Chapter I

INTRODUCTION

Traditionally, engineers have used formulas or "laws" to describe the behavior of physical systems. Although all physical systems have randomness in their behavior, most engineering analysis treats systems as though they were deterministic (nonrandom). This practice can be very acceptable where the particles are very small and the populations of particles are very large, such as when dealing with gases or electric currents. When, however, the populations become relatively small and the particles relatively large so that the observer finds himself measuring the behavior of individual particles, attention to random properties becomes all-important. Such is the situation with automobile traffic.

This monograph deals with some of the relationships which have been found useful in handling the random properties of traffic. One principal tool is the Poisson distribution, named after Simeon Denis Poisson, a French mathematician who carried out many of the early studies on probability. As will be seen, however, the Poisson distribution has certain limitations in its application, and other distributions may provide greater accuracy in certain cases.

Nature of the Poisson Distribution

When events of a given group occur in discrete degrees (heads or tails, 1 to 6 on the face of a die, etc.) the possibility of occurrence of a particular event in a specified number of trials may be described by the Bernoulli or binomial distribution (See Appendices B and C).

As an example, let us consider an experiment consisting of five successive drawings of a ball from an urn containing uniformly mixed black and white balls, with the drawn ball being returned to the urn after each drawing. Of the five drawings which comprise a single experiment, let x be the number which produced black balls, i.e., the number of black balls in a sample of 5 where each drawing had a uniform probability. Thus x can equal 0, 1, 2, 3, 4, or 5. Let P(x) be the probability that in a given experiment the number of black balls would be exactly x. If p is the probability that a particular drawing will yield a black ball, and q $(q = 1 - p)$ is the probability that a particular drawing yields a white ball, then

$$P(x) = C_x^5 p^x q^{5-x}$$

where C_x^5 is the number of combinations of 5 things taken x at a time. It can be seen that in the above example p, the probability that a single drawing yields a black ball, is equal to the percentage (written as a decimal fraction) of black balls in the urn. With the ball being replaced after each drawing p will remain constant from drawing to drawing.

The experiment just described is an example of the "Bernoulli" or "binomial" distribution in which P(x), the probability of exactly x successes out of n trials of an event where the probability of success remains constant from event to event, is given by

$$P(x) = C_x^n p^x q^{n-x}$$

If the number of items in the sample n, becomes very large while the product $pn = m$ is a finite constant, the binomial distribution approaches the Poisson distribution as a limit. This implies that the probability of occurrence, p, becomes very small.

$$\lim_{n\to\infty} P(x) = \frac{m^x e^{-m}}{x!}$$
$$pn = m$$

where e is the Napierian base of logarithms. $(e = 2.71828 \ldots)$

The derivation of the Poisson distribution as a limiting case of the binomial distribution is given in Appendix D. (The Poisson distribution can be derived independently of the binomial distribution by more advanced concepts. See Appendix D or Fry (*1*).

The mathematical conditions of an infinite number of trials and an infinitesimal probability are never achieved in practical problems. Nevertheless, the Poisson distribution is useful as approximating the binomial distribution under appropriate conditions.*

For such practical purposes, then, the Poisson distribution may be stated as follows:

If in a given experiment the number of opportunities for an event to occur is large (e.g., $n \geqslant 50$)

and

If the probability that a particular event occurs is small (e.g., $p \leqslant 0.1$)

and

If the average number of times the event occurs has a finite value, m $(m = np)$

$$\text{Then} \quad P(x) = \frac{m^x e^{-m}}{x!}$$

where $x = 0, 1, 2, \ldots$

In this statement of the Poisson distribution an experiment may consist of such things as:

a. Observing the number of micro-organisms in a standard sample of blood, x representing the number of micro-organisms in any one sample.
b. Observing the number of alpha particles emitted during each successive interval of t seconds. The number of such intervals will be j, and $x_1, x_2 \ldots x_j$ will be the number of particles during the 1st, 2nd, . . . jth intervals.

* It may also be noted that under appropriate conditions the binomial distribution may be approximated by the normal distribution.

c. Observing the number of blowholes in each of k castings, x representing the number of holes in any one casting.
d. Observing the number of cars passing a given point during each 30-second period, i representing the number of periods observed and x representing the number of cars in any 30-second period.

Historical Background of Poisson Distribution

The first record of the use of the Poisson distribution to treat populations having the properties described is attributed to Bortkiewicz who studied the frequency of death due to the "kick of a horse" among the members of ten Prussian cavalry corps during a period of 20 years (*2*). A summary of his study is shown in Example 1, which compares actual and computed frequencies.

Example 1

The Number of Men in Ten Prussian Cavalry Corps Killed by a Horse Kick in the Twenty Years 1875–1894

Number of deaths per corps-year	*Observed number of corps-years during which the given number of deaths occurred.*	*Theoretical number of corps-years during which the given number of deaths occurred (as computed from the Poisson distribution)*
0	109	108.7
1	65	66.3
2	22	20.2
3	3	4.1
4	1	0.6
5 and over	0	0.1

Some of the earliest engineering problems treated by the Poisson distribution were telephone switching problems. The following example is based on such data (*3*):

Example 2

Connections to Wrong Number

Number of wrong connections per period	*Observed number of periods during which the given number of wrong connections occurred*	*Theoretical number of periods exhibiting the given number of wrong connections (Poisson distribution)*
0	0	0.0
1	0	0.3
2	1	1.6
3	5	4.8
4	11	10.4
5	14	18.2
6	22	26.4
7	43	33.1
8	31	36.0
9	40	35.2
10	35	30.7
11	20	24.3
12	18	17.9
13	12	12.0
14	7	7.5
15	6	4.3
16	2	2.4
>16	0	1.9

Following the pioneer work in the field of telephone applications, the Poisson distribution was gradually applied to other engineering problems. The following example adapted from Grant (*4*) shows an application to the occurrence of excessive rainfall:

Example 3

RAINSTORMS

Number of storms per station per year	*Observed number of occurrences*	*Theoretical number of occurrences (Poisson distribution)*
0	102	99.3
1	114	119.1
2	74	71.6
3	28	28.7
4	10	8.6
5	2	2.0
>5	0	0.7
Total	330	330.0

The application of the Poisson distribution to traffic problems is not new. Certain applications were discussed by Kinzer (*5*) in 1933, Adams (*6*) in 1936, and Greenshields (*7*) in 1947.

The first published examples were those of Adams. The following is one of his examples:

Example 4

RATE OF ARRIVAL (Vere St.)

(Number of vehicles arriving per 10 second interval)

Number of vehicles per 10 second period	*Observed frequency*	*Total vehicles*	*Theoretical frequency*
0	94	0	97.0
1	63	63	59.9
2	21	42	18.5
3	2	6	3.8
>3	0	0	0.8
Total	180	111	180.0

Note: Since there were 111 vehicles in 180 ten-second periods, the hourly volume was 222.

REFERENCES FOR CHAPTER I

1. Fry, Thornton C., *Probability and Its Engineering Uses,* Van Nostrand, 1965, pp. 240–243.
2. Bortkiewicz, Ladislaus von, *Das Gesetz der kleinen Zahlen,* B.G. Teubner, Leipzig, 1898.
3. Thorndike, Frances, "Application of Poisson's Probability Summation," *Bell System Technical Journal,* vol. 5, no. 4, October 1926, pp. 604–624.
4. Grant, Eugene L., "Rainfall Intensities and Frequencies," *Transactions,* American Society of Civil Engineers, vol. 103, 1938, pp. 384–388.
5. Kinzer, John P., *Application of the Theory of Probability to Problems of Highway Traffic,* thesis submitted in partial satisfaction of requirements for degree of B.C.E., Polytechnic Institute of Brooklyn, June 1, 1933.
   Abstracted in: *Proceedings,* Institute of Traffic Engineers, vol. 5, 1934, pp. 118–124.
6. Adams, William F., "Road Traffic Considered as a Random Series," *Journal,* Institution of Civil Engineers, vol. 4, Nov., 1936, pp. 121–130.
7. Greenshields, Bruce D., Schapiro, Donald, and Ericksen, Elroy L., *Traffic Performance at Urban Street Intersections,* Technical Report No. 1, Yale Bureau of Highway Traffic, 1947.

CHAPTER 2

The Poisson Process

2.1. Definition and examples

In several of the applications outlined in Chapter 1 a natural first hypothesis is that events are occurring completely randomly in time. Thus, in Example 1.1, if we wish to test the reality of the apparent trend in the accident rate, we take a null hypothesis that accidents occur randomly at constant rate. Such a hypothesis can be considered for two reasons. First, before we can have confidence in the reality of an apparent systematic effect in the series, we need to show that the effect is unlikely to have arisen just by chance. Secondly, simple methods for the comparison of the rates of occurrence in different series are available whenever the individual series can be assumed to be completely random.

As a mathematical model of a completely random series of events we consider a *Poisson process*. The definition and main probabilistic properties of the Poisson process are well-known and in particular are discussed very carefully by Feller (1957, p. 400). Therefore here we merely review these results quickly and then pass to the statistical procedures connected with a single Poisson process. These procedures are mostly standard ones, but it is useful to collect them together. In Chapter 9 we deal with methods for the comparison of two or more Poisson processes.

Consider events occurring along a line which for definiteness will be called the time axis. The reader who likes to think in terms of specific applications may consider each event to be an emission from a radioactive source. Let λ be a constant with the dimensions of the reciprocal of time. It will measure the mean rate of occurrence of events over a long period of time and will be called the *rate of occurrence* or more fully the *probability rate of occurrence* of events. Denote by $N_{t,t+h}$ the random variable defined as the number of events occurring in $(t, t+h]$, where $h > 0$. The conditions for a Poisson process of rate λ

$$\text{prob}(N_{t,t+h} = 0) = 1 - \lambda h + o(h), \quad (1)$$

$$\text{prob}(N_{t,t+h} = 1) = \lambda h + o(h), \quad (2)$$

and that the random variable $N_{t,t+h}$ is statistically independent of the number and position of events in $(0, t]$. This last condition is to hold for all $t, h > 0$. Equations (1) and (2) imply that

$$\text{prob}(N_{t,t+h} \geqslant 2) = o(h). \quad (3)$$

In these equations the terms $o(h)$ denote any quantities which tend to zero more rapidly than h as $h \to 0$, i.e. they denote functions $f(h)$ such that, as $h \to 0$, $\lim\{f(h)/h\} = 0$. Note particularly that in (1)–(3) the terms $o(h)$ represent three different quantities all having the required limiting behaviour.

The central features of the definition of the Poisson process are that

(*a*) the probabilities (1)–(3) do not vary in time, so that there is no trend in the series;

(*b*) the chance of two or more events occurring simultaneously is negligible;

(*c*) the chance of an event in $(t, t + h]$ is quite independent of what happens up to t. In particular the chance is not affected by the time that has elapsed since the last preceding event.

The Poisson process is a mathematical concept and no real phenomenon can be expected to be exactly in accord with it. Whether or not a particular series is in reasonable agreement with a Poisson process is ultimately an empirical matter, even though the key assumptions have varying degrees of plausibility in different applications.

Example 2.1. *Telephone calls*. Consider a telephone exchange serving a large number of subscribers and let an event be a request for a call. Then provided that we consider a fairly short period of time, say 15–30 min., the conditions for a Poisson process are likely to be closely fulfilled. The occurrence of a call from one subscriber at time t will in general affect the chance of a further call arriving at time $t + h$ from that same subscriber, at any rate for small h. If the total number of subscribers were small, this would mean that condition (*c*) would not usually be applicable. However, when, as is assumed here, the number of subscribers is large and different subscribers act independently of

one another, the chance of a call arising at one instant is unlikely to be appreciably affected by what has happened in preceding instants. The effect of superimposing a number of independent sequences of events is analysed mathematically in Chapter 8.

Condition (*a*) is certain to fail if a long period of time is considered, for the density of the calls will fluctuate between different times of the day. Hence it is reasonable to hope that a Poisson process with a parameter varying slowly and systematically in time will be a very good representation of the observed distribution of requested calls. Time-dependent Poisson processes will be discussed in Section 2.2(v).

Example 2.2 *Radioactive source*. Suppose that an event is an emission from a radioactive source. A Poisson process is, for substantially the reasons outlined for Example 2.1, again likely to be a good representation of the occurrences over a period during which the strength of the source remains substantially constant. The events recorded by an electronic counter, however, do not form a Poisson process. This is because of counter dead times; following each recorded event, or sometimes following each event, there is a period during which no further events can be recorded. This means that condition (*c*) is certainly not satisfied for the series of recorded events.

The probability problems connected with correcting the recorded series for loss due to dead time have been extensively treated; Smith (1958) has given a concise guide to the literature.

Example 2.3. *Stops of machine*. Let the time axis represent the running time of a machine, such as a loom, subject to intermittent stops and suppose that an event is a stop of a specified type. Here the applicability of the Poisson process will very much depend on the nature of the stop. Some types of stop such as those connected with the completion of an operation, will occur fairly regularly. Others, primarily those due to machine faults or defective material, may be expected to occur irregularly; it does not follow though that they are necessarily well described by a Poisson process. For example they may tend to occur in bursts, long stretches with very few stops being interspersed with short periods with a high rate of occurrence. In many cases, however, the Poisson process is a reasonable first approximation although the parameter λ is likely to fluctuate in time and possibly to change suddenly, for example when a new batch of raw material is introduced.

2.2. Properties of the Poisson process

(i) *Distribution of number of events*

We now review, mostly without proof, the main properties of the Poisson process.

First consider the number of events, N_t, occurring in an arbitrary interval of length t. Then N_t has a Poisson distribution of mean λt, i.e.

$$\text{prob}(N_t = r) = \frac{(\lambda t)^r e^{-\lambda t}}{r!} \quad (r = 0, 1, \ldots). \tag{1}$$

Further it is clear from property (*c*) in the definition that the numbers of events in non-overlapping time intervals are independent. Thus if we consider $\{N_{0,t}\}$, the number of events in $(0, t]$, as a stochastic process, i.e. as a random function of time, it has the property of having independent increments in non-overlapping time intervals.

To record the main properties of the Poisson distribution (Cramér, 1946, p. 203), let Z be a random variable with the distribution

$$\text{prob}(Z = r) = \frac{\mu^r e^{-\mu}}{r!} \quad (r = 0, 1, \ldots); \tag{2}$$

the corresponding probability generating function is

$$\sum_{r=0}^{\infty} \zeta^r \text{prob}(Z = r) = e^{\mu(\zeta - 1)}.$$

Then if E and var denote, respectively, expectation and variance,

$$E(Z) = \text{var}(Z) = \mu \tag{3}$$

and more generally all the semi-invariants of the distribution are equal to μ.

It easily follows from (3) with $\mu = \lambda t$ that N_t/t converges in probability to λ as $t \to \infty$, justifying the name *rate of occurrence* proposed for the parameter λ of the Poisson process.

The additive property of the Poisson distribution is that if $Z_1, \ldots, Z_m$ are mutually independent and follow Poisson distributions with means $\mu_1, \ldots, \mu_m$, then $\sum Z_i$ follows a Poisson distribution of mean $\sum \mu_i$. This follows easily from the probability generating function. The meaning in terms of a Poisson process is clear; if $\mu_i = \lambda t_i$, the random variable Z_i is distributed as the number of events in an interval of length t_i and the summed random variable has the distribution of the number of events in a single interval of length $\sum t_i$.

As μ increases, the Poisson distribution is asymptotically normal with mean and variance μ. That is, if $Z_{(\mu)}$ denotes a Poisson random variable of mean μ, then for large μ

$$\text{prob}(Z_{(\mu)} \leqslant r) \simeq \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\frac{r+\frac{1}{2}-\mu}{\sqrt{\mu}}} e^{-\frac{1}{2}t^2} dt \tag{4}$$

$$= \Phi\left(\frac{r + \frac{1}{2} - \mu}{\sqrt{\mu}}\right),$$

say. This can be proved directly, or is a consequence of the central limit theorem; the inclusion of the continuity correction of $\frac{1}{2}$ on the left-hand side is not necessary for the truth of (4), but improves the numerical approximation for moderate μ.

Table 2.1. *Adequacy of the normal approximation to the Poisson distribution*

Poisson mean, μ	Observed value, r	prob($Z_{(\mu)} \leqslant r$) Exact	Normal approx.
5	0	0·0067	0·0221
	1	0·0404	0·0588
	9	0·9682	0·9799
	11	0·9946	0·9982
10	3	0·0103	0·0200
	5	0·0671	0·0774
	15	0·9513	0·9590
	18	0·9928	0·9964
20	10	0·0108	0·0168
	13	0·0661	0·0731
	27	0·9475	0·9532
	31	0·9919	0·9949
50	34	0·0108	0·0142
	38	0·0474	0·0519
	61	0·9443	0·9481
	67	0·9911	0·9933

Table 2.1 assesses the accuracy of the normal approximation (4) to the distribution function. For each value of μ, values are given near to lower and upper 1 and 5 per cent points. The table gives the exact cumulative probability and the corresponding normal approximation,

with continuity correction. For many problems connected with significance tests and confidence intervals, the normal approximation is quite accurate enough even below $\mu = 10$. The main source of error in the approximation arises from the skewness of the Poisson distribution.

There are a number of tables both of the probabilities (2) and of the corresponding cumulative probabilities. One of the most accessible sets is in the *Biometrika* tables (Pearson and Hartley, 1956, Tables 7 and 39) and one of the more extensive is that of the General Electric Co. (1962).

(ii) *Distribution of intervals between events*

A second important group of properties of the Poisson process concern the intervals between events. Let X be the interval from the time origin to the first event. We can obtain the distribution of X from first principles (see Exercises) or derive it from (1). For no event occurs in $(0, x]$ if and only if $X > x$. Hence

$$\text{prob}(X > x) = \text{prob}(N_x = 0) = e^{-\lambda x}. \tag{5}$$

Thus $F_X(x)$, the distribution function (d.f.) of X, and $f_X(x)$, the probability density function (p.d.f.), are

$$F_X(x) = 1 - e^{-\lambda x} \quad (x \geqslant 0) \tag{6}$$

$$f_X(x) = F'_X(x) = \lambda e^{-\lambda x} \quad (x \geqslant 0). \tag{7}$$

We call this the *exponential distribution* of parameter λ.

The Laplace-Stieltjes transform of the distribution is

$$E(e^{-sX}) = \int_0^\infty e^{-sx} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda + s}. \tag{8}$$

It follows from (8), considered as a moment generating function, or directly, that

$$E(X) = 1/\lambda, \quad \text{var}(X) = 1/\lambda^2. \tag{9}$$

A very important point is that because the occurrences in any section of a Poisson process are independent of the preceding sections of the process, the origin from which X is measured may be defined in a variety of ways. Thus X may be

(*a*) the time from the original time origin to the first event;

(*b*) the time from any fixed time point to the next event;

(*c*) the time from any event to the next succeeding event, i.e. the interval between successive events;

(*d*) the time from any point t' determined by the pattern of events in $(0, t']$ to the next event.

Further if $X_1, X_2, \ldots$ are the intervals between the origin and the first event, between the first and second events, and so on, the random variables $X_1, X_2, \ldots$ are mutually independent and each with the p.d.f. $\lambda e^{-\lambda x}$. In fact the Poisson process can be defined by this property.

To stress the wide range of applicability of these results it is useful to consider two Examples.

Example 2.4. *Free path length in kinetic theory*. The exponential distribution seems first to have been considered by Clausius in connection with the kinetic theory of gases; see, for example, Jeans (1925, Section 347). Consider a molecule with given speed and let an event be a collision with another molecule. It is reasonable to postulate a constant probability rate of occurrence of events per unit time. Therefore the time between collisions, and hence the free path length of molecules of a given speed, follows an exponential distribution. The unconditional distribution of free path length can be derived when the dependence on velocity of the probability rate of occurrence is known, and this can in fact be calculated.

Example 2.5. *Length of jute fibres*. Dr H. P. Stout (personal communication) has discussed the frequency distribution after processing of the length of jute fibres. This is not in the first place an example of a series of point events in space or time, but is closely related to such a series for the following reason. Jute fibres are initially long and are repeatedly broken during processing. The lengths of the final pieces are therefore determined by the intervals between successive 'weak points' in the initial long fibres. If these 'weak points' occur approximately randomly, the final distribution of fibre length should be nearly that of the intervals between successive events in a Poisson process. Dr Stout showed empirically that the observed frequency distribution is close to the exponential form, except that in some cases the experimental curve falls slightly below the exponential for small lengths. Kolmogorov (see Aitchison and Brown, 1957, p. 101) has given a different stochastic model of breakage leading to a log normal distribution. In applications a log normal distribution with high dispersion and an exponential distribution are quite hard to distinguish.

To continue the general discussion of the distribution of intervals in a Poisson process, let T_{n_0} be the time from the origin to the n_0th following event, where n_0 is a fixed integer. In this notation, the random variable of (5) is T_1. Now T_{n_0} is the sum of n_0 independent random variables each exponentially distributed, so that the Laplace-Stieltjes transform of the distribution of T_{n_0} is, from (8), $\lambda^{n_0}/(\lambda+s)^{n_0}$, which corresponds to the p.d.f.

$$\frac{\lambda(\lambda t)^{n_0-1}e^{-\lambda t}}{(n_0-1)!} \quad (t \geqslant 0), \tag{10}$$

reducing to (7) when $n_0 = 1$. For numerical work it is useful to note from (10) that $2\lambda T_{n_0}$ has the chi-squared distribution with $2n_0$ degrees of freedom; symbolically

$$2\lambda T_{n_0} = \chi^2_{2n_0}. \tag{11}$$

An important relation between the Poisson distribution and the chi-squared distribution is obtained by noting that the events $\{N_t < n_0\}$ and $\{T_{n_0} > t\}$ are identical and therefore have the same probability. Thus

$$\sum_{r=0}^{n_0-1} \frac{(\lambda t)^r e^{-\lambda t}}{r!} = \text{prob}(\chi^2_{2n_0} > 2\lambda t),$$

or

$$\sum_{r=0}^{n_0-1} \frac{\mu^r e^{-\mu}}{r!} = \int_{\mu}^{\infty} \frac{v^{n_0-1}e^{-v}}{(n_0-1)!}dv. \tag{12}$$

Equation (12) is easily proved directly by integrating by parts. The type of probabilistic argument leading to (12) will be used again in Chapter 4.

The main properties of the random variable T_{n_0} follow from (11), or directly from the p.d.f. (10), or from the definition of T_{n_0} as the sum of n_0 random variables. Thus

$$E(T_{n_0}) = n_0/\lambda, \quad \text{var}(T_{n_0}) = n_0/\lambda^2, \tag{13}$$

and standard tables of the chi-squared distribution can be used to find the percentage points of the distribution.

For large n_0, T_{n_0} is, by the central limit theorem, asymptotically normally distributed. Thus $(\lambda T_{n_0} - n_0)/\sqrt{n_0}$ can be taken as having a standardized normal distribution or, to a closer approximation (Cramér, 1946, p. 251), $\sqrt{(\lambda T_{n_0})}$ can be taken as normally distributed with mean $\sqrt{(n_0 - \frac{1}{4})}$ and variance $\frac{1}{4}$.

For some methods of statistical analysis, especially those of Section 3.2(i), it is useful to work with $\log T_{n_0}$. By (11),

$$\log T_{n_0} = -\log(2\lambda) + \log(\chi^2_{2n_0}),$$

so that the results of Bartlett and Kendall (1946) on the log chi-squared distribution can be used. In fact from (10) we have that the semi-invariant generating function of $\log T_{n_0}$ is

$$\log E\{\exp(i\theta \log T_{n_0})\} = \log\left\{\int_0^\infty t^{i\theta}\frac{\lambda(\lambda t)^{n_0-1}e^{-\lambda t}}{(n_0-1)!}dt\right\}$$
$$= \log\left\{\frac{\Gamma(n_0+i\theta)}{\Gamma(n_0)}\right\} - i\theta\log\lambda,$$

where the Gamma function is defined by

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt.$$

Thus the semi-invariants of $\log T_{n_0}$ are given by the derivatives of the log Gamma function of argument n_0 and, in particular, taking the coefficients of $i\theta$ and of $(i\theta)^2/2!$, we have that

$$E(\log T_{n_0}) = \psi(n_0) - \log\lambda, \quad \text{var}(\log T_{n_0}) = \psi'(n_0), \tag{14}$$

where

$$\psi(x) = \frac{d\log\Gamma(x)}{dx}$$

is the digamma function (Davis, 1933, Table 9).

Bartlett and Kendall (1946) tabulate the first four semi-invariants of the log chi-squared distribution and give graphs of the p.d.f. The distribution has a moderate negative skewness and tends rather slowly