Name: ___________________________

PROBABILITY DISTRIBUTION AND QUEUING MODULE

INTRODUCTION

One of the fundamental insights of the physical and social sciences in the 20th century is the applicability of probability theory. For example, the probabilism of quantum mechanics has replaced the determinism of Newtonian mechanics. Present day analysts speak of deterministic models versus probabilistic models, This module, entitled “Probability Distribution and Queuing,” is a continuation of the module on probability. After working through this module, the reader should
a) understand the concepts of probability distributions
b) see the applicability of probability distributions in a system of queues.
c) gain some insights on using queuing theory as a decision-making tool.

PROBABILITY DISTRIBUTIONS

Definition. A variable whose value is a number determined by the outcome of an experiment or process is called a *random variable*.

Everyone is familiar from high school algebra with x as the independent variable and y as the dependent variable. Random variable is the name used for the variable, usually represented by an upper case X, in probability theory.

Illustration 1)

Consider the experiment of tossing two pennies. We could ask this question: how many of the coins show "head"? The possible outcomes of this experiment is: (HH) , (WT) , (TH) , (TT) . Readers not sure of this result please consult page one of Probability Module. The answer to the question is: "head" appears either zero, or once, or twice. If we let the random variable, X, represent the number of "head" showing in this experiment, then X has the set of possible values: 0,1, 2.

We could also ask another question: what is the probability of the random variable, X, to take on the value of 0 or 1 or 2? X has the value of “0" only when (TT) shows; therefore the probability is ¼. If needed, the reader is advised to consult pages 5 and 6 of Probability Module. X has the value of "1" only when either (HT) or (TH) shows; therefore the probability is ½. X has the value of "2" only when (HH) shows; therefore the probability is ¼. The following table summarizes the experiment:

possible outcomes	number of heads	probability of outcome
(HH)	2	¼
(HT)	1	¼
(TH)	1	¼
(TT)	0	¼

random variable X = number of heads	associated probability of random variable X
0	¼
1	¼ (one head shows up twice in the possible outcomes; therefore ¼ + ¼ = ½)
2	¼

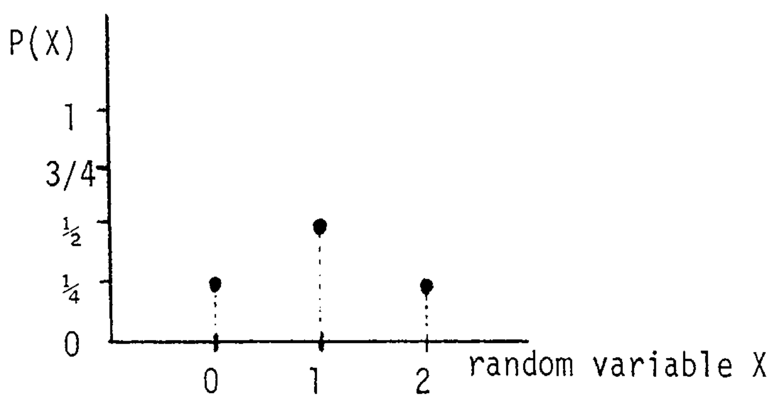
Note that the sum of the probabilities is 1.

In shorter notation we write:

P(X=0) = ¼ This reads: the probability of the random variable equals zero is ¼
P(X=1) = ½
P(X=2) = ¼

From this illustration we see that a random variable, in addition to having values (in this case 0, 1. 2) has an associated probability (this case: ¼, ½, ¼ respectively).

We could also graph the results of this experiment:



The heavy dots emphasize the functional relations, for example, at "0," P(0) is ¼, not 1/8 or 1/16, etc. The random variable X takes on the traditional independent variable axis and P(X), the probability, takes on the traditional dependent-variable axis. Note that X takes on only the values 0, 1, 2; therefore this is called a *discrete probability distribution*. P(X) is sometimes called the probability function of the discrete random variable X. In technical literature, P(X) is usual represented by f (X).

Illustration 2)

Consider the experiment of tossing one die. Let the random variable X take on the value of the number of dots showing on the top face of the die. The possible values of X are: 1, 2, 3, 4, 5, 6. Complete the chart:

random variable X = number of dots showing	associated prob. of random variable X
1	1/6
2	1/6
3	( )
4	( )
5	( )
6	( )

Complete the discrete probability distribution graph:

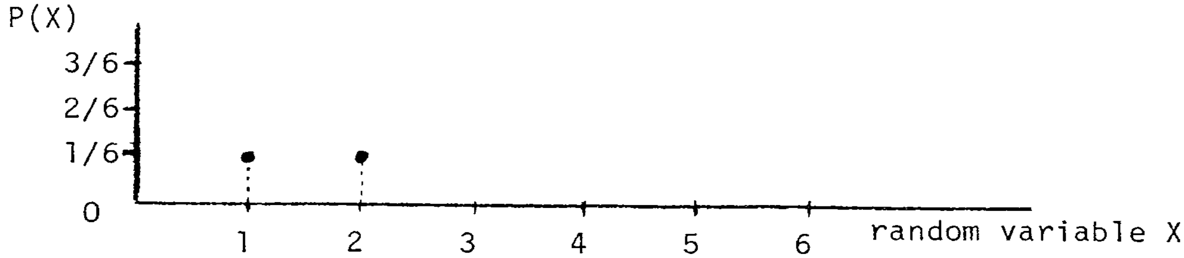
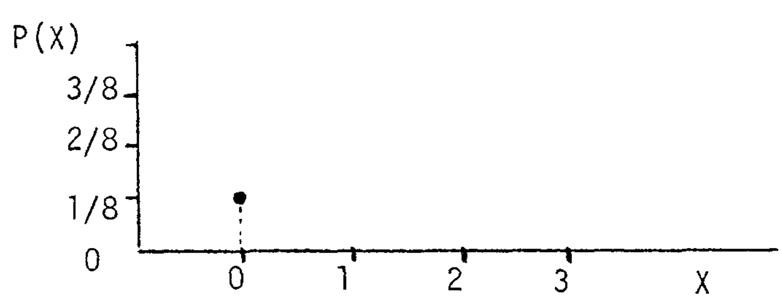


Illustration 3)

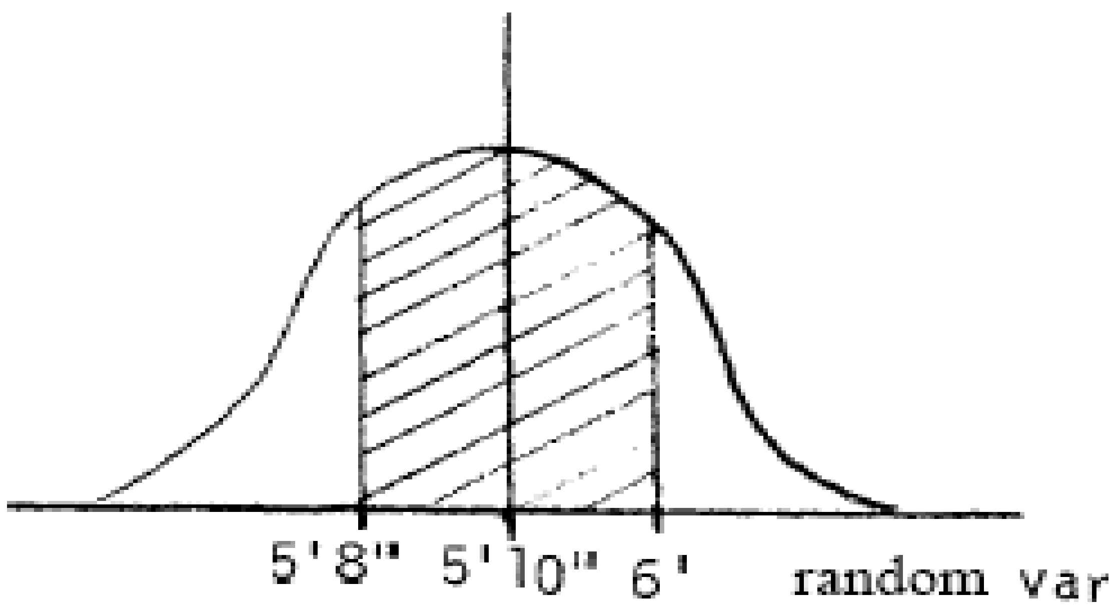
Consider the experiment of tossing three coins. Let the random variable X take on the values of the number of "heads" showing. The set of possible values for X is: 0, 1, 2, 3. If needed, please consult page 3 of Probability Module. Complete the chart and graph:

X	P(X)
0	1/8
1	( )
2	( )
3	( )



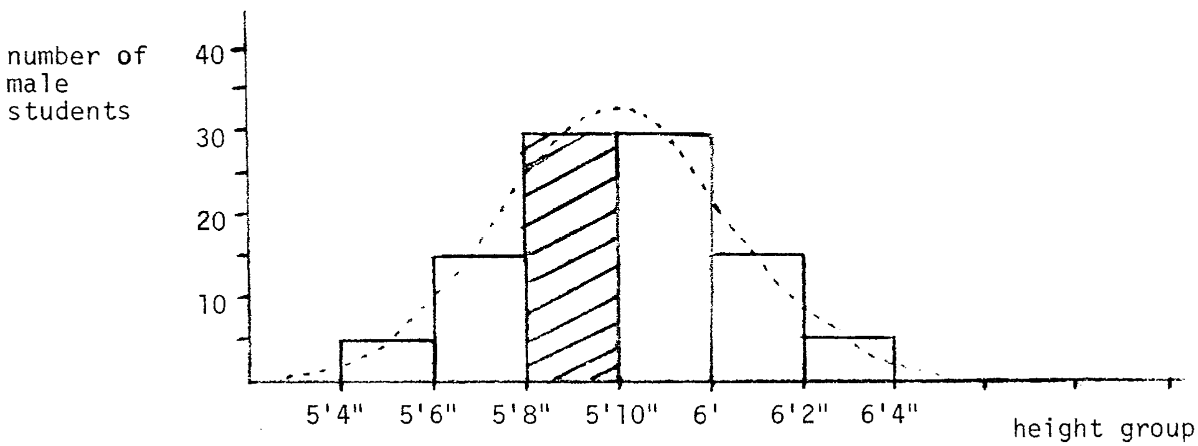
Consider the experiment of measuring the heights of every student on campus. The random variable in this experiment is the measured heights. What is the probability of finding a student *exactly* 6 feet tall? "Exactly 6 feet tall" means 6.0000000... feet tall. The probability is zero because we have no chance at all of finding a student 6.0000000 ... feet tall. We may be able to find some students between the intervals 6.0000000 ... ft and 5.0000000 ... ft; or between the intervals 5.5000000... ft and 5.0000000... ft. This experiment illustrates the concept of continuous probability distribution. It is called continuous because its random variable takes on continuous numerical values and not discrete numerical values. Whereas in a discrete probability distribution, of a random variable is represented by the height of the graph, for a continuous probability distribution, we represent the probability as the area between two intervals. The most famous continuous probability distribution is the bell curve, technically called the normal distribution.

This graph could represent the heights of all males on campus. The average height is 5'10". The probability of a male 5'10" is zero. The probability of a male between 5' 8" and 6' is the area of the shaded region. One of the characteristics of a probability distribution is the sum of the probabilities is one; therefore the total area under the curve of the normal distribution is one. The reader should verify this characteristic for the discrete distributions by going back to illustrations 1, 2, 3 and for each distribution, sum up the probabilities for values of the discrete random variable. They should sum to one.



Consider this procedure. We measure the heights of 100 male students. We plot the results into a histogram where the horizontal axis is the height groups and the vertical axis is the number of male students. The step-wise discrete histogram can be approximated by a continuous bell curve, as shown. The continuous curve can be standardized into the curve shown on page 3 with an area of one.

Height group	Number of male students
5'4" - 5'6"	5
5'6" - 5'8"	15
5'8" - 5'10"	30
5'10" - 6'	30
6' - 6'2"	15
6'2" - 6'4"	5



We could ask this question: what is the probability of picking out from this 100 male student group a student between 5'8" and 5'10"? The area of this height group on the histogram divided by the total area of the histogram gives the probability of picking one student with the height. The answer is 30 / 100 = 0.3

THREE FREQUENTLY USED DISTRIBUTIONS

In this section, we study three frequently used distributions: the binomial, the Poisson, the exponential. The binomial distribution is one of the most frequently used discrete probability distributions. This distribution is frequently used to <u>model</u> experiments or processes in which events occur either "success" or "failure." There are four characteristics of a binomial distribution model:

1) Each trial of the experiment falls into one of two categories: either "success" or "failure."

2) There are a fixed number of trials in the experiment. Call the number "n."

3) The outcomes of the trials are *independent* of one another. For example, getting a "head" on tossing a penny will not "influence" the next toss of the penny to get either a "head" or "tail."

4) For each trial, the probability of a success is the same constant, p. The binomial distribution, in functional form, is: $P(x) = \binom{n}{x} p^{x} (1 - p)^{n-x}$

where X is discrete and is the numerical value of the random variable, and

$\binom{n}{x}$ = n! / [x! (n−x)!]

Notice $\binom{n}{x}$ is not the same as n / x.

Illustration 4)

Let us use the binomial distribution to model the tossing of a penny. "n" is the number of times we toss the penny. Let X be the random variable taking on the value 0 when no head shows, 1 when a head shows, 2 when 2 heads show, 3 when 3 heads show, etc. We call "a head shows on a toss" a success; therefore p = ½.

a) What is the probability of getting a head in one toss?

n = 1, X = 1, p = ½

$P(X = 1) = \binom{1}{1} (1/2)^{1} (1/2)^{0}$

We expect this result from experience: the probability of getting a head on one toss is ½.

b) What is the probability of getting one head in two tosses? Just one head, not two heads.

n = 2, X = 1, p = ½

$P(X=1) = \binom{2}{1} (1/2)^{1} (1/2)^{1} = 2 \times ½ \times ½ = ½.$

c) What is the probability of getting exactly two heads in five tosses?

n = 5, X = ( ), p = ½

P(X=) = $\binom{5}{\ }$ $(1/2)^{(\ )}$ $(1/2)^{(\ )}$

d) What is the probability of getting exactly five heads in five tosses?

Illustration 5)

The Peekskill Precision Glass Co. is having trouble because only 90% of its glass is non-defective,

a) If we randomly 10 glass products, what is the probability of finding exactly eight non-defective products?

n = ( ), X = ( ), p = 0.9

P(X=) = ( ) $(0.9)^{(\ )}$ $(\ )^{(\ )}$ = ( )

b) If we select 10 glass products, what is the probability of finding exactly nine non-defective products?

n = ( ), X = ( ), p = 0.9

P(X=) = ( ) $(0.9)^{(\ )}$ $(\ )^{(\ )}$ = ( )

In these two examples, the random variable is the number of non-defective products, 8 and 9 respectively.

Another discrete distribution that's frequently used to model events is the Poisson distribution, named after its discoverer, the 19th century French mathematician Poisson, The Poisson distribution has the form

$P(X) = e^{-m} m^{x} / x!$

where X is the discrete random variable taking on the values: 1, 2, 3, 4, ... m is a positive number representing the average of the occurrence of the modeled event. e is the base of the natural log, 2.718 ....

This distribution is frequently used to model the occurrences of events in a fixed period of time or space. For example, the Poisson can be used to model:

- The number of wrong numbers dialed in a telephone exchange in a one-minute period.

- The number of defects in a newly manufactured airplane.
- The number of cars passing a point in a road in an one minute period.
- The number of customers arriving at a checkout counter in an one minute period.

Illustration 6)

Suppose a checkout counter at a supermarket has an average of two customer arrivals per five-minute interval.
a) What is the probability of one customer arrival per five minute?
b) What is the probability of four customer arrivals per five minute?

a) The random variable, X, is the number of customers arriving in the 5-minute interval.
m = 2, the average in a 5-minute interval.
X = 1
$P\ (X=1) = (e^{-2})\ (2^{1})\ /\ (1!) = (e^{-2})\ (2) = 0.27$

b) m = 2, the average in a 5-minute interval; X = 4.
$P\ (X = 4) = (e^{-2})\ (2^{4})\ /\ (4!) = (e^{-2})\ (2) = 0.09$

Illustration 7)

Suppose at a toll booth the average arrival of cars is one per one minute interval.

a) What is the probability of two cars arriving in a one-minute interval?

m = l; X = (   )
$P(X= \quad) = (e^{-1})\ (1^{(\quad)})\ /\ (\quad !) = (\qquad)$

b) What is the probability of three cars arriving in a one-minute interval?

c) What is the probability of four cars arriving in a one-minute interval?

This exercise illustrates our common sense expectation: if on the average one car arrives per one-minute interval, then the probability would decrease if we expect two cars to arrive, or three cars to arrive, or four cars to arrive, and so on.

The third probability distribution we examine is the exponential distribution:

$P(X) = a\ e^{-a\,x}$

where x is greater than or equal to zero. x can take on any values equal to or greater than zero; thus this is a *continuous probability distribution*. "e" is the base of the natural log, “1/a” is the average value of the random variable X.
A more useful form of the exponential distribution is:

$P(X \le t) = 1 - e^{-a x}$

This equation is read: the probability of the random variable taking on values less than or equal to time t is $1 - e^{-a x}$.

Illustration 8)

At a service counter, the average interarrival time of customers is five minutes. Interarrival time is the time between consecutive arrivals.
a) What is the probability of one customer arriving within a one-minute interval of the previous customer?
b) What is the probability of one customer arriving within a two-minute interval of the previous customer?
c) What is the probability of one customer arriving within a three-minute interval of the previous customer?

The random variable is the time interval in minutes.

a) $1/a = 5$ minutes; $a = 0.2$; $P(X \le 1) = 1 - e^{-(0.2)(1)} = 0.181$

b) $P(X \quad) = 1 - e^{-(\quad)} = 0.330$

c) $P(X \quad) = 1 - e^{-(\quad)} = 0.451$

These results match our common sense expectation that given a 5 minute average interarrival time of customers, as time increases from the "0" time, the probability of an arrival also increases. For this example, there is a probability of 0.181 of an arrival within a time period of one minute, probability of 0.330 of an arrival within a time period of two minutes, probability of 0.451 of an arrival within a time period of three minutes, and so no. In other words, if we wait long enough, eventually we will have an arrival.

Illustration 9)

At a service counter, the average service time per customer is 2 minutes.
a) What is the probability of servicing a customer in less than or equal to 3 minutes?
b) What is the probability of servicing a customer in less than or equal to 5 minutes?
c) What is the probability of servicing a customer in less than or equal to 10 minutes?

The random variable is the service time in minutes.

a) $1/a = 2$ minutes; $a = 0.5$; $P(X \le 3) = 1 - e^{-(0.5)(3)} = 0.777$

b) $P(X \quad) = 1 - e^{-(\quad)} = 0.918$

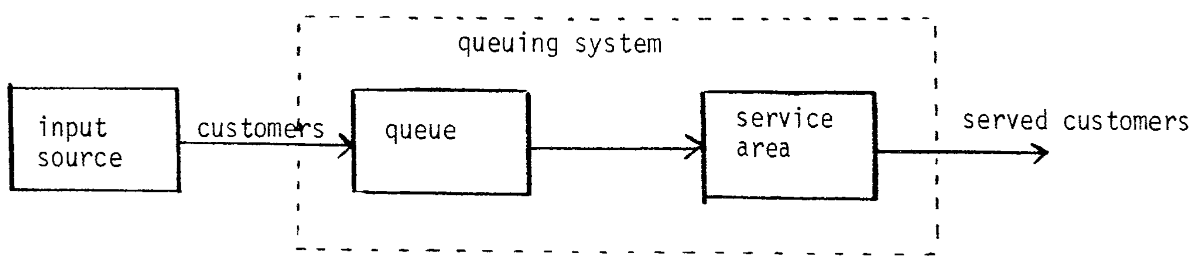
c) $P(X \quad) = 1 - e^{-(\quad)} = 0.993$

These results match our common sense expectation that given an average service time per customer of 2 minutes, as time increases from the initial time (the time at which the service begins), the probability of completing the service on the customer increases. In this example, the probability is 0.777 of servicing the customer in 3 minutes or less; probability is 0.918 for 5 minutes or less; probability is 0.993 for 10 minutes or less. Given a long enough time, once began, the customer will be serviced.

QUEUING MODEL

The reader should realize by now that the exponential distribution is the most useful probability distribution in modeling queues. The formulas in the textbook are derived from this distribution. The two most important statistical properties of a queuing system are the probability distributions of interarrival times and of service times. In real queuing systems, the distributions of interarrival times and service times can take on almost any form. However, to a queuing model, the exponential distribution is used to model the interarrival times and service tines. From illustrations 8, 9, the reader should convince himself or herself that the exponential distribution is sufficiently realistic and it also has the virtue of being mathematically simple.

An elementary queuing model can be diagramed:



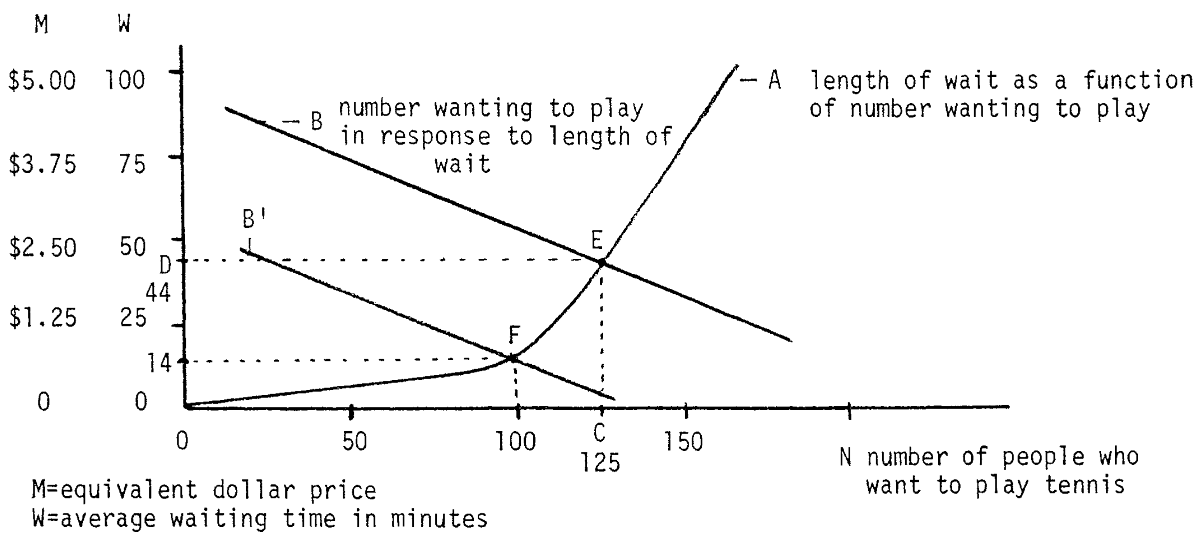
One of the assumptions of this model is the input source produces customers "randomly." Randomly means each arrival is independent of the other arrivals. An example of *non-random* arrivals is the exit from a subway train station. Each person leaving the train station is "dependent" to each other because they all arrived at the station on the same train. Another assumption is that the input source has an infinite supply of customers. This assumption is to allow easier mathematical calculations. Another assumption is that the generation of new customers is not affected by the number of customers ready in the queuing system. The point is clear: in every model, assumptions must be made and clearly explained.

All elementary queuing models are said to have a *Poisson input* and an *exponential service time*. However, to model the input process, the exponential distribution is used in place of the Poisson distribution because of the former's mathematical simplicity. The exponential distribution's modeling of *interarrival time* is equivalent to the Poisson distribution's modeling of *an average arrival rate*.

In the service area, the elementary queuing model's service discipline is first come first served, otherwise, more mathematical complications. The exponential distribution again is used to model the service time. The final major assumption for the service-area component is the distribution is applied to all servers and all customers.

QUEUING THEORY AS A DECISION MAKING TOOL: CASE STUDY[1]

.
Everyone hates standing in a queue. No one derives any benefits or pleasure from a queue because the time spent in a queue is wasted time, and to the economist and to almost everyone, the wasted time represents wasted productivity and lost money. Consider the case of the Peekskill Country Club. The Club has 500 members each paying an annual base fee of $300. Its Board of Directors decided to charge no fee for its limited tennis facilities. The Board also decided not to have an advance sign up system and decided to have a simple first-come-first-play policy. During the warm season, everyone wants to play tennis and obviously queues are developed everyday and result in wasted time.



In the graph, curve A represents the average waiting time as a function of the number of people who want to play tennis. The shape of the curve is reasonable because we expect waiting time to increase with an increase in the number of people who want to play. Curve B represents the number of people who want to play in response to the average waiting time. It is reasonable to assume that as the average waiting time increases, fewer people want to play. Point E is the equilibrium, the intersection of curves A and B. At E, "C" number of people want to play, with a “D” average waiting time. C is 125 and D is approximately 44 minutes.

We assume everyone values his or her waiting time at $3.00 per hour. In other words, a person doesn't care whether he or she pays a $3.00 fee to play immediately or waits an hour to play for free, Accordingly, another vertical scale, M = equivalent dollar

[1]Stokey, E.: Zeckhauser, R. (1978). *A Primer for Policy Analysis*. Norton, New York.

price, is put on the graph. For example, a person either pays \$3.75 to play immediately or waits 75 minutes; either choice is indifferent to the person. At point E, the average wait of 44 minutes is worth \$2.20 for each person waiting; \$2.20 is also the value for the marginal tennis player. A marginal player would pay either \$2.20 to play immediately or wait 44 minutes; but a more dedicated tennis player may be willing to pay, for example, \$3.75 to play immediately or wait 75 minutes.

The 44-minute wait was too long and everybody started to complain to the Board. The Board investigated the matter and realized the situation fitted the characteristics of a queuing model. The 44 minutes waiting time represented a so called "deadweight loss," i.e., no one benefits from the waiting time. The Board decided to impose a \$2.00 fee per person per hour of play time. With this \$2.00 fee, the tennis players naturally will not be willing to wait as long. Therefore, the curve of number wanting to play in response to length of wait goes down to B' (see graph). This results in a new equilibrium at F. At the point F, 100 people wait on an average of 14 minutes to play. So now everyone is happy. The average waiting time is diminished by 30 minutes, a decrease in "deadweight loss." The country club takes in 100 x \$2 = \$200 per day, assuming each person plays one hour. The dedicated tennis players keep paying the \$2.00 while the less dedicated tennis players decide to do other things.