CHAPTER II

THEORETICAL CONSIDERATION

It has been stated earlier that the main problem of this research is to generate the arrivals and services of vehicles at the intersection. It is the concept of traffic situation that the vehicles are random run into the intersection with various form of probability distributions, and there is tendency to be served by a quite specific form. It has been accepted in the traffic flow theory that the generation of some continuous and discrete distribution functions are close to the real-world of traffic flow. They have been observed and used for applying in simulations, especially in generation of the arrivals and services of vehicles. It is very important to know that every generation for such distribution is based on the generator of random numbers. It is necessary since it is the control function of the generation in the computer programming. This has already been prepared for using in many systems of digital computer. Because this research is studied under the IBM 1130 computing system, so a specific subroutine for generating the random numbers, called RANDU is going to be used. Some specifications and detail about the usage of this subroutine are listed in APPENDIX B.

The general ideas and techniques of generating the probability distributions which are to be used in the research are explained in the following section.

Continuous Probability Distributions

The Uniform Distribution Perhaps the simplest continuous probability density function is the one that is constant over the interval (a,b) and is zero otherwise. This density function defines what is known as the uniform or rectangular distribution. The uniform distribution may arise in the study of rounding errors when measurements are recorded to a certain accuracy. For example, if measurements of weights are recorded to the nearest gram, one might assume that the difference in grams between the actual weight and the recorded weight is some number between -0.5 and +0.5 and that the error is uniformly distributed throughout this interval. The principal value of the uniform distribution for simulation techniques lies in its simplicity and the fact that it can be used to simulate random variables from almost any kind of probability distribution.

Mathematically the uniform density function is defined as follows:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases} \qquad (2.1)$$

Here X is a random variable defined over the interval (a,b). The graph of the uniform distribution is illustrated by Fig. 2-1.

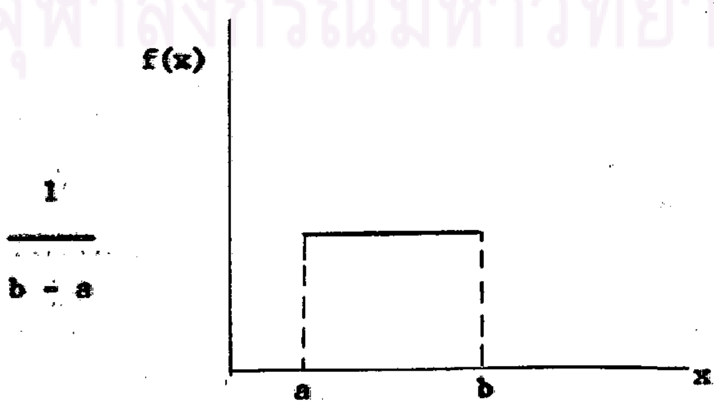

Fig. 2-1

The cumulative distribution function F(x) for a uniformly distributed random variable X is

$$F(x) = \int_a^x \frac{1}{b-a}\,dt = \frac{x-a}{b-a} \qquad (2.2)$$

$$0 \leq F(x) \leq 1.$$

The expected value and variance of a uniformly distributed random variable are given by,

$$EX = \int_a^b \frac{1}{b-a}\,x\,dx = \frac{b+a}{2} \qquad (2.3)$$

$$VX = \int_a^b \frac{(x-EX)^2}{b-a}\,dx = \frac{(b-a)^2}{12} \qquad (2.4)$$

In actual applications the parameters of the uniform density function (2.1) (i.e., the numerical value of a and b) may not necessarily be known directly. Typically, although not for uniform distributions, we know only the expected value and variance of the statistics to be generated. In this case the values of the parameters must be derived by solving the equation system consisting of Eqs. 2.3 and 2.4 for a and b, since EX and VX are assumed to be known. This procedure-similar to an estimation technique known in statistical literature as " the method of moments " _ provides the following two expressions:

$$a = EX - \sqrt{3VX} \qquad (2.5)$$

$$b = 2EX - a \qquad (2.6)$$

To simulate a uniform distribution over some given domain (a,b) we must first obtain the inverse transformation[2] for Eq. 2.2

$$x = a + (b - a)r \qquad 0 \leq r \leq 1. \qquad (2.7)$$

It is then generated a set of random numbers corresponding to the range of cumulative probabilities, i.e., uniform random variates defined over the range 0 to 1. Each random number r determines uniquely a uniformly distributed variate x.

A graphical explanation will perhaps serve to clarify the issues here. Fig. 2-2 illustrates that each generated value of r is associated with one and only one value of x. For example, the specific value of the cumulative distribution function at r_0 fixes the value of x at x_0. Obviously, this procedure can be repeated as many times as desired, each time generating a new value of x. Generating random variates through the use of cumulative probabilities will also be followed in simulating several other distributions in this paper. Furthuermore this technique serves as the basis for developing the more general Monte Carlo methods, discussed later.

Figure 2-3 contains a flow chart of the logic that must be utilized in simulating a uniform distribution for a given range (a,b), if it is to be programmed for use on a computer. The flow chart has been formulated in such a manner that it is compatible with the FORTRAN subroutine, which follows in Fig. 2-4.

```
1. SUBROUTINE UNFRM ( A,B,X )
2. R = RND(R)
3. X = A + (B - A) * R
4. RETURN
```

Fig. 2-4 Generation of Uniform variates, FORTRAN subroutine

[2] Thomas H. Naylor, Computer Simulation Techniques : " Generation of Stochastic Variates for Simulation ", 1966, pp 70-73.

<u>The Exponential Distribution</u> Throughout our daily life we observe time intervals between the occurences of distinct random events. We receive information about numerous events that take place in our environment such as births, deaths, accidents, and world conflicts on the basis of a completely independent time schdule. If the probability that an event will occur in a small time interval is very small,and if the occurence of this type is exponentially distributed.Whether a stochastic process in the real world actually yields exponential variates or not is an empirical question, whose answer depends on the degree to which the assumptions underlying the exponential distribution are satisfied. Specifically, the following assumptions must be satisfied by exponential variates.

1. The probability that an event occurs during the time interval $[t,(t + \Delta t)]$ is $\alpha\Delta t$.

2. α is a constant and independent of t and other factors.

3. The probability that more than one event will occur during the time interval $[t,(t + \Delta t)]$ approaches 0 as $\Delta t \to 0$ and is of a smaller order of magnitude than $\alpha\Delta t$.

Curiously enough, the behavior of a number of time dependent processes has been found to satisfy these rather strong assumptions. For example, the time interval between accidents in factory, the arrival of orders at a firm, the arrival of patients in hospital, and the arrival of vehicles at intersections have been found to follow the exponential distribution.

A random variable X is said to have an exponential distribution if its density function is defined as,

$$f(x) = \alpha e^{-\alpha x} \qquad (2.8)$$

for $\alpha > 0$ and $x \geq 0$.

The cumulative distribution function of X is

$$F(x) = \int_0^x \alpha e^{-\alpha t}\, dt = 1 - e^{-\alpha x}, \qquad (2.9)$$

and the expected value and variance of X are given by the following formulas

$$EX = \int_0^\infty x \alpha e^{-\alpha x}\, dx = \frac{1}{\alpha} \qquad (2.10)$$

$$VX = \int_0^\infty \left(x - \frac{1}{\alpha}\right)^2 \alpha e^{-\alpha x}\, dx = \frac{1}{\alpha^2} = (EX)^2 \qquad (2.11)$$

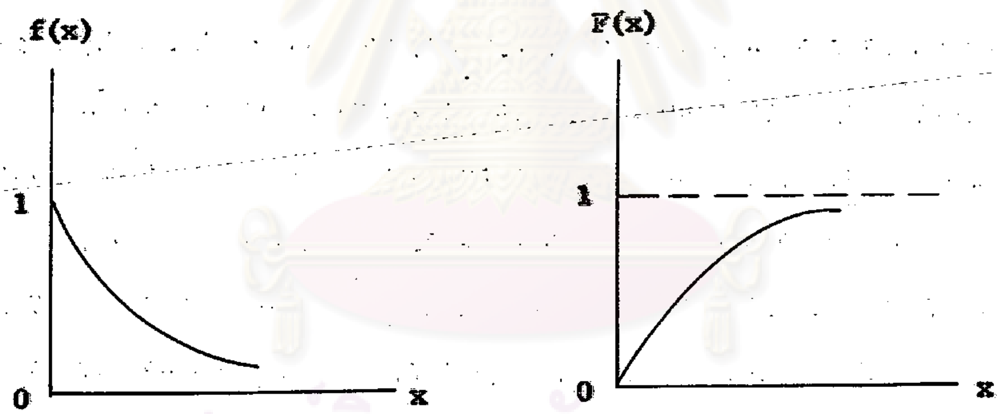

Fig. 2-5

Graphically the exponential distribution appears as in Fig. 2-5. Since the exponential distribution has only one parameter it is possible to express as,

$$\alpha = \frac{1}{EX} \qquad (2.12)$$

The generation of exponential random variates can be accomplished in a number of different ways. Since F(x) exists in explicit form, the inverse transformation technique provides a straightforward method. Because of the symmetry of the uniform distribution, F(x) and 1-F(x) are interchangeable. Therefore,

$$r = e^{-\alpha x} \tag{2.13}$$

and consequently,

$$x = -\left(\frac{1}{\alpha}\right) \log r = - EX \log r \ . \tag{2.14}$$

Thus for each value of the pseudorandom number r, a unique value of x is determined, which will take only nonnegative value (since $\log r \leq 0$, for $0 \leq r \leq 1$) and will follow the exponential density function (2.8) with expected value EX. Although this technique seem very simple, the reader is remined that the computation of the natural logarithm on digital computers includes a power series expansion (or some equivalent approximation technique) for each uniform variate generated.

Fig. 2-6 contains a flow chart for generating exponential variates, and Fig. 2-7 contains the corresponding FORTRAN subroutine. The name of the subroutine is EXPOL.

Discrete Probability Distributions

A significant number of probability distributions are defined on random variates that only discrete, non-negative integer values. The cumulative probability distribution for a discrete random variable X is defined as follow:

$$F(x) = P\ (\ X \leq x\ ) = \sum_{X=0}^{x} f(x) \tag{2.15}$$

where f(x) is the frequency function of X defined for integer x values such that

$$f(x) = P\ (X = x) \qquad (2.16)$$

for x = 0, 1, 2, ...

Discrete probability distributions serve as stochastic models for certain counting processes over either finite or infinite samples, where the presence or absence of a binary attribute is governed by chance. Empirically, discretedistributions may also occur as a result of rounding continuous measurements on a discrete scale. Strictly speaking, however, discrete probability distributions are appropriate models of random phenomena only if the values of the random variates are measurable by counting.

The following section contain descriptions of techniques for generating stochastic variates from which is applicable for the research.

Poisson Distribution If we take a series of n independent Bernoulli trials, in each of which there is small probability p of an event occurring, then as n approaches infinity, the probability of x occurences is given by the Poisson distribution

$$f(x) = e^{-\lambda} \frac{\lambda^x}{x!} \qquad x = 0, 1, 2, ... \qquad (2.17)$$

$$\lambda > 0,$$

when p is allowed to approach zero in such a manner that $\lambda = np$ remains fixed. It is known that np is the expected value of the binomial distribution, and it can be shown that λ is the expected value for the Poisson distribution. In fact, both the expected value and the variance of Poisson distribution are equal to λ. It can be shown that if x is a Poisson variable with parameter λ, then for large values of λ, ($\lambda > 10$).

the normal distribution with $EX = \lambda$ and $VX = \lambda$ can be used to approximate the distribution of x.

Poisson distributed events frequently occur in the real world. For example, the number of aircraft arriving at an airport during a twenty-four-hour period can be very large. Yet theprobability of an aircraft arriving during a particular second is very small. Hence, we might expect the probability of 0,1,2, ... aircraft arriving in a given period of time to follow a Poisson distribution. The Poisson distribution is particularly useful in dealing with the occurence of isolated events over a continuation of time, or when it is possible to prescribe the number of times an event occurs but not the number of times it does not occur.

To simulate a Poisson distribution with a parameter λ, it is taken the advantage of the well-known relationship between the exponential and Poisson distributions. It can be shown that if (1) the total number of events occurring during any given time interval is independent of the number of events that have already occurred prior to the beginning of the interval and (2) the probability of an event occurring in the interval t to $t + \Delta t$ is approximately $\lambda \Delta t$ for all values of t, then (a) the density function of interval t between the occurence of consecutive events is $f(t) = \lambda e^{-\lambda t}$, and (b) the probability of x events occurring time t is

$$f(x) = e^{-\lambda t} \frac{(\lambda t)^x}{x!} \qquad \text{for all x and t} \qquad (2.18)$$

Consider a time horizon (beginning at reference point 0) that has been divided into unit time intervals as illustrated in Fig. 2-8. Events are assumed to occur along the time horizon and are denoted by the symbol ($\wedge$). The time interval t between events is assumed to have an exponential distribution withcwith expected value equal to $1/\lambda$. This implies that the number of events x occurring during a unit time interval follows a Poisson distribution with expected value equal to λ . One method of generating Poisson variates involves generating exponentially distributed time intervals $t_1, t_2, t_3, \ldots$ with expected value equal to 1. These random time intervals are accumulated as they are generated until their sum exceeds.

In mathematical terms the Poisson variate x is determined by the inequality

$$\sum_{i=0}^{x} t_i \leq \lambda < \sum_{i=0}^{x+1} t_i \qquad (x = 0, 1, 2, \ldots), \quad (2.19)$$

where the exponential variates t_i are generated by the formula

$$t_i = - \log r_i \qquad (2.20)$$

with unit expectation. A faster method of generating Poisson variates x call for rewriting of Eq. 2.19 as

$$\prod_{i=0}^{x} r_i \geq e^{-\lambda} > \prod_{i=0}^{x+1} r_i \qquad (2.21)$$

The FORTRAN subroutine for Poisson distribution is not necessary for this research since the Exponential distribution is capable for generating the arrivals. All explanation above leads to the understanding about the relation of Exponential distribution and Poisson distribution only.

<u>Empirical Discrete Distributions</u> All distribution described before is concerning about the standard probability distribution, now turn the point to somewhat more general method that can be used to simulate: (1) any empirical distribution, (2) any distribution, and (3) any continuous distribution that can be approximated by discrete distribution. However, in general, it would not be used this method to generate variates from the standard probability distributions because one of the methods described previously would be expected to yield "better" results from the standpoint of computation speed, ease of programming, and memory requirements. In other words the method propose in this section is a method to use when no other alternative is available.

Let X be a discrete random variable with $P\ (X = b_i) = p_i$, such as the random variable in the following table.

b_i	$P(x = b_i) = p_i$
b_1	0.273
b_2	0.037
b_3	0.195
b_4	0.009
b_5	0.124
b_6	0.058
b_7	0.062
b_8	0.151
b_9	0.047
b_{10}	0.044

Clearly one method of generating x on a computer is to generate a uniform (0,1) random variate r and set $x = b_i$ if

$$p_1 + \ldots\ldots + p_{i-1} < r \leq p_1 + \ldots\ldots + p_i \qquad (2.22)$$

Although a number of search techniques based on this method have been developed, most of them involves relatively complicated programs requiring excessive computational time .

All of probability distributions mentioned, are applied to the subprograms named TEXPA, TNORS, DISCR and UNFRM. The procedure will be discussed in the next chapter. In addition to making the comupter programming more efficient, some techniques, described later, will be needed to combine with those FORTRAN subprograms previously mentioned.