. Hildreth has been most generous in giving encouragement and inspiration to the author.

The author's heartfelt appreciation goes to M.W.L. for the invaluable role she played in helping the author to complete his graduate work.

****

USE OF QUEUING THEORY IN DETERMINING OPTIMAL
SUPER MARKET CHECK-OUT FACILITIES

By

John Y. Lu

AN ABSTRACT

Submitted to the School for Advanced Graduate Studies
of Michigan State University of Agriculture and
Applied Science in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

Department of Agricultural Economics

Year 1959

Approved Clifford Hildreth

ABSTRACT

The check-out service provided by a super market appears to have all the necessary features that make up a typical queuing problem. There are customers who demand service; as each customer reaches a service channel, he receives a service. After a certain service time, he leaves. If the service channel is not immediately available to him, he joins a queue. In most cases, the management has little control over the arrivals of customers to the service area either in quantity or in time; it can, however, expand or contract facilities to meet a certain prespecified optimum criterion. The problem involves a balancing of the cost incurred by providing a certain amount of check-out facilities for a given period of time against the cost of losing customers in the future because of inferior service standards.

Recent developments in queuing theory provide the basis for systematically and quantitatively analyzing such a problem. The procedure is to first estimate, for given incoming traffic pattern and service practice, the probability distribution of the check-out system's being in each of all the possible states which are specified by the number of customers present in the system. Next, the cost associated with the system in each state is calculated. When the cost as well as the probability of the system being in each state is known, the expected cost per unit time of operating the check-out service under given conditions can be calculated. This process is repeated until the conditions which would result in the minimum expected cost are found.

The formal model used to estimate the state probability distribution is characterized as follows:

Input - the number of customers arriving per unit of time is a Poisson variate.

Service mechanism - the time required to serve a customer at a counter follows one of two different negative exponential distributions depending on whether or not a package boy is assisting a checker.

Queue discipline - "first come, first served."

This procedure was applied to data obtained at one of the large super markets in the Detroit area. The week was subdivided into five periods. The state probability distribution was estimated for each period and the quantities of interest such as the expected length of queue and the probability of more than a certain number of customers in a queue were calculated. Based on these quantities, expected costs were obtained and the service facilities which would generate the minimum expected cost were found.

Sensitivity of the choice of optimum service facilities to changes in estimated average arrival rate and average service rate was examined and was found not to be serious. The assumptions of Poisson inputs and exponential service times were tested. The number of customers arriving at the check-out area per unit of time follows closely a Poisson distribution. The negative exponential distribution did not appear to give the best fit to observations on service time. The assumption that service times are distributed by a gamma function was found to be more plausible. An alternative procedure based on the Monte Carlo method was proposed to take account of this fact.

TABLE OF CONTENTS

CHAPTER Page

I INTRODUCTION.. 1

What is a Queuing Problem?........................... 1
Objectives of this Study............................. 5

II QUEUING MODEL TO BE USED IN THIS STUDY..................... 7

A Simplified Representation of the Check-out Operation at a Super Market.................................. 7
Probability Distributions of Customer Arrivals and Service Time.. 9
Expected Values of Various Quantities of Practical Interest in the Problem Under Study.................. 10
Study of Volume of Daily Sales at One Detroit Super Market Under Study................................ 13
Review of Literature................................. 15

III SOME STATISTICAL ANALYSES OF CUSTOMER ARRIVALS AND SERVICE TIMES.. 21

The Goodness-of-fit Test............................. 22
Statistical Properties of Average Durations of Service Times... 30
Confidence Interval.................................. 32

IV OPTIMAL RULES FOR CHECK-OUT OPERATION...................... 36

A Criterion of Optimality............................ 37
General Method of Solution........................... 37
Some Specific Solutions.............................. 40
Optimal Check-out Rules for the Super Market in Detroit. 42

V SUMMARY OF RESULTS AND SUGGESTIONS FOR FURTHER STUDY....... 57

LITERATURE CITED.. 66

APPENDIX A -- STATE PROBABILITY DISTRIBUTION................ 68

APPENDIX B -- THE DISTRIBUTION FUNCTION OF ESTIMATED AVERAGE SERVICE TIME..................................... 78

LIST OF TABLES

TABLE		Page
I	Customer Arrival Goodness-of-Fit	28
II	Confidence Limits of Average Service Time	35
III	Estimates of Average Arrival Rates and Average Service Rates, and Their Confidence Intervals	35
IV	Optimum Check-Out Rules for Monday, Tuesday, and Wednesday, $\hat{\lambda} = 0.91$	49
V	Optimum Check-Out Rules for Thursday, $\hat{\lambda} = 1.53$	50
VI	Optimum Check-Out Rules for Friday, before 6 P.M., $\hat{\lambda} = 1.85$	51
VII	Optimum Check-Out Rules for Friday, after 6 P.M., $\hat{\lambda} = 2.55$	52
VIII	Optimum Check-Out Rules for Saturday, $\hat{\lambda} = 2.44$	53
IX	Optimum Check-Out Rules for Different Average Arrival Rates, and Different Criteria of Optimality	55
X	Effect of Variation in Estimates of λ and μ on Optimum Check-Out Rules	56

LIST OF FIGURES

FIGURE Page

1. Total Variable Cost of Operating a Queuing Process.......... 2

2. Daily Sales Volume at One Super Market in Detroit, Jan. 5-Feb. 21, 1959.. 14

3. Actual Frequency Distribution of Customer Arrivals.......... 23

4. Theoretical Frequency Distribution of Customer Arrivals..... 23

5. Comparison of Actual and Theoretical Customer Arrivals on Monday, Tuesday and Wednesday................................ 25

6. Comparison of Actual and Theoretical Customer Arrivals on Thursday.. 25

7. Comparison of Actual and Theoretical Customer Arrivals for Friday, before 6 P.M... 26

8. Comparison of Actual and Theoretical Customer Arrivals for Friday, after 6 P.M.. 26

9. Comparison of Actual and Theoretical Customer Arrivals for Saturday.. 27

10. Diagramatic Determination of the Penalty Coefficient d...... 46

CHAPTER I

INTRODUCTION

Shopping at a super market is a familiar experience for the general public. Every shopper is, to some extent, interested in super market operations whether he is an operations research analyst or not because it requires little technical training to understand the operations. As he goes through aisles, pushing a cart, selecting his purchases, and finally ending up at the check-out stand, he cannot help but wonder if there is a way of improving at least some phases of the operations.

In this study, the author is thinking along the same line as this super market customer except in a more direct manner. The main concern of the study is how to provide a high grade check-out service in the most economical manner.

What is a Queuing Problem?

There are many operational problems involving flow of customers[1] in which the following two types of condition are observed:

1) Units that require service must wait for service because there is a shortage of service facilities.
2) The servicing facilities must remain unused, not only because of the lack of customers in quantity, but also because of the nature of the time spacing between customer arrivals.

---

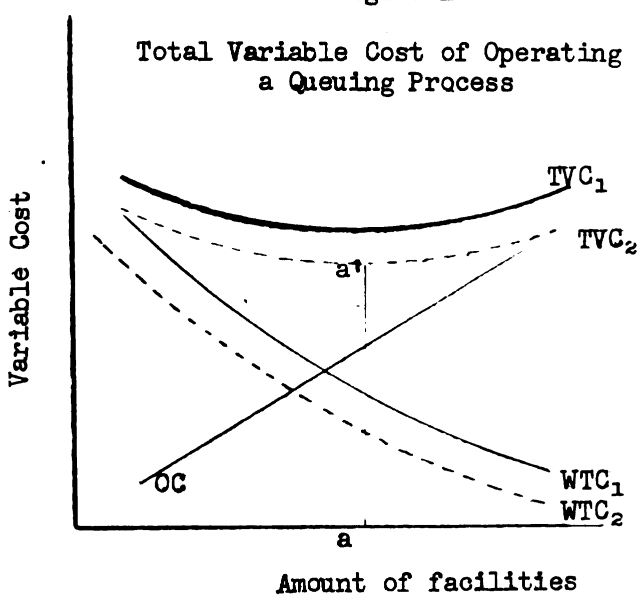
[1]In queuing theory, the term customer is not restricted to a person. It can mean an aircraft waiting to land at an airport or an automobile waiting to pass through a toll gate.

Either of these conditions results in the formulation of a waiting line. In the first instance it would be a queue of input units, and in the latter the units in a waiting line consist of service facilities. The shortage and surplus of service facilities are usually brought about by inability of a system to decide in advance a right amount of service facilities to provide to its customers; this inability, in turn, is explained by the random elements which influence demand for service both in time and quantity.

Under these conditions, a so-called queuing problem arises. The problem is to change the behavior of the arriving units, or the service facilities or both in order that the queuing process may be operated in as efficient and economic a manner as possible. This can be illustrated for the case where a queue consists of input units. Total variable costs (TVC) of operating a system that involves a queuing process are made up of operating cost (OC) and waiting time cost (WTC).

Figure 1

Total Variable Cost of Operating a Queuing Process



Here waiting time cost is defined as the cost of losing a potential customer because of insufficient service facilities. Operating cost may be assumed to increase approximately in proportion to the amount of facilities provided. On the other hand waiting time cost (for units being serviced) probably decreases at a decreasing rate as more service facilities are added to the system. It should be noted that waiting time cost is a joint function of service facilities and the behavior of arriving units. For each change in the behavior of arriving units, a new waiting time cost function and a corresponding total variable cost function can be drawn as they are indicated by dotted curves in the above Fig. 1.

There are at least three ways of minimizing the total variable costs.

1) For a given amount of facilities, find what behavior of arriving units will give a minimum variable cost. In Fig. 1, aa' is the minimum total variable cost.
2) Find what is a right amount of service facilities which corresponds to the minimum point of the total variable cost curve for a given behavior of arriving units.
3) By varying the two variables (behavior of arriving units and service facilities), the minimum point along the lowest total variable cost curve can be found.

In order to quantitatively solve a problem of this type, the tools of probability theory can be used to provide convenient methods of determining the relations between the flow of arriving units, amount of service facilities, and grade of service. With this knowledge the optimum grade of service can be determined in a logical manner; the amount of service facilities required at any time of day as well as the flow of customers to be permitted to the system can be specified in advance.

The probability theory can be applied in two different ways in solving a queuing problem. They are usually referred to as the mathematical and simulated sampling approaches.

The mathematical or analytical approach begins with specifications of probability distributions regarding customer arrivals and service times. Based on these specifications, relationships that describe the queuing process are derived by writing down statements about the probability of there being a given number of customers in waiting line under various conditions. These relationships can be solved for such quantities of interest as average number of customers in the system, expected length of waiting lines, etc. When costs arising from waiting time and operation of service facilities are known, one can analytically arrive at the conditions under which a minimum cost is attainable.

Solution of the problem by the mathematical approach has a neat appearance. However, sometimes it is difficult to specify explicitly arrival and service distributions. Even if they can be represented in terms of probability distributions, frequently a researcher is not able to arrive at mathematical statements describing the queuing process. If this is the case, the second approach may be used.

In the simulated sampling approach, a procedure known as the Monte Carlo technique is commonly used. By the use of a table of random numbers and of empirically determined probability distributions, statistics on arrivals and service time are duplicated in a mechanical way on a high-speed electronic computer. This method allows a research worker to study the effect of changing conditions without waiting for actual data over a long period of time. The simulated sampling approach has become more useful as a result of developments in high-speed computers.

In the current study, the first approach was followed and total variable cost of the check-out operation was minimized with respect to the amount of service facilities for a given arriving pattern of incoming units.

Objectives of This Study

The study attempts to test the applicability of a queuing model to the check-out operation of a super market. The check-out operation seems to have all the necessary features that make up a typical queuing problem. There are customers who demand service; as each customer reaches a service point, he receives a service. After a certain service time, he leaves. If the service point is not immediately available to him, he must wait his turn; in other words, he joins a queue. How long he will have to be in the waiting line depends on how many other customers have been to the store before him and also depends on how many service points are in operation. Management can do little about the flow of customers to a store but it can adjust the number of check-out counters available to customers so that a certain criterion can be met. In most super markets, allocation of man power necessary in providing check-out service as well as control of the number of check-out stands opened at any time is left to the discretion of the store manager. In making these decisions, he mainly relies on his experience and a rule-of-thumb work standard.

Hence it appears plausible to analyze the check-out service by means of queuing theory so that the operational problem involved can be dealt in quantitative terms. It is felt that this approach would be a supplement to a subjective way of dealing with the problem as in the past.

There are certain crucial assumptions on which a queuing model is based. These assumptions must be carefully checked in the light of the actual check-out operation to see if they are reasonable. Above all they must yield fruitful results.

To begin with, the analysis by queuing theory is based upon: (1) the length of time between two successive customers' arrivals and (2) the time used for servicing a customer. These two quantities are specified in terms of probability distributions. It is then natural to apply some statistical tests based on the hypothesized probability distributions to empirically derived distribution functions.

Like many a mathematical model, the queuing model need not correspond exactly to a real situation; if it can be regarded as an approximation to the actual check-out operation, then it can be used to provide decision criteria for the operation. A general method of determining the amount of check-out facilities which will minimize the expected total variable cost of the check-out operation is indicated, and three specific approaches based on this general method were proposed and they were applied to the data obtained at one of the large super markets in Detroit.

CHAPTER II

QUEUING MODEL TO BE USED IN THIS STUDY

Various queuing models can be constructed by altering specifications in regards to: (1) the number of service "channels," (2) probability distributions of customer arrivals and service times, (3) queue discipline, and (4) queue length. In the current study, the formulae of operational interest for the queuing model which has multiple exponential "channels" with Poisson arrivals, infinite queue and strict queue discipline[1] were applied to the super market check-out operation in order to determine an optimal way to provide check-out service to customers from the standpoint of management.

A Simplified Representation of the Check-out Operation at a Super Market

In order to apply the above model, first consider a super market which has a finite number of check-out stands. These check-out stands have identical service mechanisms and each of them can operate independently of the others. Furthermore, it is assumed that each stand

---

[1]Multiple exponential channels mean that there are more than one service point and the time needed for a customer to go through a service point follows the negative exponential distribution. Poisson arrivals refer to the assumption that the number of arriving customers per unit time is a Poisson variate. Infinite queue refers to situations in which every arriving customer must join the queue no matter how long it happens to be. Theoretically the queue may become infinite. Strict queue discipline is the so-called "first come first served" rule.

can be operated at the two different levels of average check-out rate, say μ_1 and μ_2. In this analysis, the number of check-out stands to be made available to customers at any given time will be less than or equal to the number of check-out stands that the super market has at the outset of the analysis.

Next, customers are considered to arrive at the check-out area at a certain average rate. Suppose there are m-k check-out stands which have the average service rate of μ_1 and there are k check-out stands with the average service rate of μ_2. A rule adopted here is that $\frac{m-k}{m}$ x 100 percent of the incoming customers will be served by the check-out stands with the average service rate μ_1, and the remaining portion of customers will be serviced through those stands which have the average service rate μ_2.

It is shown in Appendix A that adopting this rule will reduce the problem to a more familiar case of m equivalent service channels, i.e., each of the m check-out stands has the average service rate, say μ, where μ is

$$\mu = \frac{(m-k)\mu_1 + k\mu_2}{m} .$$

As soon as each customer arrives at the check-out area, he moves into any check-out lane which is free to service him. If he finds all the lanes are occupied, he is assumed to form a hypothetical common queue. Under these conditions, a queue exists when the number of customers in the check-out area exceeds the number of operating check-out stands at any time and their difference is the length of queue.

Probability Distributions of Customer Arrivals and Service Time

The main feature of a queuing model rests in its characterization of the input process and capacity of service mechanism in terms of probability distributions. It is assumed that the customers arrive at "random," i.e., the number of arrivals per unit time is a Poisson variable. Perhaps this is the simplest hypothesis about the input process. From this assumption, it follows that the time interval between two consecutive arrivals has the negative exponential distribution.[2] As to the service time, i.e., the time necessary for a particular customer to be served, the assumption is that successive service times are statistically independent of one another and each is distributed by the negative exponential distribution. It should be noted that two types of service time are considered here.

At first, these artificial schemes of representing customers' arrivals and service times may not seem realistic, because the information that no customer has arrived at the check-out area for, say, ten minutes will generally increase our expectation that a customer will show up in the next minute. It is also not natural to assume that service time obeys a negative exponential law, because one would intuitively feel that the probability of service time approaching a very short duration of time must be close to zero. They have been, however, found useful in many situations which seem to have features similar to the check-out operation. Besides, considerable mathematical

---

[2]See Feller, (1950), p. 364.

simplifications can be achieved by these assumptions. One of the main objectives of this study, as stated at the outset, is to examine the applicability of these schemes to the check-out operation. These assumptions are stated more completely in Appendix A.

Expected Values of Various Quantities of Practical Interest in the Problem Under Study

After the fluctuations of customer arrivals and service times have been expressed in terms of probabilities, other measurable quantities associated with the check-out operation will be considered as stochastic variables which vary with time about some average values. There are two possible ways of studying these stochastic variables. The first deals with the steady state or stationary phenomena of the variables. The main purpose of this first approach is to determine how the variables behave in the long run. The second approach has to do with the transient behavior of the variables, i.e., the exact behavior of the stochastic variables as a function of time.

In the current study, the first approach was adopted to study the stationary structure of the check-out process. Intuitively, one feels that the check-out process will approach a steady state, because there are restoring forces within the system which attempt to keep down the length of queue. In many cases, a probabilistic picture of steady state will give sufficient insight for one to be able to calculate important quantities of the system and predict the system's overall behavior.

Suppose the whole check-out process is considered as a system. This system can be in a number of possible states, which are specified by the number of units in the system, waiting for service, in service, etc. The steady state solution would give us the probability that the system is in each of the possible states. From these probabilities, average values of the various quantities of interest (mean number of customers in the system, average length of queue, etc.) and derived probabilities such as the probability that there are more than a certain number of customers waiting in each check-out lane can be calculated.

The steady state distribution is given in the following. The derivation of this distribution is set forth in Appendix A. Although a minor modification had to be made to take account of the fact that not all the service points have the identical average service rate, the mathematical development in Appendix A essentially follows the presentation of Feller (1950). The author assembled all the assumptions which are needed to derive the steady state distribution and presented the logic underlying the derivation of steady state distribution in detailed form. The purpose is to facilitate the understanding of the technique used in this study.

Let

P_n : The probability that there are n customers in the check-out system.

ρ : The ratio of the average arrival rate to the average service rate per channel. λ/μ. μ is defined as in P. 8.

m : The number of check-out stands that are rendering service to customers.[3]

---

[3]The quantity ρ/m is often called the relative traffic intensity. The term "relative" implies that the traffic intensity is measured in relation to the capacity of the system.

The steady state distribution is as follows:

$$P_o^{-1} = \sum_{j=0}^{m-2} \frac{\rho^j}{j!} + \frac{\rho^{m-1}}{(m-1)!} \frac{m}{m-\rho}$$

$$P_n = P_o \frac{\rho}{n!} \qquad 0 < n \leqq m \qquad (2.1)$$

$$P_n = P_o \frac{\rho}{m!m^{n-m}} \qquad n \geqq m$$

Given the explicit expressions for the state probabilities, as in (2.1), it is relatively straightforward to obtain the expressions for expected number of customers in the system, expected number of customers waiting for service, derived probabilities such as the probability of the length of queue being less than a certain fixed number. Some useful formulae are presented below and their derivations are also given in Appendix A, pp. 76-77.

Let

L_s : The expected number of customers in the system.

L_q : The expected number of customers in waiting line.

q^* : A given constant.

$$L_s = P_o \sum_{n=o}^{m-1} \frac{\rho^m}{(n-1)!} + \frac{\rho^{m+1}}{(m-1)!(m-\rho)^2} + \frac{\rho^m m}{(m-1)!(m-\rho)} \qquad (2.2)$$

$$L_q = P_o \frac{\rho^{m+1}}{(m-1)!(m-\rho)^2} \qquad (2.3)$$

$$\text{Prob. (length of queue} > q^*) = \frac{P_o \; \rho^{m+q^*+1}}{m!m^{q^*}(m-\rho)} \qquad (2.4)$$

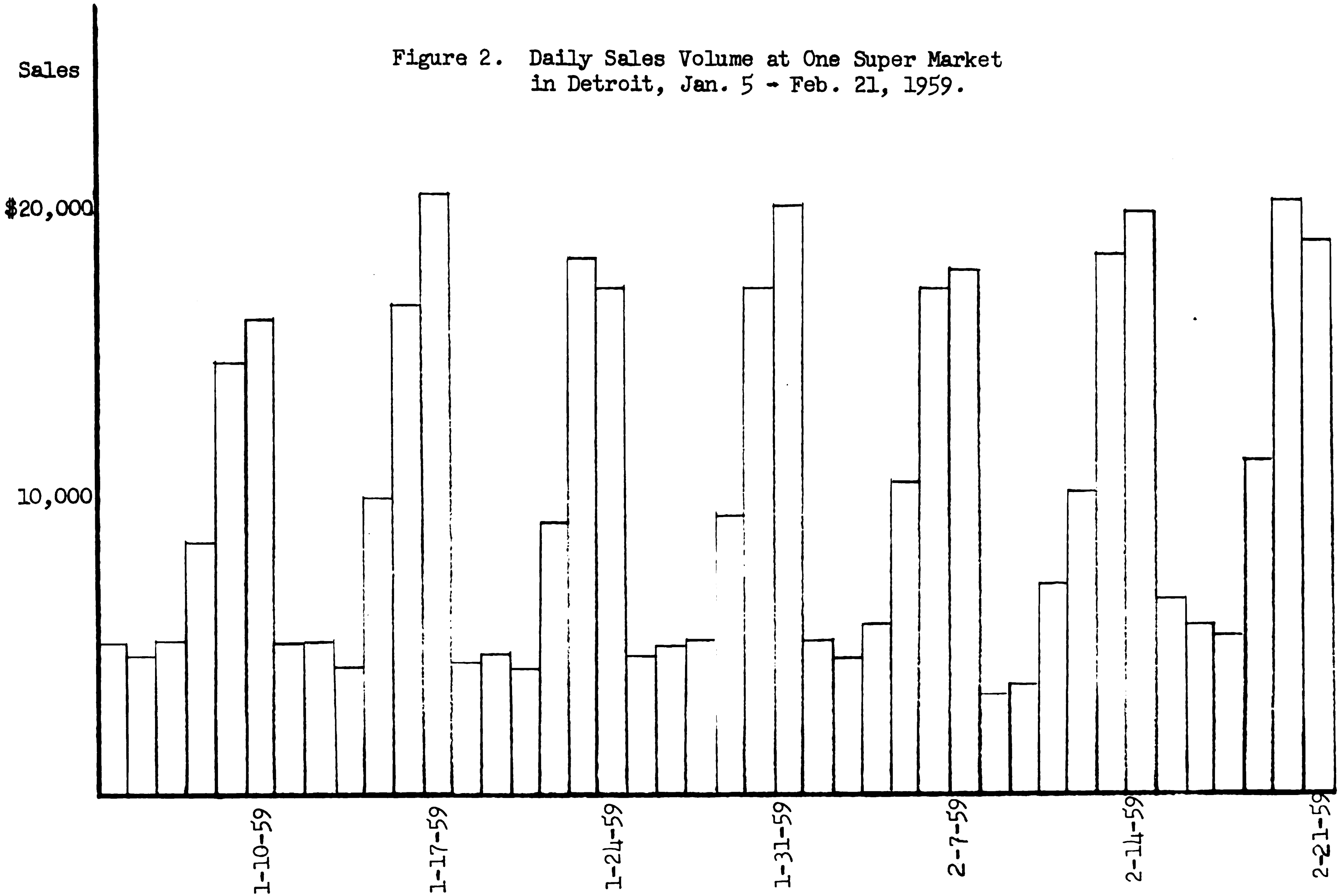
Study of Volume of Daily Sales at the Detroit Super Market Under Study

In order for the steady state solution to be valid, the average customer arrival rate and the average service rate must, roughly speaking, be neither rising nor falling. If they are stable, i.e., their fluctuations are short-term fluctuations around their constant mean values, the state probabilities and derived average values will be independent of time.

As long as there is no basic change in check-out facilities, it is not too unrealistic to assume that the average service rate remains stable. However, average customer arrival rate does fluctuate. In fact, it is closely related to sales volume per unit time. Inspection of daily sales volume at this store in Detroit for a period of about two months indicated that variation is nearly periodic with periodicity of one week. This is shown in the following diagram in which sales volume for each week day in January and February of 1959 is plotted. Each vertical block represents a volume of sales for one day. A low sales volume for the first part of the week is in contrast with heavy week-end sales. Sales on Thursday seem to constitute a group by themselves.

Study of this diagram suggests that the week could be subdivided into at least four periods so that within each period the customer arrival rate may be reasonably stable. The first three days of the week were combined into one period, and the remaining three days were analyzed separately. Initially it was conjectured that Friday and Saturday would have almost identical traffic intensity, judging from the volume of their daily sales. Discussion with the store manager and

Figure 2. Daily Sales Volume at One Super Market in Detroit, Jan. 5 - Feb. 21, 1959.



the others who are familiar with the check-out operation, however, revealed that the traffic pattern on Friday is quite different from that of Saturday. On Friday before six o'clock, the incoming traffic is usually not much faster than an ordinary week day, but the traffic picks up considerably after six. Actually, more than half of the day's sale is usually made in the last two or three hours before the store closes. Hence, eventually Friday was not only studied separately from Saturday, but within Friday two separate analyses were made.

Review of Literature

In this section, the main sources of the theoretical study about the queuing model used here as well as the sources for other types of models are mentioned. Some applied works which have come to the author's attention and were found useful in connection with the preparation of this thesis are also discussed.

The pioneering work in queuing theory was undertaken by A. K. Erlang some fifty years ago at Copenhagen in connection with telephone switching problems. Most of his papers are contained in a memoir prepared by Brockmeyer _et al._, (1948). An account of Erlang's work is also given by Fry (1928).

More recently, Feller (1950) discussed the queuing process as an example of simple time-dependent stochastic process and derived the state probability distribution in the case where the number of arrivals per unit time is a Poisson variable and service time is distributed by a negative exponential distribution. This was the model tested in this thesis.

An excellent survey of queuing theory is also to be found in the two articles by Kendall (1951, 1953). He considered, in the first paper, a queuing process with Poisson input and no specific assumption with respect to service time, i.e., the service time distribution can be one of the following three distributions: 1) constant or regular, 2) negative exponential, and 3) Erlangian.[4] The distribution of queue size in statistical equilibrium was derived. In the second paper, the assumption regarding the input process was stated in a more general form--the inter-arrival times were independently and identically distributed in an arbitrary manner. However, the service mechanism was supposed to be characterized by the negative exponential service time. The analysis of the system was carried out by applying the method of the imbedded Markov chain.[5] The ergodic[6] behavior of this Markov chain was investigated.

A queuing system of the more general type was considered by Lindley (1952). No specific assumptions with respect to input and service mechanism were made except input is independently and identically distributed and service time is distributed by one of the three

---

[4]Let t, $0 < t < \infty$, denote the inter-arrival time and A(t) its c.d.f. The density of the Erlangian distribution is

$$dA(t) = \frac{(k\lambda)^k}{\Gamma(k)} e^{-k\lambda t} t^{k-1} dt.$$

For $k = 1$, $dA(t) \rightarrow \lambda e^{-\lambda t}$ negative exponential.

For $k = \infty$, $dA(t) \rightarrow 0$ constant.

[5]It is a method of transforming a sequence of random variables into a process which satisfied the Markovian property. See Kendall (1953), pp. 341-342.

[6]Loosely translated, it means asymptotic properties.

distributions mentioned in connection with the papers by Kendall

The queuing systems considered by the aforementioned authors are all based on the strict queue discipline. The system becomes increasingly complex to analyze as one allows for a more flexible rule. Although there is one article by Holly (1954) on this subject, the problem does not seem to have been treated extensively by the statistician.

The theory of queues has a surprisingly wide range of applications. The four studies reviewed here represent only a small sample.

The first empirical study is an example of the classical application of queuing theory. The study was conducted by Molina (1927). The objective of his study was to find a way to operate the telephone trunking system economically yet consistent with good telephone service to the subscriber. In other words, the purpose of the study is to find a compromise between the number of circuits and the amount of equipment and the time necessary to complete a call. He assumed the number of incoming calls per unit time to be a Poisson variable and considered two alternative assumptions regarding holding time,[7] viz., exponential holding time and constant holding time. It is to be noted that in the case of the constant holding time, he was able to obtain only approximate solutions. The probability of a delay, which is greater than an interval of a certain length, was calculated for a given number of trunks under various arrival rates of incoming calls. These probabilities were used as action criteria in deciding if it was necessary to

[7]"Holding time" is a terminology of the telephone industry, and it is equivalent to service time.

add more circuits. Molina did not explicitly introduce the concept of cost into his study.

Edie (1955) applied the theory in analyzing traffic delays at toll booths operated by the New York Port Authority at the Lincoln tunnel, George Washington bridge, etc. Since the major portion of expenses necessary in manning these toll booths is the salary of toll collectors, it was natural to consider a way to economize the toll collecting operation by reducing the number of collectors, yet at the same time maintain the policy of giving uniformly good service to the public. In the past the allocation of manpower and controlling the number of toll booths opened at any time were left to the discretion of the toll sergeants. The quantity of the service tended to vary appreciably from time to time. By means of the queuing theory Edie was able to provide methods for dealing with the problem in quantitative terms and was able to reduce toll collection expenses without impairing the quality of service. The queuing system used is similar to the one considered by Feller (1952). He made careful analyses of the incoming traffic, but not much was said about service time distributions. Again, the problem seemed to have been tackled mainly from the engineer's point of view, because there was little discussion on the cost aspect of the problem.

In contrast to the first two studies, the cost concept was directly utilized together with the queuing theory in solving an inventory problem in the study made by Flagle (1956). The stochastic element involved in the inventory problem is the random receipt of orders for the product. The problem was to determine the planned initial stock level that yields a minimum sum of expected storage and depletion costs. The first step in solving the problem was to calculate the probability of experiencing each of the possible inventory states during a reorder period. Next the costs associated with the system in each of the possible states were estimated. There were three major expense items that comprised the cost. They were interest charges being incurred for carrying a certain level of stock, storage and handling fees, and loss of orders due to the inability of the system (shortage of stock) to fill an order. The costs associated with the first two items are proportional to the size of the inventory. The cost incurred by a loss of a customer or a potential customer is inversely proportional to the inventory level. After state probabilities and associated state costs have been obtained, expected cost was calculated for a given value of average order rate and depletion reserves. The calculation was repeated by varying the value of depletion reserve until that value of depletion reserve which yielded a minimum expected cost was found. This technique was applied in minimizing the expected cost arising from the queuing system studied here.

In spite of its wide use by the engineer, the theory of queues has not been applied frequently in the field of agricultural economics. As far as the author is aware, there has been only one report on an applied queuing problem in the Journal of Farm Economics. Cox et al., (1958) applied the theory in determining livestock unloading facilities. The problem was to decide on the number of docks needed to meet certain management requirements which were specified in terms of either maximum allowable waiting time for a truck to unload or the length of the maximum allowable waiting line. The simulated sampling approach was adopted.

First probability distributions of arrivals and service time were estimated. Then an arrival number was drawn at random from the estimated distribution. For each arriving truck, a service time was also drawn at random from the estimated service time distribution. And a record of occupancy of the dock was kept. The process was repeated on a high speed computer until the probabilistic features of a waiting line formation were known. In this study, the authors were interested only in the probability distribution of the number of trucks waiting in line and the cost associated with this probability distribution was not discussed.

In closing this section, it should be noted that a comprehensive bibliography of queuing problems can be found in the book edited by McClosky and Coppinger (1956).

CHAPTER III

SOME STATISTICAL ANALYSES OF CUSTOMER ARRIVALS AND SERVICE TIMES

The analysis by queuing theory is based upon: (1) the length of time between two successive customers' arrivals, and (2) the time used for servicing a customer. These two quantities are specified in terms of probability distributions as mentioned in the previous chapter. It is then natural to apply some statistical tests based on the hypothesized probability distributions to empirically derived distribution functions. Although the hypothesis is that the probability distributions are negative exponential in both cases, empirically derived distributions of the first quantity were not directly tested against the negative exponential functions for goodness-of-fit. Instead, empirical probability distributions of the number of customers arriving per unit time were checked against Poisson distributions. As mentioned briefly in the previous chapter, assuming the customer arrivals per unit time to be distributed by a Poisson distribution implies that the length of time between two consecutive arrivals has a negative exponential distribution.[1] Furthermore, it is simpler to count the number of customers arriving at the check-out area per unit time than to directly measure how much later a customer arrives after his predecessor.

---

[1]This was demonstrated by Feller in his Probability Theory and Its Applications, pp. 363-367

The Goodness-of-fit Test

The first type of data recorded was customer arrivals at the check-out area. Observations were taken by counting the number of customers arriving per one minute interval, with an exception of Thursday.[2] An interval of one minute was used because it was about the shortest that permitted the observer to make recordings without losing the count. Observations were taken for each of the five periods within the week, and an actual frequency distribution was constructed from observations for each period by computing occurrences of each arrival class as a percentage of the total intervals observed. These percentages were then plotted against the arrival classes, as shown in Fig. 3, and frequency polygons were drawn. The frequency distribution with the higher average traffic volume per unit time tends to flatten out, and at its right-hand tail there is a tendency for the frequency to be higher.

In order to compare these actual frequency distributions with the theoretical distributions (in this case, Poisson), the theoretical distribution corresponding to each group of observations was obtained by estimating the mean from the observations and it was plotted in Fig. 4. The similarity of the curves in Fig. 3 to those in Fig. 4 is quite evident. An empirical distribution based on observations on Thursday's traffic arrivals was not included in Fig. 3 because observations were based on different time units.

---

[2]The arrival classes were based on a five-minute interval on Thursday.

Figure 3. Actual frequency distributions of customer arrivals.

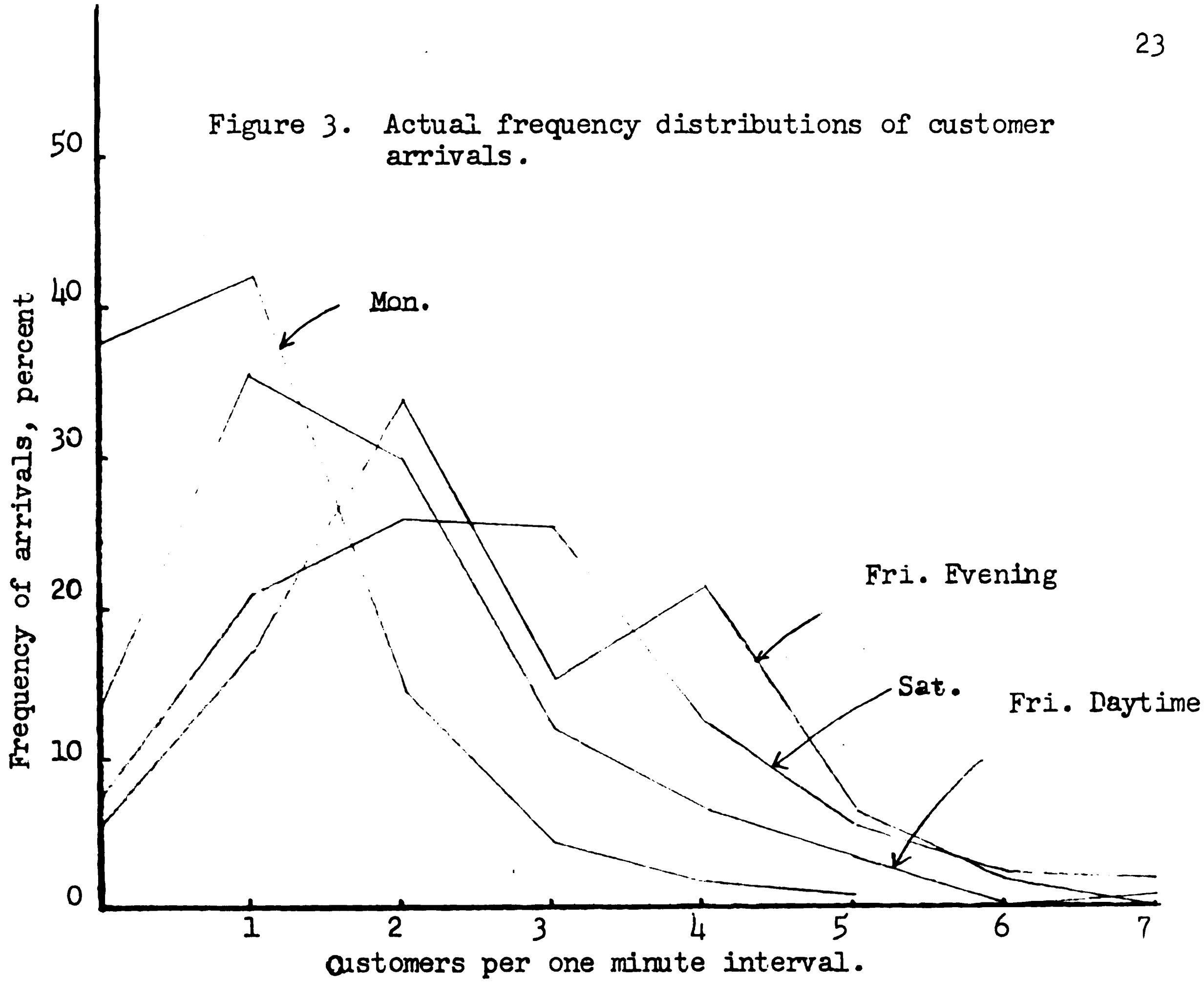
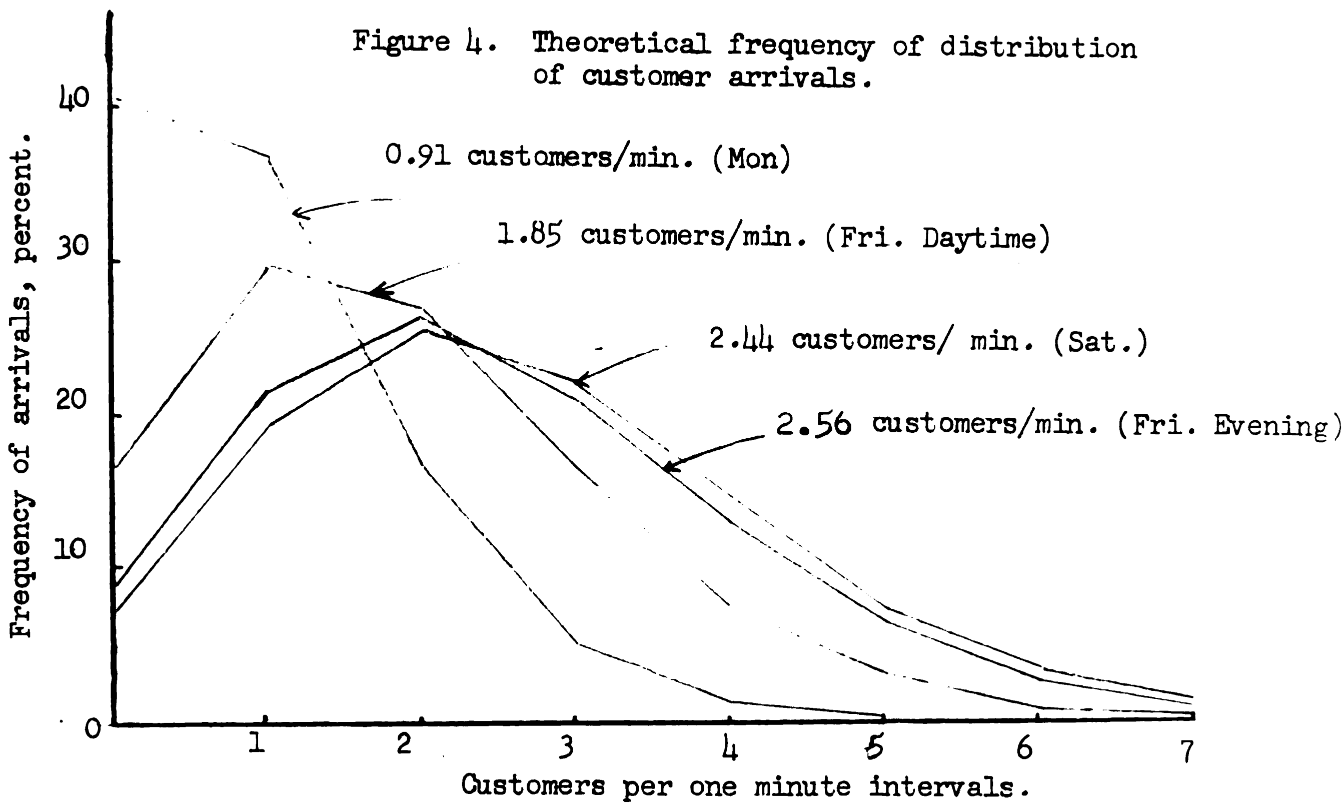


Figure 4. Theoretical frequency of distribution of customer arrivals.



An easier comparison between the actual and the two theoretical distributions, Poisson and normal, can be made by referring to Fig. 5, 6, 7, 8 and 9. In each diagram, the actual distribution and two theoretical distributions estimated from the same set of observations are plotted together. The mean arrival rate and standard deviation are also presented in the diagram. One feature to be noted is that the normal curve appears to fit slightly better than Poisson to observations on Thursday. This is probably due to the fact that arrival classes on Thursday's observations are based on a five-minute interval instead of on the one-minute interval. When the duration of observation interval is prolonged, frequency of occurrences of the event that an extremely small number of customers coming in during a five-minute interval would tend to be small. Hence, the empirical distribution resembles a familiar bell-shaped normal curve.

In addition to plotting frequency polygons, a statistical test was applied to see which of the two theoretical distributions gives the better fit to the data. The test statistic used was the familiar quadratic expression whose values are larger the farther the observed frequencies differ from their means as calculated under the hypothesized distribution. It is known to have asymptotically a chi-square distribution with T-K-1 degrees of freedom where T and K are the number of classes and estimated parameters respectively. When the test statistic has been calculated, a probability level of fit can be found in a table of chi-square distribution. A perfect fit would show a probability level of 1.00, but a fit showing a probability level better than 0.05

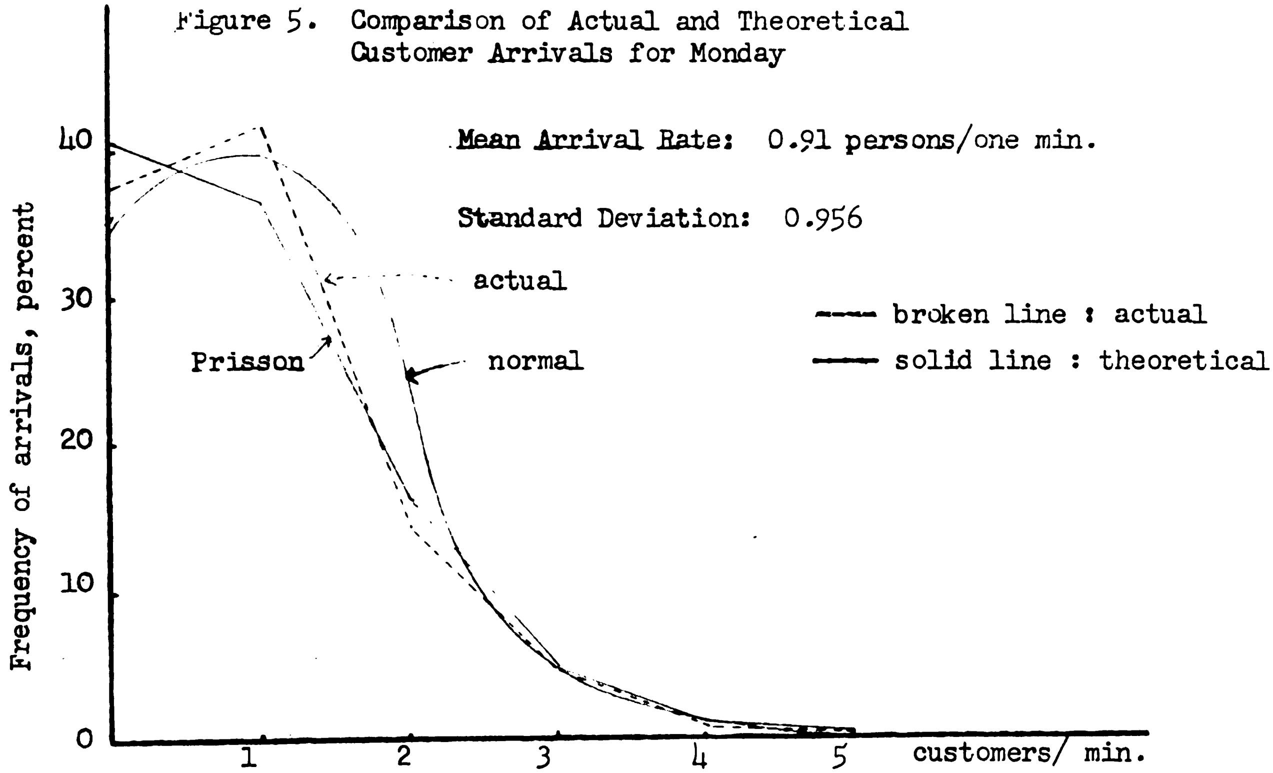


Figure 5. Comparison of Actual and Theoretical Customer Arrivals for Monday

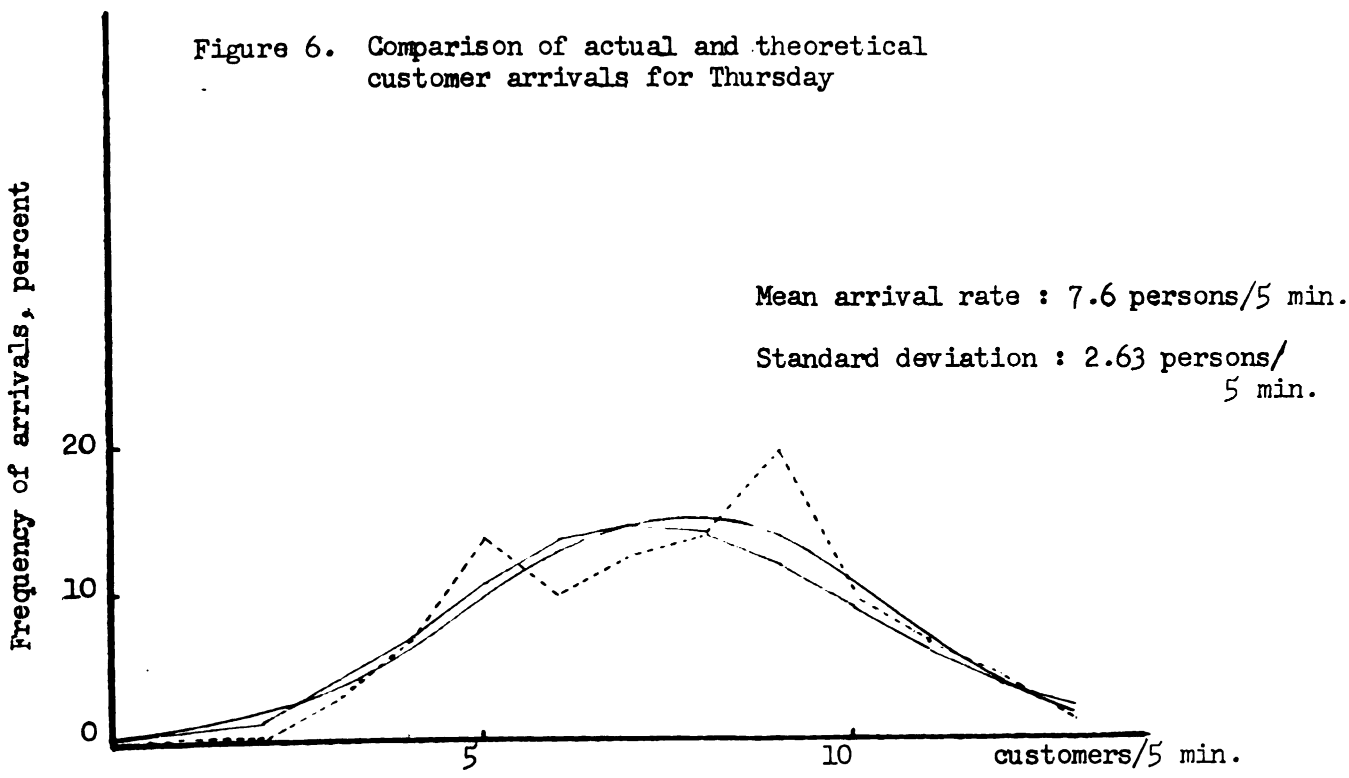


Figure 6. Comparison of actual and theoretical customer arrivals for Thursday

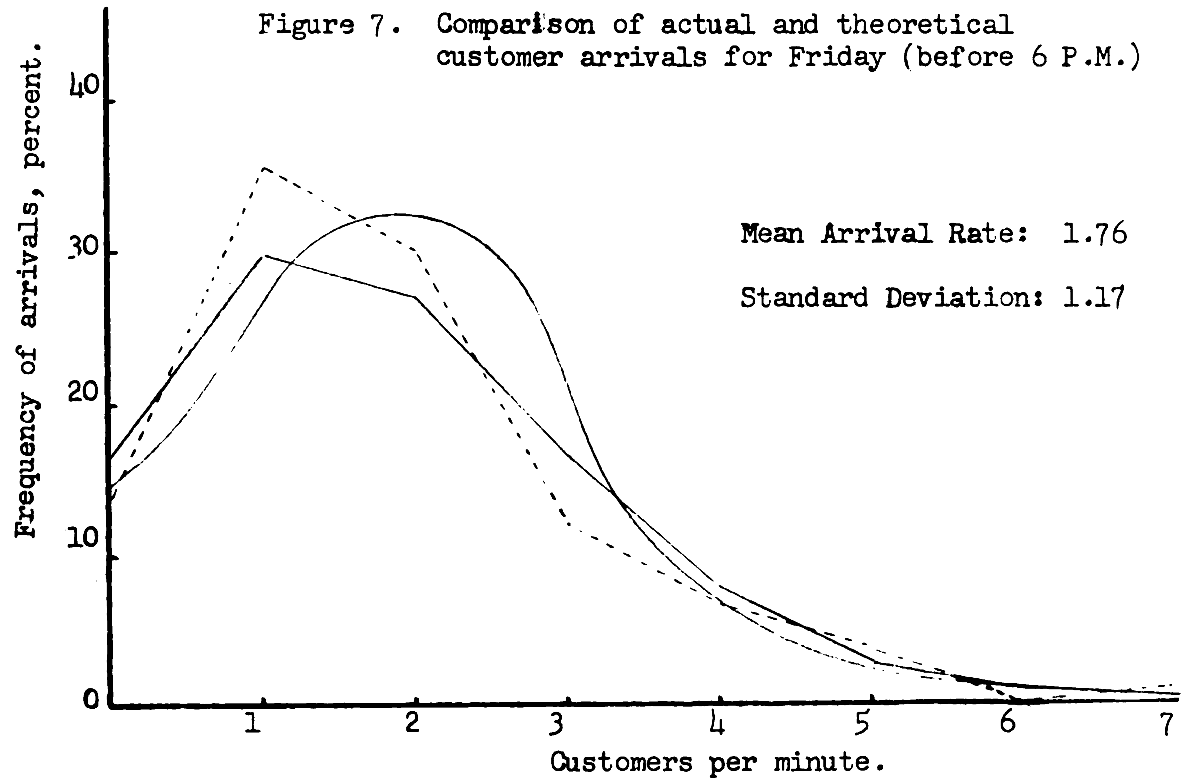


Figure 7. Comparison of actual and theoretical customer arrivals for Friday (before 6 P.M.)

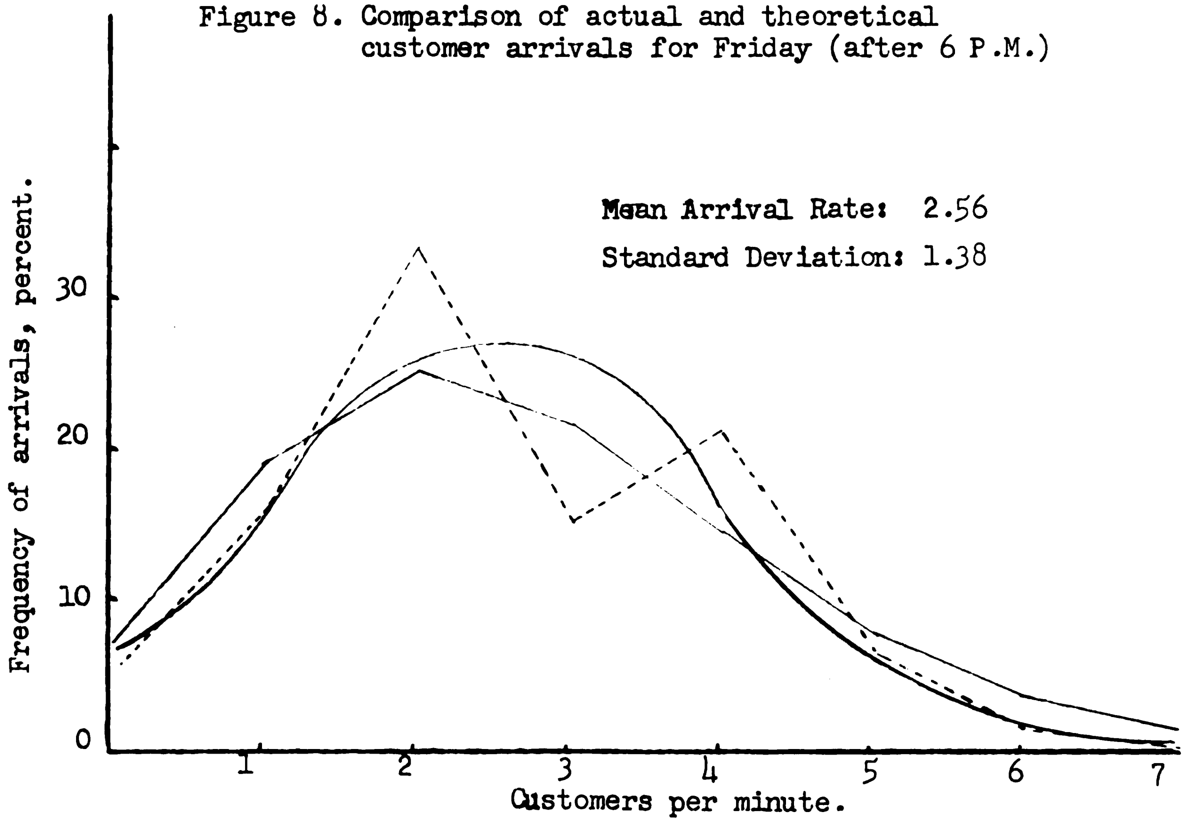


Figure 8. Comparison of actual and theoretical customer arrivals for Friday (after 6 P.M.)

TABLE I

CUSTOMER ARRIVAL GOODNESS-OF-FIT

Period	Average Arrival Rate Per Min.	Number of One Minute Intervals	Chi-square Statistic				The Significance Level at which H_0 is Rejected	
			Poisson	d.f.	Normal	d.f.	Poisson	Normal
Mon. Tues. Wed.	0.91	194	3.897	4	39.232	3	99%	*
Thurs.	1.50	365	7.155	12	6.681	11	90	90
Fri. before 6	1.77	145	8.107	6	18.858	5	25	*
Fri. after 6	2.56	113	13.193	6	10.484	5	5	5
Sat.	2.44	257	3.828	6	12.822	5	75	*

Abbreviations: * = The significance level is less than 0.05.
d.f. = Degrees of freedom.

a theoretical distribution seem to support the reasonableness of the assumption that the incoming traffic of customers is of the Poisson type. And they also seem to suggest that statistical tests based on the normal approximation would be adequate for a large sample.

Observations on another important quantity of the queuing theory, average service time, were also tested by the chi-square test. Since service time is a continuous variable, it was tested against a negative exponential function. As mentioned previously, two kinds of check-out service were considered; 1) a checker alone at a check-out stand, and 2) a checker as well as package boy is at the stand. In each case, service time is directly measured by clocking the duration of time taken by a checker when she starts to register a customer's purchases until

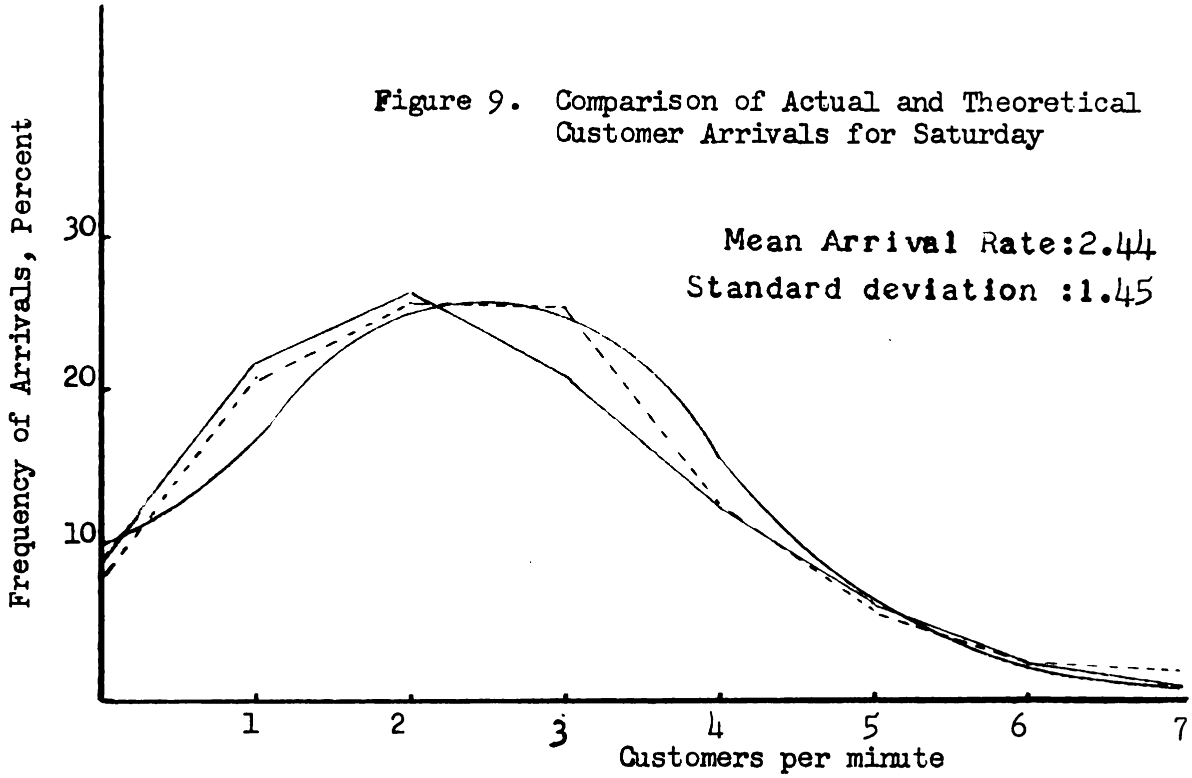


Figure 9. Comparison of Actual and Theoretical Customer Arrivals for Saturday

is generally taken to mean that the hypothesized distribution need not be rejected.

Results of applying the test are given in Table I. The fit of the Poisson distributions seems to be very good, especially at the low traffic volume per unit time. The fit of the normal curves are not as satisfactory as the Poisson, although it appears to show some improvement when the number of intervals observed is relatively large.

As indicated by the size of calculated chi-square statistic in Table 1, the fit of a Poisson distribution to an empirically derived arrival distribution tends to deteriorate for a relatively large average arrival rate. This may be explained by the fact that when the traffic becomes extremely congested the distribution has a tendency to become constant.

Results of chi-square tests as well as an inspection of those diagrams in which an empirical frequency distribution is plotted against she finishes packing groceries into paper bags. Computed chi-square statistics were relatively large which suggest that the data fit poorly to the negative exponential function. They were significant at the one percent level or less; in other words, values as large as the calculated statistics would be observed only one time out of a hundred trials by chance. This might have been expected because the assumption of a negative exponential function implies that the frequency of occurrences of service time being zero will be the largest, i.e., $\mu e^{-\mu t}$ reaches its maximum at $t = 0$. Ordinarily one would expect the highest frequency of check-out service times to be clustered around a certain average value which is different from zero. Inspection of the data indicated that they might have fitted better to another member of the gamma function family.[3]

Since the assumption of exponential service times makes the analytical approach to a queuing problem manageable, the state probabilities were calculated based on this assumption. Some alternative assumptions as well as possible different approaches to the problem are mentioned in the last chapter.

---

[3]Observations on the check-out time per channel (which is run by a checker alone) were fitted experimentally to a gamma function of the following form:

$$f(t) = \beta \frac{e^{-\beta t}(\beta t)^{r-1}}{(r-1)!}$$

The parameter r determines the shape of f(t); the parameter β determines its scale, $(r > 0, \beta > 0)$. Estimates of the parameters r and β by the method of moment were 1.843 and 0.957 respectively. The calculated chi-square statistic against this distribution was 8.793. Since the degree of freedom was 9, the null hypothesis that the observations came from the gamma distribution with $r = 1.843$ and $\beta = 0.957$ need not be rejected at the 50 per cent level.

Statistical Properties of Average Durations of Service Times

In a theoretical study of the probability structure associated with a queuing problem, values of average arrival and service rates are assumed to be known a priori. In application of queuing theory, however, these quantities must be first estimated from observations. The estimates are subject to sampling fluctuations. In this section some statistical properties of estimates of average duration of service times are discussed. Remarks made in this section on service time can be equally applied to customer's "idle" time,[4] because they are defined in an analogous manner.

First the sampling distribution function of an estimated average service duration was derived. From this sampling distribution, a confidence interval about the estimated average service duration can be obtained to give us some measure of assurance that the true parameter does lie within the interval. At the same time, an appropriate sample size can be determined in order that the probability that the sample mean will lie within a fixed distance from the population mean will meet at least a certain prespecified level. The confidence limits and sample sizes based on the exact sampling distribution are compared with those derived from the normal approximation. The normal approximation was found to be satisfactory in general.

The sampling distribution was obtained by first forming a joint density function of all the observations on service times and then

---

[4]Customer's idle time refers to a time interval between two consecutive customers' arrivals.

applying an appropriate transformation to the variables of the joint density function, and finally integrating out all the irrelevant variables from the transformed joint density function. Details of the above procedure are presented in Appendix B. The desired cumulative distribution function and density function are respectively as follows:

$$\text{Prob.}(\hat{\theta} < y) = 1 - e^{-n\mu y} \sum_{i=1}^{n} \frac{(n\mu y)^{n-1}}{(i-1)!} \qquad (3.1)$$

$$g(\hat{\theta}) = \frac{(n\mu)^{n}}{(n-1)!} \hat{\theta}^{\,n-1} e^{-n\mu\theta} \qquad (3.2)$$

where

$\hat{\theta}$: The maximum likelihood estimate of average service time. It is defined as

$$\hat{\theta} = \frac{\sum_{i=1}^{n} t_i}{n}$$

where t_i is the i-th observation on the service time.

n : The total number of observed service durations.

μ : The average service rate. Note that $\frac{1}{\mu} = \theta$ where θ is the "true" average service time.

The cumulative distribution function (3.1) is an incomplete gamma function. It can also be considered as a right-hand tail of the Poisson distribution with the parameter $(n\mu y)$. From the density function (3.2) the mean and variance of $\hat{\theta}$ can be evaluated.

$$E\hat{\theta} = \int_0^{\infty} \hat{\theta}\, g(\hat{\theta})\, d\hat{\theta} = \frac{1}{\mu}$$

$$\text{Var.}(\hat{\theta}) = \int_0^{\infty} (\hat{\theta} - E\hat{\theta})^2 g(\hat{\theta})\, d\hat{\theta} = \frac{1}{n\mu^2}$$

Since the mean and variance of $\hat{\theta}$ are known functions of the parameter μ, the normal approximation can be applied as a short cut in calculating a confidence interval of $\hat{\theta}$ as well as in determining a required sample size, when the sample size is large.

Confidence Interval

Although the sampling distribution function of $\hat{\theta}$ has been found, it is not independent of an unknown parameter. As can be seen from the equation (3.1), it involves the unknown parameter μ. In order to make a probability statement of the form:

$$\text{Prob.}\ (a<\hat{\theta}<b) = \int_a^b g(\hat{\theta})d\hat{\theta} = \gamma$$

where γ is the fiducial probability or confidence coefficient, it is convenient to have a distribution of $\hat{\theta}$ which is entirely free of the unknown parameter. This can be done by transforming the density function (3.2) according to the following formula:

$$\hat{\theta} = \frac{z}{\mu\sqrt{n}} + \frac{1}{\mu}$$

The new density function, say $\varphi(z)$, is as follows:

$$\varphi(z) = \frac{(\sqrt{n})^n(z+\sqrt{n})^{n-1}}{(n-1)!} e^{-\sqrt{n}\,(z+\sqrt{n})} \qquad (3.3)$$

Given this new density function, confidence limits which are independent of the unknown parameter μ can be obtained. To find the bounds of integration a and b in the following integral:

$$\int_a^b \varphi(z)dz = \gamma$$

or $$\text{Prob.}(z < b) - \text{Prob.}(z < a) = \gamma \qquad (3.4)$$

we would have to find the cumulative distribution function of z. From the transformation formula, z can be expressed in terms of $\hat{\theta}$.

$$z = \mu\sqrt{n}\ (\hat{\theta} - \frac{1}{\mu})$$

Hence

$$\text{Prob.}\ (z < b) = \text{Prob.}\ [\mu\sqrt{n}\ (\hat{\theta} - \frac{1}{\mu}) < b\ ]$$

$$= \text{Prob.}\ [\hat{\theta} < \frac{b+\sqrt{n}}{\mu\sqrt{n}}]$$

Therefore, the cumulative distribution of z can be easily obtained from that of $\hat{\theta}$.

$$\text{Prob.}\ (z < b) = 1 - e^{-\sqrt{n}(b+\sqrt{n})} \sum_{i=1}^{n} \frac{[\sqrt{n}(b+\sqrt{n})]^{i-1}}{(i-1)!} \qquad (3.5)$$

By means of the distribution (3.5), constants a and b were determined to enable us to make the following statements:

$$\text{Prob.}\ (z < b) = \gamma_1 \text{ and Prob.}\ (z < a) = \gamma_2 \qquad (3.6)$$

such that $\gamma_1 - \gamma_2 = \gamma$. There are infinitely many ways to choose γ_1 and γ_2 which will meet the above condition. Usually they are chosen in such a way that $1 - \gamma_1 = \gamma_2$.

Each of the equations (3.6) was solved for b and a respectively by use of a chart showing the cumulative Poisson distributipn for a number of different values for the parameter. This chart was prepared by

Bell Telephone engineers.[5] The function, $1 - e^{-\beta} \sum_{i=1}^{n} \frac{\beta^{i-1}}{(i-1)!}$ for values of β ranging from 0 to 200 and i from 0 to 270, is drawn in the chart. For instance, if the probability level γ_1 was set at 0.95 and the number of observations was 150, the corresponding value of β, say β_0, could be read from the chart. The constant b was calculated by simply solving the following equation:

$$-\sqrt{n}\,(b+\sqrt{n}\,) = \beta_0 \qquad (3.7)$$

Confidence limits of $1/\hat{\theta}$, which is the reciprocal of estimated average service time, for $\gamma_1 = 0.95$ and $\gamma_2 = 0.05$ are presented in Table II. They are based both on the exact sampling distribution and normal approximation, and are calculated for several different sample sizes. It appears from the table that the normal approximation is fairly adequate even for a sample size of less than 200.

In Table III, the estimators of average arrival rates for different periods of the week and of average service rate of a check-out stand which is operated by a cashier alone as well as that of the check-out stand which is manned by a cashier and bag boy are given. In the same table, their confidence limits are also presented. The confidence limits calculated were based on the normal approximation, because it was felt that the approximation would yield satisfactory results for sample sizes at hand. Besides, calculations could be considerably simplified by using the approximate method.

---

[5]G. A. Campbell, "Bell System Technical Journal," Jan. 1923.

TABLE II

CONFIDENCE LIMITS OF AVERAGE SERVICE TIME

Sample Size	Confidence Limits Exact Distribution		Normal Approximation	
110	$1/\hat{\theta}$ (0.845)	$1/\hat{\theta}$ (1.1636)	$1/\hat{\theta}$ (0.8432)	$1/\hat{\theta}$ (1.1568)
120	$1/\hat{\theta}$ (0.850)	$1/\hat{\theta}$ (1.1584)	$1/\hat{\theta}$ (0.8498)	$1/\hat{\theta}$ (1.1502)
130	$1/\hat{\theta}$ (0.862)	$1/\hat{\theta}$ (1.1538)	$1/\hat{\theta}$ (0.8549)	$1/\hat{\theta}$ (1.1451)
140	$1/\hat{\theta}$ (0.865)	$1/\hat{\theta}$ (1.1500)	$1/\hat{\theta}$ (0.8610)	$1/\hat{\theta}$ (1.1390)
150	$1/\hat{\theta}$ (0.870)	$1/\hat{\theta}$ (1.1467)	$1/\hat{\theta}$ (0.8657)	$1/\hat{\theta}$ (1.1343)
160	$1/\hat{\theta}$ (0.872)	$1/\hat{\theta}$ (1.1438)	$1/\hat{\theta}$ (0.8700)	$1/\hat{\theta}$ (1.1300)

TABLE III

ESTIMATES OF AVERAGE ARRIVAL RATES AND AVERAGE SERVICE RATES, AND THEIR CONFIDENCE INTERVALS

Period	Average Arrival Rate customers/minute	90 Per Cent Confidence Interval	
Mon., Tues., Wed.	0.91	0.798	1.022
Thurs.	1.53	1.423	1.636
Fri., before 6	1.85	1.628	2.038
Fri., after 6	2.55	2.313	2.809
Sat.	2.44	2.282	2.588
Type	**Average Service Rate customers/minute**	**90 Per Cent Confidence Interval**	
Cashier alone	0.40	0.367	0.442
Cashier and bag boy	0.82	0.730	0.885

CHAPTER IV

OPTIMAL RULES FOR CHECK-OUT OPERATION

In order to decide on what grade of check-out service that management should provide to its customers, knowledge about the probability structure of a queuing process associated with the check-out operation alone is not sufficient. It is necessary to bring a cost concept into the problem. By superimposing a cost structure on the queuing process, expected variable cost involved in the check-out operation can be calculated.

In the short run, management has little control over the flow of customers to the store. It cannot change the physical set-up of the store to install more check-out counters either. It can, however, regulate the average service rate of its check-out operation by varying the number of check-out stands to be operated as well as the number of bag boys to assist cashiers. These are two controllable variables that management would like to adjust in such a way that a given criterion of optimality is met.

Since the main interest of this study is determination of an optimal combination of the controllable variables under given conditions, little attention was given to institutional and managerial arrangements necessary to secure the number of cashiers or bag boys required to meet the criterion of optimality.

A Criterion of Optimality

The criterion adopted here is the minimization of expected costs incurred to the super market management in providing check-out service per unit time. Notion of expected values is introduced so that probability calculus can be used to take account of random fluctuations of important variables, such as the number of arriving customers and departing customers per unit time, in solving the problem.

Some simplifying restrictions were imposed when applying the above optimal criterion. It is quite common for a super market to set up an express check-out stand to accommodate those customers with a relatively small number of purchases. Since the volume of daily sales made at this counter is small in comparison to the total daily sales of the store, and the stand is kept open most of the time, it was not included in the present analysis. The second restriction is that there should not be more package boys than checkers at any time. This restriction is not too unreasonable because it is in accord with the general practice. The third restriction is that the presence of a long queue has no effect on the speed of service. This restriction can be somewhat relaxed if one allows different service rates for different types of traffic intensity.

General Method of Solution

The first step is to calculate a stationary distribution which describes the equilibrium probability of the system being in each of all the possible states. This stationary distribution, as explained in

Appendix A, does not change with time and is almost always independent of where the system was in time 0. The stationary distribution depends on the probability distribution of customers' demand for service and also on a rate at which customers are serviced. Since this service rate is a function of how many cashiers and bag boys are at work which are controllable by management, a proper selection of these variables can be made to satisfy the optimal criterion.

Next consider a cost structure which is generated by choosing a certain combination of the controllable variables. For each selection, certain costs must be incurred. The most relevant portion of these costs can be classified into two types. They are: 1) wages of cashiers and package boys who are providing the check-out service, and 2) cost incurred by a loss of the store's good will due to frequent formations of an unnecessarily long queue which is caused by inadequate check-out facilities. The first kind of cost can be readily estimated. It is independent of the incoming traffic intensity. On the other hand, the second type of cost is rather imponderable. The thing to be quantified is the adverse effect of keeping a customer waiting too long and too frequently for check-out service. He may become impatient and decide that he will not come back in the future. This cost can be considered to be a function of the expected length of queue and probably it is reasonable to say that the cost will increase at an increasing rate as the expected queue length becomes large.

The above discussion can be more conveniently and precisely stated with use of symbols.

Let

$P_n[\lambda, \mu(m,k)]$: Probability that there are n customers in the system.[1]

$C_1[m,k]$: Cost per unit time of employing m cashiers and k bag boys, where $m \geqq k$.

$C_2[L_q(\lambda, \mu)]$: Penalty cost per unit time.

Then the expected cost for providing the check-out service, given fixed values of λ and μ is

$$C_{\lambda,\mu} = \sum_n (C_1 + C_2)\, P_n \qquad (4.1)$$

The optimal criterion dictates that the expected cost (4.1) is to be minimized with respect to $\mu(m,k)$ for a given value of λ. Hence the equation (4.1) is evaluated at all the feasible values that m and k may take.

Evaluation of the expected cost by means of (4.1) can be somewhat simplified, if one notes the fact that costs of employing a certain number of cashiers and bag boys are independent of the stationary distribution P_n. The equation (4.1) can be reduced to the following form:

$$\begin{aligned} C_{\lambda,\mu} &= C_1 \sum_n P_n + \sum_n C_2 P_n \\ &= C_1 + \sum_n C_2 P_n \end{aligned} \qquad (4.2)$$

---

[1]The functional notation is to emphasize the dependence of the state probability on the average arrival rate and average service rate. The average service rate, in turn, is a function of the number of checkers and package boys, (see p. 41 of this Chapter). Similar remarks apply to the cost functions C_1 and C_2.

Hence the problem may be stated as follows:

Minimize $C_{\lambda,\mu}$ with respect to $\mu(m,k)$.

Because of one of the simplifying assumptions that there should not be more bag boys than cashiers at any time, the function $\mu(m,k)$ would be defined at most at $\frac{m(m+1)}{2}$ points. Since in most super markets the number of check-out stands will rarely exceed ten, evaluation of the cost function $C_{\lambda,\mu}$ is not unmanageable.

Some Specific Solutions

Three different objective functions based on (4.2) were formulated in this study. Their differences lie in the nature of the penalty cost function assumed. In the first formulation, a specific penalty function was introduced into the objective function. In the remaining two, the penalty cost was not included in terms of the dollar values of the "good will" lost as a result of keeping one customer waiting in line for one time unit.[2] However, the length of queue was incorporated indirectly into the cost function as a constraint of minimization.

1) Minimize $(m - k)W_1 + kW_2 + \delta(L_q)$
   with respect to m and k, (4.3)

where

m: The number of counters in operation or the number of checkers.

k: The number of bag boys, $(0 \leq k \leq m)$

---

[2]"Good will" lost can be considered as the discounted net value of future business lost.

W_1: Cost per unit time of operating a counter without a bag boy

W_2: Cost per unit time of operating a counter with a cashier as well as a bag boy

L_q: Expected length of queue which can be computed by formula (A.18) in Appendix A.

δ : Penalty cost function which is assumed known to management.

Suppose δ is a linear function, the cost function (4.3), may be written as follows:

$$(m-k)W_1 + kW_2 + d \frac{P_0 \rho^{m+1}}{(m-1)!(m-\rho)^2} , \qquad (4.4)$$

where d is considered as the dollar value of the good will lost as a result of keeping one customer waiting in line for one unit time and $\rho = \lambda/\mu$. Since there is an average of L_q customers waiting per unit time, the average penalty cost per unit time is dL_q. In the above formulation the explicit expression for L_q was inserted. The average service rate μ is a function of m and k. The function is specified as follows:

$$\mu = \frac{(m-k)\mu_1 + k\mu_2}{m} \qquad (4.4')$$

The symbols μ_1 and μ_2 refer respectively to the average service rate of one check-out stand operated by a cashier alone and the average service rate of a stand when it is run by a checker and bag boy.

2) Minimize $(m - k)W_1 + kW_2$ (4.5)

with respect to m and k, subject to

$$L_q < L^*$$

where L^* is the maximum allowable average length of queue. It is a parameter set by management.

3) Minimize $(m - K)W_1 + W_2$ (4.6)

with respect to m and k, subject to

$$\text{Prob.}\ (q > q^*) = \frac{P_0 \rho^{m+q^*+1}}{m!\ m^{q^*}(m-\rho)} < P^*$$

where

q : The number of customers in waiting line.

q^* : A fixed length of queue per service channel.

P^* : Probability level specified by management.

These last two formulations appear to be more adaptable to management's traditional or intuitive way of thinking, because the management was more willing to specify parameters such as L^*, q^*, and P^* than to specify a penalty function $\delta(L_q)$. This does not mean that solutions obtained by these two methods are always optimal in the sense that the expected cost per unit time is minimized. Although assigning a proper weight to the penalty cost function may involve a great deal of technical difficulties, management should strive to get as accurate an estimate as it can about parameters of the penalty cost function. Some suggestions as to how the parameters may be estimated are given in the last chapter.

Optimal Check-out Rules for the Super Market in Detroit

First, a check-out rule is defined as follows: for given average customer arrival rate, any choice of a combination of the controllable variables (viz., the number of checkers and the number of package boys) in order to provide the check-out service to the incoming customers is a check-out rule, and that rule which would minimize the expected cost as defined in pp. 37-40 is an optimal rule.

The three methods of solution as described in the previous section were applied to data obtained at one of the large super markets in Detroit. This store is a little above average in size and does a weekly business of better than $50,000. It is considered to be typical by the management as far as the purchasing patterns of its customers are concerned. There is no dominant nationality group among the customers who might show their preference for some particular shopping times. Variation in sales volume is closely associated with time of month or week in relation to pay days of customers. In this store, approximately one-fourth of the total weekly sales occur during the first three days of the week; Friday and Saturday together usually account for nearly one-half of the weekly sales. The store has seven regular check-out stands and one express check-out stand. The latter is responsible for only about five per cent of the total weekly sales. Like many a super market, the operation of check-out service is based on the composite judgment of the manager as to what is a uniformly good service and what is an economical check-out service.

As previously mentioned, the week was subdivided into five periods so that within each period the average customer arrival rate may be reasonably assumed to be stable and the steady state solution valid. The optimal rules were obtained separately for each of the five periods.

Those values of μ at which cost functions are to be evaluated were calculated by (4.4'). In this study, estimated values of μ_1 and μ_2

were 0.40 and 0.81 respectively, and the number of checkers m goes from 0 to 7. The next step was to find the value of ρ corresponding to each combination of m and k (the number of package boys). This was done by dividing an estimated average arrival rate for the period under study by each of those calculated values of μ. Calculated values of ρ were then used as a preliminary step in eliminating some undesirable check-out rules. The criterion adopted to eliminate undesirable check-out rules was that $m > \frac{\lambda}{\mu(m,k)}$. This is a necessary condition for the convergence of state probability distribution as shown in Appendix A, p.75. Hence, those rules which did not satisfy this criterion were eliminated from further consideration.

Minimization of the three objective functions for each period of the week is summarized in a tabular form in the remainder of this section. In the first column of the table, feasible values of ρ are given. In the next two columns, the number of cashiers and the number of bag boys are indicated. In the column 4, the operating cost of check-out facilities is given. It was calculated by the following formula:

$$(m - k)W_1 + kW_2$$

where W_1 is the wage of a cashier per minute and W_2 is the combined wage of a cashier and bag boy per minute. They were 2.81¢ and 4.37¢ respectively in this case, and were based on the hourly wage of $1.69 for a cashier and $0.90 for a bag boy.

Expected length of queue and probability of more than two persons waiting in each check-out lane are presented in the columns 5 and 6 respectively. When the a priori impression of those who are more informed about check-out operations was sought in regard to the maximum length of queue that could be tolerated, they indicated that if there would not be more than two persons waiting in each lane probably little adverse effect would result. They also felt that customers seem to consider a waiting line of, say, two persons when all the check-out lanes are open is qualitatively different from the same length of a waiting line with only one check-out lane open. Apparently the customer is less irritated by a long wait if he sees that the management is making a reasonable effort to handle the traffic by opening more counters or putting on more package boys. This is one of the factors that one has to consider in assigning a proper weight to the penalty cost function. However, in the present analysis, this information was not used because it was not available in quantitative terms.

In the 7th column, the cost function (4.4) which consist of operating cost and penalty cost is evaluated. The penalty cost was based on a linear penalty function for simplicity. Since the management was not able to suggest the coefficient d of this function, the author used what he considered to be a reasonable value for the coefficient. The coefficient d was determined as follows:

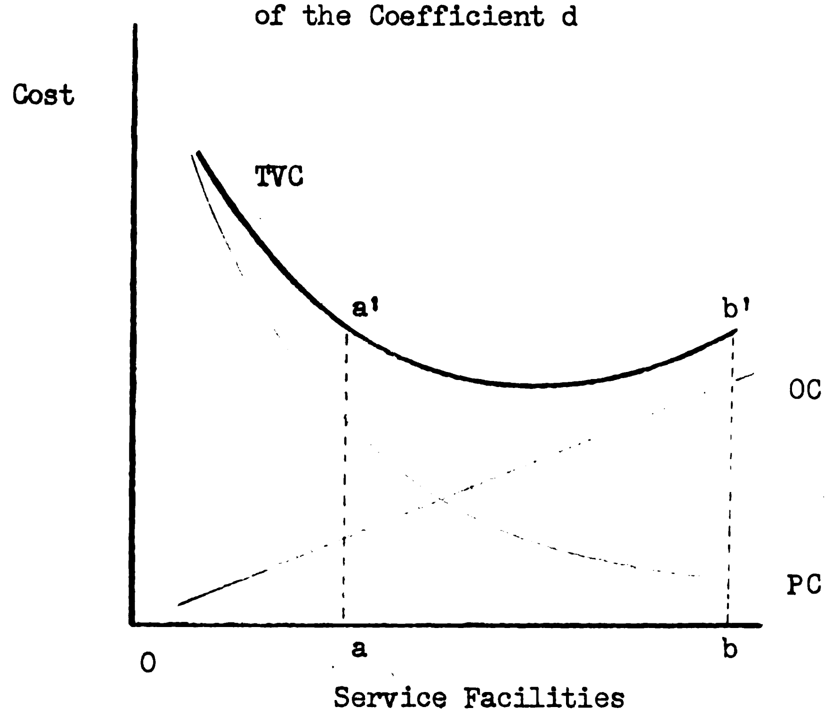
Let (m_0,k_0) and (m_1,k_1) be those combinations of checkers and package boys which would result in the highest and lowest operating costs respectively within a period under study. By evaluating the cost function (4.4) at these points, one would obtain two expressions which are functions of the coefficient d. Finally set these functions equal to each other and solve it for d.

$$(m_0-k_0)W_1 + k_0W_2 + d\ L_q(m_0,k_0) = (m_1-k_1)W_1 + k_1W_2 + d\ L_q(m_1,k_1) \tag{4.7}$$

Let the value of d which satisfies the above equality be d^*. The value of d thus obtained ensures that an optimal rule will be found somewhere between these two extremes, (m_0,k_0) and (m_1,k_1). This can be seen in the following heuristic example. Let the penalty cost, operating cost and total variable cost be labeled PC, OC, and TVC respectively in the diagram below. It is assumed that PC decreases monotonically as more service facilities are added. On the contrary, OC is assumed to be a montonically increasing function of the service facilities. Note that TVC = OC + PC. The procedure indicated in (4.7) can be shown diagramatically.

Figure 10

The Diagramatic Determination
of the Coefficient d



Select two extreme grades of check-out service. They are represented by the two points a and b along the coordinate. The points a and b may be considered to represent those check-out facilities that are called for by the combinations (m_o, k_o) and (m_1, k_1) respectively. A principle adopted here in choosing the coefficient d is to find that shape of the curve such that aa' = bb'. Now a minimum point of the TVC curve will lie within the range ab.

The actual values of d* used for calculating TVC range from about 2¢ per minute to 10¢ per minute. They do not appear to be too far off from what might be a customer's subjective judgment of how much his time is worth.

If the coefficient d was assigned a value greater than d*, it would imply that more weight is given to the penalty cost, (in fact, if it were excessively large, it would be meaningless to apply the theory of queues to this problem). On the contrary, if the coefficient takes a value smaller than d*, then operating cost becomes more important in consideration of minimizing the total variable cost. It is hoped that more research will be undertaken in the future so that the information pertaining to the parameter d will be more readily available. No entry in the 7th column of the table should be interpreted that the total variable cost for that particular check-out rule is higher than that corresponding to the optimal rule.

In the last column, optimum check-out rules according to the three criteria are indicated.

Before presenting the summary tables, the abbreviations that appear in the tables are explained again for the sake of clarity.

$\hat{\rho}$: $\hat{\lambda}/\hat{\mu}$, i.e., the ratio of the estimated average arrival rate to estimated average service rate.

m : The number of cashiers.

k : The number of bag boys.

C : The total operating cost, cents per minute.

L_q : The expected length of queue.

Prob. $(q > 2m)$: The probability of more than two customers waiting in each check-out lane.

T.V.C. : The total variable cost evaluated by means of (4.4), cents per minute.

(1) : The optimum check-out rule obtained by the first method, i.e.,

$$\min_{m,k} \quad C + d\,L_q = \text{T.V.C.}$$

(2) : The optimum rule obtained by the second method, i.e.,

$$\min_{m,k} \quad C \text{ subject to } L_q < 2m$$

(3) : The optimum rule obtained by the third method, i.e.,

$$\min_{m,k} \quad C \text{ subject to Prob. } (q > 2m) < 0.05.$$

TABLE IV

OPTIMUM CHECK-OUT RULES FOR MONDAY, TUESDAY, AND WEDNESDAY
(λ = 0.91)

$\hat{\rho}$	m	k	C	L_q	Prob(q > 2m)	T.V.C. (¢/min.)	Optimum Rule
2.25	3	0	8.43	1.699	7.80%	10.51	
1.50	2	1	7.18	1.929	15.22	9.55	(2)
2.25	4	0	11.24	0.310	-	11.62	
1.69	3	1	9.99	0.399	-	10.48	
1.13	2	2	8.74	0.530	2.38	9.39	(1),(3)
2.25	5	0	14.05	0.074	-	14.14	
1.80	4	1	12.80	0.105	-	12.93	
1.35	3	2	11.55	0.152	-	11.73	
2.25	6	0	16.86	0.018	-	16.88	
1.87	5	1	15.61	0.028	-	15.64	
1.50	4	2	14.36	0.045	-	14.42	
1.13	3	3	13.11	0.074	-	13.20	
2.25	7	0	19.67	0.004	-	19.67	
1.93	6	1	18.42	0.007	-	18.43	
1.61	5	2	17.17	0.016	-	17.19	
1.28	4	3	15.92	0.021	-	15.95	
1.97	7	1	21.23	0.002	-	21.23	
1.69	6	2	19.98	0.003	-	19.98	
1.41	5	3	18.73	0.006	-	18.73	
1.13	4	4	17.48	0.012	-	17.49	
1.75	7	2	22.79	*	-		
1.50	6	3	21.54	0.002	-		
1.24	5	4	20.29	0.003	-		
1.57	7	3	24.35	*	-		
1.35	6	4	23.10	0.001	-		
1.13	5	5	21.85	0.002	-		
1.43	7	4	25.91	*	-		
1.23	6	5	24.66	*	-		
1.31	7	5	27.47	*	-		
1.13	6	6	26.22	*	-		
1.21	7	6	29.03	*	-		
1.13	7	7	30.59	*	-		

* : L_q is less than 0.001.

- : Prob(q > 2m) is less than 1%.

TABLE V

OPTIMUM CHECK-OUT RULES FOR THURSDAY
($\hat{\lambda}$ = 1.53)

$\hat{\rho}$	m	k	C	L_q	Prob(q > 2m)	T.V.C. (¢/min.)	Optimum Rule
3.78	4	0	11.24	15.134	52.94	32.43	
2.84	3	1	9.99	16.009	61.37	32.40	
1.89	2	2	8.74	16.790	66.38	32.24	
3.78	5	0	14.05	1.462	2.17	16.10	
3.02	4	1	12.80	1.596	4.14	15.03	(3)
2.27	3	2	11.55	1.799	8.21	14.07	(2)
3.78	6	0	16.86	0.399	-	17.42	
3.15	5	1	15.61	0.469	-	16.23	
2.52	4	2	14.36	0.556	-	15.14	
1.89	3	3	13.11	0.671	1.53	14.05	(1)
3.78	7	0	19.67	0.125	-	19.84	
3.24	6	1	18.42	0.156	-	18.64	
2.70	5	2	17.17	0.198	-	17.45	
2.16	4	3	15.92	0.253	-	16.27	
3.31	7	1	21.23	0.053	-	21.30	
2.84	6	2	19.98	0.072	-	20.08	
2.36	5	3	18.73	0.096	-	18.86	
1.89	4	4	17.48	0.133	-	17.66	
2.74	7	2	22.79	0.025	-	22.82	
2.52	6	3	21.54	0.035	-	21.58	
2.08	5	4	20.29	0.049	-	20.36	
2.65	7	3	24.35	0.013	-		
2.27	6	4	23.10	0.019	-		
1.89	5	5	21.85	0.030	-		
2.41	7	4	25.91	0.007	-		
2.06	6	5	24.66	0.011	-		
2.21	7	5	27.47	0.004	-		
1.89	6	6	26.22	0.006	-		
2.04	7	6	29.03	0.002	-		
1.89	7	7	30.59	0.001	-		

- : Prob (q > 2m) < 0.01.

TABLE VI

OPTIMUM CHECK-OUT RULES FOR FRIDAY BEFORE 6 P.M.
(λ = 1.85)

$\hat{\rho}$	m	k	C	L_q	Prob(q > 2m)	T.V.C. (¢/min.)	Optimum Rule
4.57	5	0	14.05	8.417	29.52	32.57	
3.65	4	1	12.80	8.461	35.82	31.41	
2.74	3	2	11.55	8.880	44.45	31.09	(2)
4.57	6	0	16.86	1.425	1.26	20.00	
3.81	5	1	15.61	1.526	2.94	18.97	
3.04	4	2	14.36	1.667	4.13	18.03	(3)
2.28	3	3	13.11	1.848	8.51	17.17	(1)
4.57	7	0	19.67	0.434	-	20.63	
3.91	6	1	18.42	0.493	-	19.50	
3.26	5	2	17.17	0.572	-	18.43	
2.61	4	3	15.92	0.672	1.06	17.40	
4.00	7	1	21.23	0.180	-		
3.43	6	2	19.98	0.220	-		
2.85	5	3	18.73	0.266	-		
2.28	4	4	17.48	0.332	-		
3.55	7	2	22.79	0.084	-		
3.04	6	3	21.54	0.107	-		
2.51	5	4	20.29	0.133	-		
3.20	7	3	24.35	0.043	-		
2.74	6	4	23.10	0.058	-		
2.28	5	5	21.85	0.080	-		
2.91	7	4	25.91	0.023	-		
2.49	6	5	24.66	0.034	-		
2.66	7	5	27.47	0.013	-		
2.28	6	6	26.22	0.020	-		
2.46	7	6	29.03	0.008	-		
2.28	7	7	30.59	0.005	-		

- : Prob (q > 2m) is less than 1%.

TABLE VII

OPTIMUM CHECK-OUT RULES FOR FRIDAY AFTER 6 P.M.
($\hat{\lambda}$ = 2.55)

$\hat{\rho}$	m	k	C	L_q	Prof(q > 2m)	T.V.C. (¢/min.)	Optimum Rule
6.32	7	0	19.67	6.880	14.18	33.49	
5.42	6	1	18.42	7.148	18.17	32.79	
4.51	5	2	17.17	7.210	27.86	31.66	
3.61	4	3	15.92	7.323	31.64	30.63	(2)
5.53	7	1	21.23	1.743	1.78	24.73	
4.74	6	2	19.98	1.862	2.83	23.72	
3.95	5	3	18.73	2.010	4.16	22.77	(3)
3.16	4	4	17.48	2.173	7.07	21.84	(1)
4.92	7	2	22.79	0.727	-	24.25	
4.21	6	3	21.54	0.799	-	23.15	
3.48	5	4	20.29	0.852	-	22.00	
4.42	7	3	24.35	0.347	-	25.04	
3.79	6	4	23.10	0.404	-	23.91	
3.16	5	5	21.85	0.477	-	22.80	
4.02	7	4	25.91	0.185	-		
3.45	6	5	24.66	0.228	-		
3.69	7	5	27.47	0.107	-		
3.16	6	6	26.22	0.135	-		
3.40	7	6	29.03	0.063	-		
3.16	7	7	30.59	0.040	-		

- : Prob(q > 2m) is less than 1%.

TABLE VIII

OPTIMUM CHECK-OUT RULES FOR SATURDAY
($\hat{\lambda}$ = 2.44)

$\hat{\rho}$	m	k	C	L_q	Prof(q > 2m)	T.V.C. (¢/min.)	Optimum Rule
6.03	7	0	19.67	3.870	6.91	32.44	
5.17	6	1	18.42	4.060	10.89	31.82	
4.31	5	2	17.17	4.263	13.30	31.24	
3.46	4	3	15.92	4.416	18.67	30.49	(2)
5.28	7	1	21.23	1.219	-	25.35	
4.52	6	2	19.98	1.305	1.16	24.29	
3.77	5	3	18.73	1.439	1.97	23.49	
3.01	4	4	17.48	1.562	4.16	22.63	(1),(3)
4.69	7	2	22.79	0.517	-	24.50	
4.02	6	3	21.54	0.587	-	23.48	
3.31	5	4	20.29	0.625	-	22.35	
4.22	7	3	24.35	0.254	-	25.19	
3.62	6	4	23.10	0.304	-	24.10	
3.01	5	5	21.85	0.360	-	23.03	
3.84	7	4	25.91	0.138	-		
3.29	6	5	24.66	0.171	-		
3.52	7	5	27.47	0.078	-		
3.01	6	6	26.22	0.101	-		
3.25	7	6	29.03	0.047	-		
3.01	7	7	30.59	0.028	-		

- : Prob(q > 2m) is less than 1%.

From the foregoing tables, the optimum check-out rules for each period of the week can be determined. However, even if the management is unable to specify parameters necessary for optimization, these tables would still provide useful information because they have narrowed down a range of selection for optimal rules.

As mentioned previously, rules considered here are limited to those which satisfy one of the conditions for the convergence of the stationary probability distribution. This condition states that $\rho > m$. Hence, the rules in the table may be considered to have passed the initial screening.

In most cases, a check-out rule with a higher operating cost would tend to be associated with a better grade of service as indicated by a shorter expected length of queue and smaller Prob.$(q > 2m)$. However, there are some exceptions to this tendency. One may note that a number of check-out rules in the foregoing tables illustrates this point. For instance, on Saturday, the check-out rule $(m = 5, k = 3)$ is to be preferred to the rule $(m = 7, k = o)$, because the application of the former would bring about the better service with smaller expenses as compared to the latter. Hence a rule such as $(m = 7, k = 0)$ has to be excluded from further consideration. The job can be always done more efficiently and economically by adopting the rule $(m = 5, k = 3)$ in place of the rule $(m = 7, k = 0)$.

It may be worth noting that for given average arrival rate and cost function, the optimal rule (m,k) is not highly sensitive to the choice of the criterion of optimality (at least not for the three criteria used here). This can be seen in the following table.

TABLE IX

OPTIMUM RULES FOR DIFFERENT AVERAGE ARRIVAL RATE AND DIFFERENT CRITERIA OF OPTIMALITY

Criterion \ λ	0.91		1.53		1.85		2.44		2.55	
	m	k	m	k	m	k	m	k	m	k
(1)	2	2	3	3	3	3	4	4	4	4
(2)	2	1	3	2	3	2	4	3	4	3
(3)	2	2	4	1	4	2	4	4	5	3

Before concluding the discussion on optimum check-out rules, a question may be raised as to what effect the sampling fluctuations of estimates of average arrival rate and average service rate would have on determination of optimal check-out rules. For this purpose, some results obtained in the second half of Chapter III will be used.

The first item to be considered is the variation in $\hat{\rho}$ as estimates of λ and μ fluctuate. This is done by obtaining upper and lower limits of $\hat{\rho}$ from confidence limits of $\hat{\lambda}$ and $\hat{\mu}$.

Let λ^*, and λ_* be the upper and lower confidence limits, respectively, of $\hat{\lambda}$ for a given size of confidence coefficient; similarly μ^* and μ_* are the upper and lower limits of $\hat{\mu}$.[3] $\hat{\rho}$ would then vary between its upper limit $\frac{\lambda^*}{\mu_*}$ and lower limit $\frac{\lambda_*}{\mu^*}$. The effect of variations in $\hat{\lambda}$ and $\hat{\mu}$ on optimum check-rule can be studied by obtaining the optimum rules corresponding to those limits.

[3]Since μ is calculated by the formula $\mu = \frac{(m-k)\mu_1 + k\mu_2}{m}$, (4.4') μ^* is obtained by inserting upper confidence limits of μ_1 and μ_2 into (4.4') and μ_* is obtained by inserting lower confidence limits of μ_1 and μ_2 into the equation.

In Table X, optimum rules for Saturday based on the upper and lower limits of $\hat{\rho}$, which in turn were obtained from the 90 percent confidence limits of $\hat{\lambda}$ and $\hat{\mu}$, as presented in Table III, page 35, are compared to those given in Table VIII of this chapter.

TABLE X

EFFECT OF VARIATION IN ESTIMATES OF λ AND μ ON OPTIMUM CHECK-OUT RULES

	Method (1)		Method (2)		Method (3)	
	m	k	m	k	m	k
Upper limit of $\hat{\rho}$	5	5	4	4	5	4
$\hat{\rho}$	4	4	4	3	4	4
Lower limit of $\hat{\rho}$	4	4	3	3	4	3

This example suggests that selection of optimal rules by the methods proposed in this study is fortunately not very sensitive to variation in estimates of λ and μ due to sampling error.

CHAPTER V

SUMMARY OF RESULTS AND SUGGESTIONS FOR FURTHER STUDY

The applications of queuing theory have been many and varied. They range from the design of airports to the scheduling of patients in clinics.

Despite the fact that the check-out service at super markets appears to have all the necessary features that make up a typical queuing problem, as far as the author is aware, no attempt has been made to answer some important questions in connection with the efficient operation of check-out service more rigorously. For example, how many check-out stands should there be in operation to handle the customers ready for check-out at any given time? Or, given that six employees are available, is it better to have six check-out stands operating or only three check-out stands with one package boy each?

This thesis is chiefly concerned with procedures that can be used to answer these questions in quantitative terms and in a logical manner. The procedures proposed here were applied to data obtained at one large super market in the Detroit area.

An application of the procedures essentially involves a balancing of the cost incurred by providing a certain amount of check-out facilities (the cost mainly comprises the wages of checkers and package boys) for a given time period against the cost of losing customers in the future because of inferior service standards. The first type of cost is essentially a function of the number of checkers and the number of package boys attending check-out stands during that time period. The second type of cost depends on the following three factors: 1) the number of checkers and the number of package boys, 2) the number of customers arriving at the check-out area during the time period, and 3) the rate at which each arriving customer is served. The management can not regulate the flow of customers to suit its available labor and facilities; however, it can decide on the amount of labor and facility requirements to meet a given flow of customers. Hence, if some functional relationships between the sum of these two costs and the three factors mentioned above can be specified, it will then be possible to determine the cost for given values of the factors. It was shown in this thesis that such functional relationships could be specified in a logical manner.

The last two factors which comprise a main portion of the second type of cost are not known in advance. In order to get around this difficulty the probability distributions of these two quantities are estimated. From these stochastic quantities, an expected cost can be calculated by means of the theory of queues. The specific procedures followed in this thesis are briefly described in the following.

First, a queuing system which is inherent in any check-out operation at super markets was characterized in the following manner:

Input: time intervals between two consecutive arrivals are independently and identically distributed by a negative exponential distribution.

Service mechanism: number of check-out counters available is finite; time required to serve one customer at a counter has one of two different probability distributions depending on whether a package boy is at the counter or not; both distributions are assumed to be negative exponential distributions.

Queue discipline: "first come, first served."

The system can be in any of possible "states," specified by the number of customers in the queue, the number of customers being served, or the total number of customers in the system. From the probabilistic characterizations of input and service mechanisms, the probabilities that the system is in each of the possible states independently of time were calculated. These probabilities are called steady state probabilities.

Next the costs associated with the system in each state were estimated. From the steady state probabilities and their associated costs, the expected costs could be readily computed.

The steady state probabilities are valid only when the average arrival rate and the average service rate can reasonably be considered stable. It is not unrealistic to assume that the latter will remain constant so long as there is no basic change in the service mechanism. On the other hand, the average arrival rate varies from day to day, if not from hour to hour. It is known to be closely related to sales per unit of time. In order to apply the procedures developed in this study to the check-out operation at a super market in Detroit, sales records of the store were examined; and the week was divided into five periods so that within each period the assumption of stable average arrival rate would be more tenable.

Optimum check-out rules for each of the five periods were obtained separately and were shown in tabular forms. From the table, the effect of alternative optimal criteria on the rules can be readily seen.

It should be noted that the thesis is chiefly concerned with derivation of procedures in obtaining certain check-out rules which satisfy given optimum criteria. As such, institutional and managerial arrangements necessary in adopting the optimal check-out rules to the actual check-out operation were considered outside the scope of this thesis.

Prior to calculating expected costs of the check-out operation, some statistical analyses were made of customer arrivals and service times in order to determine how well the assumptions of Poisson arrival and exponential service time will approximate the actual situation. The chi-square test of goodness-of-fit was used. The test results indicated that in general the number of arriving customers per minute follows closely a Poisson distribution. The negative exponential distribution, however, did not seem to give the best fit to observations on service time. Since the empirically derived distribution was skewed to the left, another member of the gamma function family would probably have given a better fit. Nevertheless, the steady state probabilities were calculated, based on this exponential service time hypothesis, because the hypothesis renders the mathematics manageable.

In the theoretical study of a queuing problem, the average arrival rate and the average service rate are treated as though they were known *a priori*. In an applied problem, they have to be estimated by some statistical procedure; hence, they are subject to sampling fluctuations.

Since a choice of an optimum check-out rule depends on the estimates of these quantities, it is natural to examine the sensitivity of the method of choosing optimal check-out rules to changes in the estimated average arrival rates and the estimated average service rate. This was done by first obtaining the sampling distributions of these estimates and then calculating their confidence limits. From these limits, a range of variation in check-out rules was examined. This range was found to be very small when the procedures were applied to the data obtained at the store in Detroit.

Although use of such a queuing model enables a research worker to analyze a given problem without going to the expense of actually duplicating the situation, there are several points that ought to be further examined for the model to be more useful.

The calculated state probability distribution p_n, for $n = 0, 1, \ldots$, can be checked for its accuracy by observing the frequency of the formation of queues of all possible lengths. One could operate the check-out stands for a period with fixed number of checkers and package boys, and observe the number of customers waiting for service at the end of every interval, say one minute, during the period. Experiment such as this is conceptually possible; however, the resource available to the author did not permit carrying out the experiment at this time.

Little is known about the coefficient of the penalty function. Probably some ideas can be obtained if one were to ask the customer in the waiting line how much he considers his time is worth. Accurate information about this coefficient is much needed in obtaining the valid solution of a queuing problem such as this. It might be desirable to consider enlisting the help of a psychologist in designing an experiment for such a purpose.

The hypothesis of exponential service time was used throughout this study, because of its mathematical simplicity. An alternative hypothesis is that of constant service time. A queuing model based on this hypothesis is briefly described below. The hypotheses with respect to input and queue discipline are the same as before.

Let

m : the number of check-out stands to be put in operation.

k : the number of package boys.

λ : the average number of customers arriving per fixed time interval. In particular, the time interval is set equal to the service time. It should be noted that when λ is measured in this manner, it will depend on the controllable variables m and k because they affect the length of average service time.

Each of the following set of equations relates the state probability at the beginning and end of the interval which is equal to service time:

$$P(n) = P(n+m)e^{-\lambda} + P(n+m-1)\lambda e^{-\lambda} + \ldots + P(m+1)\frac{\lambda^{n-1}}{(n-1)!}e^{-\lambda} + \left[\sum_{i=o}^{m} P(i)\frac{\lambda^{n}}{n!}\right]e^{-\lambda} \qquad \text{for } n = o, 1, \ldots \qquad (5.1)$$

Equation (5.1) states that the probability that there are n customers in the system at the end of the fixed interval is equal to the sum of the probabilities of the following n+1 contingencies: 1) there are n+m customers at the beginning of the interval and no customer comes during the interval...........n) there are m+1 customers at the beginning of the interval and exactly n-1 customers arrive during the interval, n+1) there are less than $m+1$ people at the beginning of the interval but n people arrive during the interval, so that at the end there are still n customers waiting.

It is not easy to solve this infinite set of equations.[1] If the solution is obtainable at a reasonable cost, the state probability distribution may be calculated for each of the feasible average service times. The expected cost corresponding to each service time can also be calculated as in the model with exponential service time.

If an assumption in regard to the service mechanism is such that an analytical approach to the queuing problem becomes so complex that it is impossible to obtain a solution, a simulated sampling approach may be adopted. Briefly, the procedure to be followed is this. Observations on customer arrivals and service times are examined and parameters of the theoretical distributions which are most likely to give the best fit to the data are estimated. By means of a table of random numbers, one can construct a scheme such that the drawing of a random number will be from the population which has a known probability distribution. This is done for both customer arrivals and service times, using their respective estimated probability distributions. For instance, a number is drawn at random from the table of random numbers at the end of every fixed interval and if the number drawn falls between 0 and 9000 it is interpreted that no customer has arrived during that period; if it is between 9001 and 9999, then one customer

---

[1]Everett (1953) suggested an iterative procedure to obtain an approximate solution.

has arrived. In order to generate a sequence of random service times from a known distribution, all the four-digit numbers may be grouped according to the estimated service time distribution, and a certain service time is assigned to each group of numbers. Repeated drawings of numbers from the table will then give a sequence of service times with the desired probability distribution. Next, the first random service time is assigned to the first customer; the second random service time to the second customer and so on. In this way, the congestion situation at the check-out area can be simulated.

If such an experiment is carried out on a high-speed computer, enough data can be obtained to construct a frequency distribution of the system in each of all the possible states in a very short period of time. The expected cost can be calculated based on this frequency distribution. The process is repeated for different values of the parameters of the arrival distribution and service time distribution until that expected cost which meets a given optimal criterion is found.

The above procedure is for the case of one service channel. If the problem involves more than one service channel and if each of them has a different probability distribution, then as many sets of service times as the number of service channels have to be drawn. A realistic rule has to be set up to determine a manner in which a random service time is to be assigned to each of the incoming customers, and the congestion situation can be simulated as before.

In the light of recent developments of high-speed computers, there is much to be recommended for adopting the simulated sampling approach to a queuing problem. The advantage of this approach is that there are more flexibilities in regard to the assumptions about input and service mechanism. It is quite likely that the future study of queuing problems by the Monte Carlo method will yield many fruitful results.

LITERATURE CITED

Arrow, K., S. Karlin and H. Scarf, (1958), Studies in the Mathematical Theory of Inventory and Production, Stanford University Press, Stanford.

Brockmeyer, E., H. Halstrom and A. Jensen, (1948), The Life and Works of A. K. Erlang, Copenhagan.

Cox, C. B., A. Glockstein and J. H. Greene, (1958), "Application of Queuing Theory in Determining Livestock Unloading Facilities," Journal of Farm Economics, XL: 104-116.

Edie, L. C. (1955), "Traffic Delays at Toll Booths," Journal of Operations Research Society of America, Vol. 2.

Everett, J. (1953), "State Probabilities in Congestion Problems Characterized by Constant Holding Times," Operations Research Society of America, November 1953, 1, 259-62.

Feller, W., (1950), An Introduction to Probability Theory and its Applications, John Wiley and Sons, Inc., New York

Flagle, C. D., (1956), "Queuing Theory and Cost Concepts Applied to a Problem in Inventory Control," Operation Research for Management, Vol. II, Baltimore, Johns Hopkins Press, 160-177.

Fry, T. C. (1928), Probability and Its Engineering Use, New York: van Nostrand.

Holley, Julian L. (1954), "Waiting Line Subject to Priorities," Journal of Operations Research Society of America, 2, 341-343.

Kendall, D. G. (1953), "Stochastic Processes Occurring in the Theory of Queues and Their Analysis by the Method of the Imbedded Markov Chair," Annals of Mathematical Statistics, 24: 338-354.

Kendall, M. G., (1951), "Some Problems in the Theory of Queues," Journal of the Royal Statistical Society, series B. Vol. XIII, 151-171.

Lindley, D. V., (1952), "The Theory of Queues with a Single Server," Proceedings of the Cambridge Philosophical Society, 48: 277-289.

Marshall, B. O. (1954), "Queuing Theory," Operations Research for Management, Vol. 1, Baltimore, Johns Hopkins Press, 134 -148.

McClosky and Coppinger, (1956), Operations Research for Management, Vol. II, Johns Hopkins Press.

Molina, E. C., (1927), "Application of the Theory of Probability to Telephone Trunking Business," Bell System Technical Journal, Vol. 6, 461-494.

Molina, E. C., Poisson's Exponential Binomial Limit, Table I. New York: van Nostrand Company.

Salvosa, L. R., (1930), Tables of Pearson's Type III Functions, Edwards Brothers, Inc., Ann Arbor.

Satty, T. L., (1957), "Resume of Useful Formulas in Queuing Theory," Operation Research Society of America, Vol. 5, No. 2.

APPENDICES

APPENDIX A

STATE PROBABILITY DISTRIBUTION

The techniques of deriving state probability distributions have been presented by a number of authors[1] in a highly technical and often abbreviated form.

In this appendix, the assumptions and definitions upon which the derivation of state probability distributions will proceed are set forth in detail and the derivation is presented step by step in order to facilitate an understanding of the logic behind these techniques.

First of all, hypotheses about the service mechanism and input process have to be specified.

Service time distribution:

Let the service time be defined as the time which elapses while a particular customer is being served. The hypothesis about the service time is as follows:

$$\text{Prob.}\left[\begin{array}{l}\text{A check-out service}\\ \text{extends beyond time } t\end{array}\middle|\begin{array}{l}\text{It has started}\\ \text{at time } o\end{array}\right] = e^{-\mu t} \quad (o < t < \infty) \tag{A.1}$$

For an interpretation of the parameter μ, we take the expected value of t. Since the equation (A.1) defines the cumulative distribution function (c.d.f.) of service time in terms of its upper tail, we have to first obtain the c.d.f. in its usual form in order to calculate mean of the random variate t.

---

[1]See Feller (1950), Kendall (1951) and Lindley (1952).

$F(T \leq t) = 1 - e^{-\mu t}$ the c.d.f. of t in the usual form.

$\frac{dF}{dt} = \mu e^{-\mu t}$ the density function

Hence $E(t) = \int_0^\infty t\mu e^{-\mu t}\, dt = \frac{1}{\mu}$ (A.2)

Since E(t) is the mean service time, μ can be considered as the average number of customers served per unit time. It is called the mean service rate.

From (A.1), the conditional probability that a check-out service is completed between t and $t + \Delta t$ given that the service was being rendered at time t can be calculated.

$$\text{Prob.}\left[\begin{array}{l}\text{Check-out service ends}\\ \text{between } t \text{ and } t + \Delta t\end{array}\middle|\begin{array}{l}\text{It did not end}\\ \text{before time } t\end{array}\right]$$

$$= \frac{\text{Prob.}\left[\begin{array}{l}\text{Check-out service does not terminate before } t,\\ \text{but ends between } t \text{ and } t + \Delta t\end{array}\right]}{\text{Prob.}\left[\text{Check-out service did not end before } t\right]}$$

$$= \frac{e^{-\mu t} - e^{-\mu(t + \Delta t)}}{e^{-\mu t}}$$

$$= 1 - e^{-\mu \Delta t}$$

$$= 1 - \left[1 - u\,\Delta t + \frac{(\mu\Delta t)^2}{2!} - \frac{(\mu\Delta t)^3}{3!} + \dots\right]$$

$$= \mu\Delta t + o(\Delta t) \qquad (A.3)$$

The equation (A.3) is interpreted as follows: If at time t a check-out stand is occupied, then the probability that the stand would terminate its operation during $(t, t + \Delta t)$ is $\mu\Delta t$ plus terms which approach zero faster than Δt.

Arrival Distribution:

The hypothesis adopted throughout this analysis is that the customer arrivals consist of a Poisson process, i.e., the number of arrivals in time t being a Poisson variable with mean λt, say. This implies that the time interval between two consecutive arrivals has the negative exponential distribution.[2] This time interval is often referred to as customer "idle" time.

$$\text{Prob.} \begin{bmatrix} \text{Every customer is idle at time o} \\ \text{and still is idle at time t} \end{bmatrix}$$

$$= \text{Prob.} [\text{No customer coming in during that period}]$$

$$= e^{-\lambda t} \qquad \text{(A.4)}$$

Using (A.4), we can calculate the conditional probability of exactly one arrival between $(t, t + \Delta t)$ given that there is no arrival during (o,t).

$$\text{Prob.} \left[\begin{matrix} \text{Exactly one arrival} \\ \text{during } (t, t + \Delta t) \end{matrix} \middle| \begin{matrix} \text{No arrival} \\ \text{between } (o,t) \end{matrix} \right]$$

$$= 1 - e^{-\lambda \Delta t}$$

$$= \lambda \Delta t + o(\Delta t) \qquad \text{(A.5)}$$

λ is the mean arrival rate.

After the incoming traffic and the service mechanism have been characterized in terms of probability as above (Poisson arrival and exponential service time), the next thing to specify is the queue discipline. It is assumed that a check-out system follows the so-called

[2]A proof of this statement can be found in Feller (1950), pp. 364-367.

strict queue discipline. Since check-out stands, in the current study, are classified into two groups depending on whether a package boy is assisting a checker or not, it is necessary to adopt a rule in order to decide how each arriving customer moves into a check-out lane. Let μ_1 and μ_2 denote respectively the average service rate of one check-out stand that is operated by a checker alone and by a package boy and a checker together. Suppose there are m-k check-out stands with the average service rate μ_1 and k check-out stands with the average service rate μ_2. The rule is that $\frac{m-k}{m}$ fraction of the incoming customers will go through the counters with μ_1 average service rate and the remaining fraction $(\frac{k}{m})$ will be served by the counters with μ_2 average service rate. In other words, the assignment of each customer to a service channel is decided by tossing a coin with a ratio $\frac{m-k}{m} : \frac{k}{m}$.

This completes the specification of the system. With these assumptions, one can proceed to write equations which represent the detailed balancing of transitions between states for a stationary steady state.

Two contingencies have to be recognized. In the first situation, the number of customers (n) present in the system is less than the number of check-out stands (m) which are in operation. The second situation is that the former is at least as great as the latter.

Adopting the terminology of the theory of stochastic process, we shall say that the system is in state E_n at time t when there are n customers in the check-out system at that time.

The system will be in state E_n at time $t + \Delta t$ only under the following conditions:

1. The system is in E_n at time t and during $(t, t + \Delta t)$ no customer arrives or departs.

2. The system is in E_{n-1} at time t and exactly one customer comes in during $(t, t + \Delta t)$.

3. The system is in E_{n+1} at time t, and exactly one customer leaves the system.

4. During $(t, t + \Delta t)$, two or more customers arrive or depart.

It is assumed that the probability of the last contingency becomes insignificant as $\Delta t \longrightarrow 0$. Hence, it was not considered in calculation of the state probabilities. Since the first three situations are mutually exclusive, the probabilities of these contingencies are additive.

Let $P_n(t)$ denote the probability that the system is in state E_n at time t. First consider the situation $n \leq m$. There the probability of exactly n customers in the system at time $t + \Delta t$ is as follows:

$$P_n(t+\Delta t) = [1 - \lambda \Delta t - (\frac{n(m-k)}{m} \mu_1 + \frac{nk}{m} \mu_2) \Delta t] P_n(t) + \lambda \Delta t P_{n-1}(t)$$

$$+ [(\frac{(n+1)(m-k)}{m} \mu_1 + \frac{(n+1)k}{m} \mu_2) \Delta t] P_{n+1}(t) + o(\Delta t). \qquad (A.6)$$

From left to right, each term on the right hand side of (A.6) corresponds to the probability of a contingency in the order listed in the previous page. When the term $P_n(t)$ is transposed to the left hand side of (A.6) and the resulting expression is divided by Δt, we have the derivative of $P_n(t)$ with respect to t by definition upon taking a limit of the ratio as $\Delta t \longrightarrow 0$.

To simplify notations, let

$$\mu = \frac{(m-k)\mu_1 + k\mu_2}{m} \tag{A.6}$$

Then this derivative can be written as follows:

$$\frac{dP_n(t)}{dt} = -(\lambda + n\mu)\, P_n(t) + \lambda P_{n-1}(t) + (n+1)\mu P_{n+1}(t) \tag{A.7}$$

$$\text{for } n \leqq m$$

This is the same set of differential equations for the case where all the service channels have equivalent service capacities.

Since the state probability $P_{n-1}(t)$ is undefined for $n-1 < 0$, the differential equation (A.7) is valid only for $m \geqq n > 1$. For $n = o$, an equation similar to (A.6) has to be formulated.

$$P_0(t + \Delta t) = (1 - \lambda \Delta t)\, P_0(t) + \mu \Delta t P_1(t) + o\,(\Delta t) \tag{A.8}$$

The above relation says that the probability of no customer in the system at time $t + \Delta t$ is the sum of two independent probabilities: 1) probability of no customer in the system at time t and no arrival during $(t, t + \Delta t)$, and 2) probability of exactly one customer in the system at time t and he will leave the system during $(t, t + \Delta t)$. From this equation, the derivative $\frac{dP_0(t)}{dt}$ is obtained in a similar manner as in (A.7).

$$\frac{dP_0(t)}{dt} = \lambda P_0(t) + \mu P_1(t) \tag{A.9}$$

Theoretically transient solutions, i.e., the solutions which depend on the variate t, can be obtained by first deriving a partial differential equation for the generating function of state probability distribution.[3] The mathematics, however, becomes rather involved. In the current study it is assumed that a statistical equilibrium exists i.e., $\lim_{t \to \infty} P_n(t) = P_n$[4] exists, and only the so-called steady state solution is obtained from (A.7) and (A.9). This can be done by setting $\frac{dP_n(t)}{dt} = o$ and $\frac{dP_o(t)}{dt} = o$ and solving the following set of simultaneous equations by recursion for all relevant P_n's in terms of P_o:

$$\lambda P_o = \mu P_1$$

$$(\lambda + n\mu)P_n = \lambda P_n + (n+1)\,\mu P_{n+1} \qquad 0 < n \leq m \qquad (A.10)$$

The result of solving these relations relating the state probabilities is

$$P_n = \frac{P_o \rho^n}{n!}\,, \quad \text{where} \quad \rho = \frac{\lambda}{\mu} \qquad (A.11)$$

The second contingency that needs to be considered is the situation in which there are more customers than the check-out stands in operation.

The basic system of differential equations for $n \geq m$ is derived from

$$P_n(t + \Delta t) = (1 - \lambda\Delta t - m\mu\Delta t)P_n(t) + \lambda\Delta t\, P_{n-1}(t) +$$
$$m\,\mu\Delta t\, P_{n+1}(t) + o\,(\Delta t) \qquad (A.12)$$

The interpretation of the above equation is similar to that of (A.6) and (A.8). The differential equations are as follows:

---

[3]The partial differential equation for the generating function is given in Feller (1950), p. 396.

[4]Supression of the variable t indicates that the state probability P_n is independent of time t.

$$\frac{dP_n(t)}{dt} = -(\lambda + m\mu)\, P_n(t) + \lambda P_{n-1}(t) + m\mu P_{n+1}(t) \qquad (A.13)$$

Again a steady state solution is obtained by setting $\frac{dP_n(t)}{dt} = o.$ The following set of simultaneous equations are then solved by recursion.

$$(\lambda + m\mu)\, P_n = \lambda P_{n-1} + m\mu P_{n+1} \qquad (A.14)$$

The solution is

$$P_n = \frac{P_o \rho^n}{m!\, m^{n-m}}, \text{ for } n \geq m > o. \qquad (A.15)$$

From (A.11) and (A.15), the probability that the system is in any state except E_o can be calculated in terms of P_o. In order to determine P_o, we make use of the condition $\sum_{n=o}^{\infty} P_n = 1$

By means of (A.11) and (A.15), this condition can be more explicitly written as follows:

$$1 = P_o \left[\sum_{n=0}^{m-1} \frac{\rho^n}{n!} + \frac{m^m}{m!} \sum_{n=m}^{\infty} \frac{\rho^n}{m^n} \right]$$

then (A.16)

$$P_o^{-1} = \sum_{n=0}^{m-2} \frac{\rho^n}{n!} + \frac{\rho^{m-1}}{(m-1)!} \frac{m}{m-\rho}$$

(A.11), (A.15) and (A.16) together completely specify the distribution of a statistically steady state. It should be noted from (A.15) that the series $\sum_{n=m}^{\infty} \frac{P_n}{P_o}$ coverges only when $\rho < m$. This can be seen from the following argument: Divide both sides of (A.15) by P_o and sum it over the index n from m to ∞.

$$\sum_{n=m}^{\infty} \frac{P_n}{P_o} = \frac{m^m}{m!} \sum_{n=m}^{\infty} \left(\frac{\rho}{m}\right)^m \qquad (A.17)$$

The right hand side of (A.17) is a geometric series, and it converges to a limit $\frac{\rho^m}{(m-1)!(m-\rho)}$ only when $\rho < m$. This convergence condition $\rho < m$ was used as one of the criteria in eliminating unsatisfactory check-out rules as explained in Chapter IV.

When the state probability distributions have been obtained, one can proceed to determine means of a number of quantities which are of interest in a queuing problem as well as explicit expressions for derived probabilities. Quantities of special interest in the current study are the mean number of customers in queue and the probability that the number of people in waiting line exceeds a certain pre-specified number.

Let L_q be the mean queue length.

$$L_q = \sum_{n=m+1}^{\infty} (n-m) \quad P_n$$

$$= \sum_{n} \frac{(n-m)\rho^n}{m!\, m^{n-m}} \quad P_o$$

$$= P_o \frac{\rho^{m+1}}{m!\, m} \left[1 + 2\frac{\rho}{m} + 3\left(\frac{\rho}{m}\right)^2 + \qquad \right]$$

It can be seen that the quantity within the brackets converges to a limit $\left(\frac{m}{m-\rho}\right)^2$ for $\rho < m$. Let $\frac{\rho}{m} = \pi$.

$$1 + 2\pi + 3\pi^2 + \ldots\ldots\ldots\ldots$$

$$= \quad 1 + \pi + \pi^2 + \ldots\ldots\ldots\ldots$$

$$\pi + \pi^2 + \ldots\ldots\ldots\ldots$$

$$\pi^2 + \ldots\ldots\ldots$$

$$= \frac{m}{m-\rho} + \frac{m}{m-\rho}\pi + \frac{m}{m-\rho}\pi^2 + \cdots\cdots\cdots$$

$$= \frac{m}{m-\rho}\left[1 + \pi + \pi^2 + \cdots\cdots\cdots\right]$$

$$= \left(\frac{m}{m-\rho}\right)^2$$

Hence $$L_q = \frac{P_0\, \rho^{m+1}}{(m-1)!(m-\rho)^2} \qquad (A.18)$$

Let Prob.$(q > q^*)$ denote the probability that the length of queue is greater than q^*. Since a queue exists when the number of customers exceeds the number of check-out stands in operation, i.e., $n > m$, the quantity $n-m$ is defined to be the length of queue.

Let $n-m = q$. q takes values from $0, 1, \ldots, q^*, \ldots$

The probability distribution of the random variate q is given by substituting q for $n-m$ in (A.15).

$$\text{Prob.}\,(q > q^*) = \sum_{q>q^*} P_0 \frac{\rho^{m+q}}{m!\, m^q}$$

$$= \frac{P_0}{m!} \frac{\rho^{m+q^*+1}}{m^{q^*+1}} \left[1 + \pi + \pi^2 + \cdots\cdots\cdots\right]$$

$$= \frac{P_0}{m!} \frac{\rho^{m+q^*+1}}{m^{q^*}(m-\rho)} \qquad (A.19)$$

APPENDIX B

THE DISTRIBUTION FUNCTION OF ESTIMATED AVERAGE SERVICE TIME

Let the joint density of n observations on service times be

$$h(t_1, t_2, \ldots, t_n) = \mu^n e^{-\mu \sum_{i=1}^{n} t_i} \qquad \text{(B.1)}$$

and let $\hat{\theta}$ denote the maximum likelihood estimate of average service time where

$$\hat{\theta} = \frac{\sum_{i=1}^{n} t_i}{n}$$

One way to obtain the distribution of $\hat{\theta}$ is to apply the following transformation to all the t_i's in the density (B.1):

$$\begin{pmatrix} w_1 \\ w_2 \\ \cdot \\ \cdot \\ w_n \end{pmatrix} = \begin{pmatrix} 1 & o & o & \ldots & o \\ 1 & 1 & o & \ldots & o \\ 1 & 1 & 1 & o \ldots & o \\ \vdots & & & & \vdots \\ 1 & \cdot & \cdot & \cdot & 1 \end{pmatrix} \begin{pmatrix} t_1 \\ t_2 \\ \cdot \\ \cdot \\ t_n \end{pmatrix} \qquad \text{(B.2)}$$

Then from the resulting transformed function, all the w's except w_n are integrated out. Note that $w_n = n\hat{\theta}$. From the formula (B.2), t's can be expressed in terms of w's explicitly.

$$\begin{pmatrix} t_1 \\ t_2 \\ \\ \cdot \\ \cdot \\ t_n \end{pmatrix} = \begin{pmatrix} 1 & o & . & . & o \\ -1 & 1 & o & . & o \\ o & -1 & 1 & o & o \\ \vdots & & \ddots & & \\ . & & & . & o \\ o & & & -1 & 1 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \\ \cdot \\ \cdot \\ w_n \end{pmatrix} \tag{B.3}$$

The Jacobian of the transformation in this case is

$$\left| \frac{\delta t_i}{\delta w_j} \right|_+ = 1$$

Hence we obtain the new density function of w's as follows:

$$h(t_1,\ldots,t_n) \left| \frac{\delta t_i}{\delta w_j} \right|_+ = \mu^n e^{-\mu w_n} \tag{B.4}$$

where t's are expressed in terms of w's by the relation (B.3). Next step is to find the appropriate region of integration for $w_1,\ldots,w_{n-1}$. From the first two equations in the system (B.3) and $o \leqq t_i < \infty$ for all i's, we see that $o < w_1 \leqq w_2$. Similarly from the 1st, 2nd and 3rd equations w_2 can be seen to take a value between o and w_3, etc., and finally $o < w_n \leqq y$. Hence the cumulative distribution of w_n is given by

$$\text{Prob } (w_n < y) = \mu^n \int_o^y \int_o^{w_n} \ldots\ldots\ldots\ldots \int_o^{w_2} e^{-\mu w_n} dw_1 \ldots\ldots dw_n$$

$$= \frac{n}{(n-1)!} \int_o^y e^{-\mu w} w_n^{n-1} dw_n. \tag{B.5}$$

Since we are interested in the distribution of $\frac{w_n}{n}$ we must make one more transformation.

Let $\frac{w_n}{n} = \hat{\theta}$ $\qquad \frac{dw_n}{d\hat{\theta}} = n$ .... Jacobian

Then

$$\text{Prob.}\left(\frac{w_n}{n} < y\right) = \frac{\mu^n}{(n-1)!} \int_0^y e^{-\mu n\hat{\theta}} (n\hat{\theta})^{n-1} \, n \, d\hat{\theta}$$

$$= \frac{(\mu n)^n}{(n-1)!} \int_0^y e^{-\mu n \hat{\theta}} \hat{\theta}^{n-1} d\hat{\theta} \qquad \text{(B.6)}$$

The distribution (B.6) can be integrated as follows:

$$\int_0^y e^{-\mu n \hat{\theta}} \hat{\theta}^{n-1} d\hat{\theta} = \left[- \frac{e^{-\mu n \hat{\theta}}}{(\mu n)^n} \left\{ (\mu n \hat{\theta})^{n-1} + (n-1)(\mu n\hat{\theta})^{n-2} + (n-1)(n-2)(\mu n\hat{\theta})^{n-3} + \dots + (n-1)! \right\} \right]_0^y$$

$$= \frac{(n-1)!}{(\mu n)^n} - \frac{e^{-\mu ny}}{(\mu n)^n} \left[(\mu ny)^{n-1} + \dots + (n-1)! \right] .$$

Hence

$$\frac{(\mu n)^n}{(n-1)!} \int_0^y e^{-\mu n\hat{\theta}} \hat{\theta}^{n-1} d\hat{\theta}$$

$$= 1 - e^{-\mu ny} \left[\frac{(\mu ny)^{n-1}}{(n-1)!} + \frac{(\mu ny)^{n-2}}{(n-2)!} + \dots + 1 \right]$$

$$= 1 - e^{-\mu ny} \sum_{i=1}^{n} \frac{(\mu ny)^{i-1}}{(i-1)!} \qquad \text{(B.7)}$$

The equation (B.7) is the desired distribution function. It is an incomplete gamma function and has the form of a Poisson distribution with the parameter μny.