New Statistical Method For Describing Highway Distribution of Cars

FRANK A. HAIGHT, BERTRAM F. WHISLER AND WALTER W. MOSHER, JR.,
Respectively, Associate Research Mathematician, Assistant Research Engineer, and Junior Research Engineer, Institute of Transportation and Traffic Engineering, University of California, Los Angeles

The distribution of cars on a road is classified as either (1) random, (2) equally spaced (regular), or (3) intermediate. In the first, the random arrangement may be represented by either one of two statistical distributions: (a) the Poisson distribution, which is called the "counting" distribution for the random case, or (b) the negative exponential distribution, which is the "gap" distribution for the random case. In the intermediate case the corresponding gap distribution is the Pearson Type III (also called gamma, or Erlang) and the corresponding counting distribution, which has been called the generalized Poisson distribution, is the one discussed.

The generalized Poisson distribution corresponds to Type III gaps. Tables and monographs are given which will aid in fitting data to the generalized Poisson formula, and examples of traffic data analysis by the method are presented.

• THERE are two ways to count vehicles: (a) by means of some apparatus which records the time of arrival of each vehicle at a fixed point, and (b) by some device such as aerial photography which records at a fixed time the spatial arrangement of the vehicles. In the first case, the time distribution is obtained and in the second the space distribution.

One measurement derived from the time or space distribution is the gap between vehicles. In the first case the gap is the time between the arrival of consecutive cars. For the space distribution, it is the space gap, defined as the linear distance between corresponding parts of consecutive cars.

The method of analysis presented in this paper can be applied equally to either of these two cases. To avoid confusion in terminology, the language of the time distribution is used almost exclusively, but application of this technique to the space distribution is equally valid.

There are various possible arrangements of dots on a line. The instants when cars pass a given point may be considered to be dots on the time axis. If the dots are placed independently of each other, so that any point of the line is equally likely to have a dot, the arrangement is called random.

Randomness is an extreme arrangement, and one which has been thoroughly studied. It seems to be realistic if the dots correspond to instants of radioactive decay, or of placing telephone calls, but, for reasons which will be explained subsequently, it does not exactly describe automobile traffic. The opposite extreme from randomness is regularity, in which the dots are equally spaced. Regularity is also not observed in traffic, although it may be approached during very heavy congestion.

Thus, there are limiting conditions: for extremely light traffic, the instants of arrival are nearly random; for very heavy traffic, the instants of arrival are nearly regular. The range of possibilities between randomness and regularity will be considered.

COUNTING AND GAP DISTRIBUTIONS

When any arrangement of dots is given, it can be described statistically in a variety of ways. Two methods of description, the "gap" and "counting" distributions, are used particularly. In the regular case, each gap is exactly the same length, and therefore, the gap distribution is deterministic,

$$f(x) = \begin{cases} 0, & x \neq g \\ 1, & x = g \end{cases} \tag{1}$$

where g is the length of gap. The gap distribution for the random case is the well known negative exponential

$$f(x) = (1/g)e^{-x/g}, \quad 0 < x < \infty \tag{2}$$

where g is the average gap length.

An equally exact way of characterizing an arrangement is by its counting distribution. In a number of time intervals of length T, the number of events in a time interval may be 0, 1, 2, ... In the random case, the Poisson distribution gives the probability of n dots in time T:

$$p_n = e^{-(T/g)}(T/g)^n/n! \tag{3}$$

Since an actual counting device usually leads to this distribution, it is called the counting distribution. In the regular case, the counting distribution is not so well known. If the time interval T contains at least N of the gaps of fixed length g, but does not contain $N+1$ of them, then any time interval, whenever taken, must contain either N dots, or else $N + 1$ dots, but cannot contain any other number. The probability of the latter is $(T—Ng)/g$, and therefore the counting distribution in the regular case is

$$p_n = \begin{cases} \dfrac{T-Ng}{g}, & n = N+1 \\ 1 - \dfrac{T-Ng}{g}, & n = N \end{cases} \tag{4}$$

The following points in Eqs. 1-4 should be considered:

1. There is an exact correspondence between Eq. 1 and Eq. 4, in that they are different methods of describing the identical situation, which is called regularity. Similarly Eq. 2 and Eq. 3, although mathematically different, describe the same physical situation, randomness.
2. The counting distribution is always discrete, defined over positive integers and zero; whereas the gap distribution is always continuous, defined over the real numbers from zero to infinity. (Eq. 1 is a degenerate form of this, since the whole probability is concentrated at one point.)
3. The two situations so far described are both extremes, which might never occur in practice.
4. The relationship between a given gap distribution and its corresponding counting distribution is not obvious and requires mathematical proof. In the random case, the proof is well known (*1*). In the regular case, the proof is based on straightforward reasoning consisting essentially of the argument given after Eq. 3 (*2*). A general method which

relates any counting distribution to its corresponding gap distribution, and *vice versa,* is given in Appendix A.

5. Both the counting and gap distributions are useful: the counting is convenient in practical work; the gap is easy to generalize to arrangements intermediate between regularity and randomness.

THE INTERMEDIATE CASE

Although the Poisson distribution has been useful in describing vehicular traffic, a careful analysis of the corresponding gap distribution shows that it cannot be theoretically correct, even in the case of very light traffic. Eq. 2 represents a curve which is highest at the origin, and declines gradually as x goes to infinity. This property does not agree with vehicular traffic conditions, since cars cannot arrive too near to the same time without traveling at very high speeds. There is a basic gap (the time for a very fast car to cross) which must be maintained. Although this gap may be small for high speed cars, it is not correct to say that the smaller the gap, the more likely it is to occur. On the contrary, very small gaps, although theoretically possible, should have low probability, and as the length of the gap approaches zero, so should the probability. In Eq. 2 this is not the case, since the value of the function at the origin is $1/g$. Gerlough (*3*) has suggested a remedy for this in translating the negative exponential curve away from the origin by some small amount. Another suggestion (*4*) is based on the model of a particle counter in which the mechanism is inoperative for a very short "dead time" after each registration. Both of these suggestions require small gaps to be impossible rather than merely improbable, and therefore do not adequately meet the objection.

To overcome these difficulties, a sensible gap distribution for the intermediate case should conform to the following standards: (a) be defined over $(0, \infty)$, (b) approach zero as the gap approaches zero, and (c) contain an extra parameter, which will measure the extent of randomness or regularity, and which will yield Eqs. 1 and 2 for extreme values of the parameter. The distribution which fulfills these conditions is called the Pearson Type III, and is formed by multiplying the function $e^{-x/g}$ in Eq. 2 by some appropriate power of x.

$$f(x) = \frac{\lambda^k e^{-\lambda x} x^{k-1}}{(k-1)!} \qquad (5)$$

If $k=1$, this becomes Eq. 2, with $\lambda=1/g$, and corresponds to random arrangement, or, in terms of vehicles, to very light traffic. As k approaches infinity, Eq. 5 approaches Eq. 1, although this fact is not proved. Hence a very large value of k in the Pearson Type III distribution of gap will correspond to very heavy traffic. The mean value of the Type III distribution is $k/\lambda=g$. The reciprocal λ is slightly more convenient than g even in the random case; therefore Eq. 2 is more frequently written

$$\lambda \exp\ (-\lambda x)$$

A formula for the (possible) distribution of cars in cases between randomness and regularity is proposed, but it is now necessary to see how well it corresponds to reality. It is useful, if not absolutely essential, to find the counting distribution which corresponds to the gap distribution Eq. 5. This function, which is intermediate between Eqs. 1 and 4, was apparently discovered by Goodman (*5*) and named by him the generalized Poisson distribution.

$$p_n = \sum_{i=1}^{k} \frac{e^{-\lambda T}(\lambda T)^{nk+i-1}}{(nk+i-1)!} \qquad (6a)$$

or

$$p_n = \sum_{j=nk}^{(n+1)k-1} \frac{e^{-\lambda T}(\lambda T)^j}{j!} \qquad (6b)$$

Appendix B gives a proof of this fact. A more complete treatment of the generalized Poisson distribution has been published previously (*6*).

Eq. 6b has a quite simple interpretation. It says that the probability of no cars in the time interval T is the sum of the first k terms of some Poisson series, that the probability of one car is the sum of the next k terms of the same Poisson series, etc. It is clear that the total probability is one, since it is the sum of a whole Poisson series. The question remains: which Poisson series to use? This is a problem in parameter estimation, which is discussed subsequently.

A simple example of the generalized Poisson distribution may be useful. The Poisson tables of Kitagawa (*7*) give the following values for $\lambda=1/2$

n	p_n	
0	0.6065	
1	0.3033	
2	0.0756	
3	0.0126	Mean value = 0.5000
4	0.0016	
$\geqq 5$	0.0002	

For $k=2$, this table would yield

n	p_n	
0	0.9098	
1	0.0884	Mean value = 0.0920
$\geqq 2$	0.0018	

So far the parameter k has appeared as an integer, and certainly this is necessary for Eq. 6b to make sense. However, there is an equivalent form (see Appendix C) for the generalized Poisson distribution in which fractional values of k are meaningful. One other interesting consequence of the alternative form is that k may be less than unity. Such a value would correspond to a situation "beyond" Poisson, which might be called hyper-random. No traffic situation for which such a value of k is appropriate has yet been found, but there seems to be some evidence that a multi-lane freeway with very high traffic volume might be hyper-random.

Considering the fact that $k=1$ is equivalent to randomness and $k=\infty$ to regularity, it seems reasonable to define

$$\text{coefficient of randomness} = \frac{1}{k} \;(=\beta)$$

so that $\beta=1$ is for perfect randomness and $\beta=0$ for perfect regularity. Thus β measures, on a scale between zero and one, the degree of randomness in the traffic situation. In terms of this coefficient, hyper-randomness occurs for values of β exceeding unity.

ESTIMATION OF PARAMETERS

In applying the generalized Poisson to specific data, it is first necessary to determine λ and k. Formulas for the mean and variance in terms of λ and k have been obtained (*2, 6*), but so far it has proved impossible to reverse these expressions. However, if k has been chosen, the approximate formula for λ,

$$\lambda = mk + \frac{1}{2}(k-1)$$

where m is the sample mean, has proved fairly accurate.

The nomograph (Fig. 1) has been computed with a view to making the choice of λ and k easy. To use it, it is necessary to compute the sample mean and variance, locate the corresponding coordinate point on the nomograph, and read off appropriate values of λ and k from the double family of curves. Whether or not integral values of k are chosen may depend on whether the added accu-

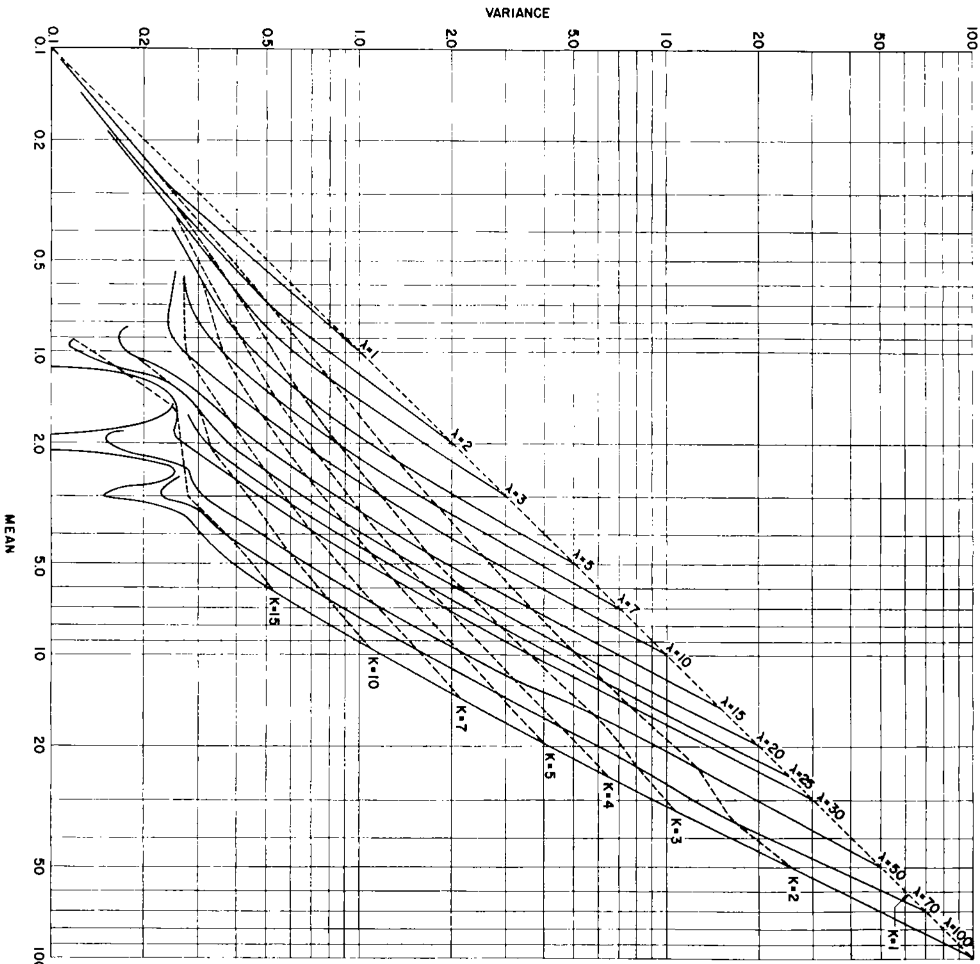

Figure 1. Nomograph of $\lambda = mk + \frac{1}{2}(k-1)$.

racy is worth the time required to use tables of the incomplete gamma function, as explained in Appendix C.

COMPUTER INPUTS FOR THE INTERMEDIATE CASE

Whenever an attempt to simulate a real time traffic situation is undertaken, the important requirement of providing traffic inputs to the simulator exists. There are two approaches to the satisfaction of this requirement: (a) supply empirical input data directly to the simulation, (b) generate input within the simulator according to some known distribution.

The first method is satisfactory in many cases if the simulator is a relatively slow device; however, when large masses of input data are required the storage and handling by this method become too time consuming and costly. In such cases, particularly when a general purpose digital computer is the simulator, the second method is superior.

Gerlough (*3*) has shown two meth-

TABLE 1

OBSERVED AND THEORETICAL VALUES FOR SHIRLEY HIGHWAY DATA

Time	6:00		6:15		6:30		6:45		7:00		7:15		7:30		7:45	
	O	T	O	T	O	T	O	T	O	T	O	T	O	T	O	T
$n = 0$	80	80	67	68	41	37	14	9	4	2	2	1	1	1	0	0
$n = 1$	18	18	29	26	28	37	26	34	19	23	18	20	15	15	10	11
$n = 2$	2	2	3	5	23	18	33	35	42	41	43	48	45	45	61	60
$n = 3$	0	0	1	1	7	6	25	16	29	25	35	26	32	31	28	27
$n > 3$	0	0	0	0	1	2	2	6	6	9	2	5	7	8	1	2
Mean	0.22		0.38		0.99		1.75		2.14		2.17		2.29		2.20	
Variance	0.21		0.36		1.01		1.09		0.86		0.66		0.71		0.38	
k	1		1		1		2		3		4		4		8	
λ	0.22		0.38		0.99		4.0		7.4		10.1		10.7		21.0	
Cars/hr	220		380		990		1750		2140		2170		2250		2200	

Time	8:00		8:15		8:30		8:45		9:00		9:15		9:30		9:45		10:00	
	O	T	O	T	O	T	O	T	O	T	O	T	O	T	O	T	O	T
$n = 0$	0	0	0	1	15	12	10	9	23	28	31	34	46	51	46	48	48	50
$n = 1$	18	19	12	16	31	38	28	33	40	36	42	37	41	34	39	35	36	35
$n = 2$	54	56	54	46	37	33	37	35	24	23	18	20	12	12	11	13	15	12
$n = 3$	27	23	30	30	16	13	22	17	13	13	7	7	1	3	4	4	1	3
$n > 3$	1	2	4	7	1	4	3	6	0	0	2	2	0	0	0	0	0	0
Mean	2.11		2.26		1.57		1.80		1.27		1.07		0.68		0.73		0.69	
Variance	0.48		0.51		0.92		0.98		0.92		1.04		0.52		0.66		0.37	
k	6		4		2		2		1		1		1		1		1	
λ	15.0		10.6		3.6		4.1		1.27		1.07		0.68		0.73		0.69	
Cars/hr	2110		2260		1570		1800		1270		1070		680		730		690	

ods for providing computer input in the random case, one based on the gap distribution (negative exponential) and one based on the counting distributing (Poisson). His procedure based on the Poisson will work equally well for the generalized Poisson. It involves the use of cumulative Poisson sums. However, the cumulative generalized Poisson sums are also cumulative Poisson sums, beginning at a different index. For example, the value of the generalized Poisson sum from n to infinity (with parameters λ and k) is just the Poisson sum with parameter λ) from nk to infinity. The procedure is one of trial and error, and can be outlined briefly as follows: First generate a random fraction R. Then begin to form the cumulative generalized Poisson term by term, comparing it to R after each term is added. When the sum first exceeds R, the corresponding index of summation is the desired variate.

SHIRLEY HIGHWAY DATA ANALYSIS

The data presented here were collected by the Bureau of Public Roads on April 2, 1958. The data were taken on northbound lane number two of the Shirley Highway at a point 1400 ft north of the Glebe Road on-ramp between 6 A.M. and 10:30 A.M. The time (to nearest hundredths of a second) of each vehicle crossing the observation point was recorded on adding machine tape.

The sample size chosen was a 6-min time interval beginning on the quarter hour, and the counting distribution was based on the value $T=1/10$ sec. The results of generalized Poisson fitting is shown in Table 1; it was possible to fit sixteen of the seventeen samples.

At the time that this analysis was performed, the nomograph was not yet prepared. Consequently the (integer) values of k were obtained by trial and error, and the correspond-

TABLE 2
OBSERVED AND THEORETICAL VALUES FOR FRACTIONAL k

	Data (7:00 sample)	Theoretical Curve (λ = 7.45, k = 3)	Theoretical Curve (λ = 7.95, k = 3.2)
$n = 0$	4	2	2
$n = 1$	19	23	22
$n = 2$	42	41	44
$n = 3$	29	25	25
$n > 3$	6	9	7
Mean	2.14	2.16	2.13
Variance	0.86	0.89	0.81
Chi-square		4.36	3.28

ing value of λ by the approximate formula (Eq. 7).

In one case the trial and error method was extended to fractional values of k, with some success as indicated in Table 2.

It will be seen (Fig. 1) that the parameter values used do not depart substantially from those indicated.

ACKNOWLEDGMENT

The calculations on which the nomograph (Fig. 1) was based were performed at Western Data Processing Center on the Los Angeles campus of the University of California. The authors wish to express their appreciation for the use of this facility. The basic data used consisted of Molina's tables (*9*) on punched cards, obtained through the courtesy of the SWAC Laboratory of the University of California.

REFERENCES

1. GERLOUGH, D. L., "The Use of of the Poisson Distribution in Highway Traffic." in Gerlough, D. L., and Schuhl, A., "Poisson and Traffic," ENO Foundation, Saugatuck, Conn. (1955).
2. WHISLER, BERTRAM F., "A Study of Certain Probability Models which Aid in Classification of Sequences of Events into Non-Random, Partially Random, or Completely Random Categories." M.A. Thesis (1960), The American University, Washington, D.C.
3. GERLOUGH, D. L., "Traffic Inputs for Simulation on a Digital Computer." *HRB Proc.,* 38: 480-492 (1959).
4. HAIGHT, F. A., "Towards a Unified Theory of Road Traffic." *Operations Research,* 6: 6, pp. 813-826 (Nov.-Dec. 1958).
5. GOODMAN, LEO A., "On the Poisson - Gamma Distribution Problem." *Ann. Inst. Statist. Math.* (Tokyo) 3:123-125 (1952).
6. HAIGHT, F. A., "The Generalized Poisson Distribution." *Ann. Inst. Statist. Math.* (Tokyo) 11: 2, pp. 101-105 (1959).
7. KITAGAWA, T., "Tables of Poisson Distribution." Baifukan, Tokyo, Japan (1952).
8. KENDALL, M. G., "The Advanced Theory of Statistics." Vol. 1, Griffin and Co., London (1943).
9. MOLINA, E. C., "Tables of Poisson's Exponential Limit." (1942).
10. PEARSON, KARL, "Tables of the Incomplete Γ Function." Cambridge University Press (1922).

APPENDIX A

RELATIONSHIP BETWEEN COUNTING AND GAP DISTRIBUTIONS

Let $p_t(n) = \text{Prob}$ (n events in time t), $n = 0, 1, 2, \cdots$ have the cumulative form

$$P_t(n) = \text{Prob}(\leqq n \text{ events in time } t), \quad n = 0, 1, 2, \cdots$$

If the density function for the distribution of lengths of individual gaps is $f(x)$, then the density function for the distribution of lengths of the sum of n gaps is known as the n-fold convolution of $f(x)$, and is denoted by $f^{n*}(x)$. This distribution is defined and explained in any standard textbook of statistics.

Now,

Prob(number of events in time t is $<n$)

$$= \text{Prob(sum of } n \text{ gaps} \geqq t)$$

which means

$$P_t(n-1) = \int_t^\infty f^{n*}(x)\,dx.$$

This gives the counting distribution in terms of the gap distribution. To obtain the reverse, it is only necessary to set $n=1$ in the preceding formula and differentiate with respect to t:

$$f(x) = -\frac{\partial}{\partial t} p_t(0) \Big|_{t=x}$$

APPENDIX B

DERIVATION OF GENERALIZED POISSON DISTRIBUTION

Let the gap distribution be Type III with parameters λ and k

$$f(x) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!}.$$

Then it is well known (see, for example, Kendall (*8*, p. 244)) that the n-fold convolution is Type III with parameters λ and nk.

$$p_t(n) = P_t(n) - P_t(n-1)$$
$$= \int_t^\infty f^{(n+1)*}(x)\,dx - \int_t^\infty f^{n*}(x)\,dx$$

and since the integral over the whole range $(0, \infty)$ is unity,

$$= \int_0^t f^{n*}(x)\,dx - \int_0^t f^{(n+1)*}(x)\,dx.$$

Writing $y = \lambda x$, with the convolution values put in,

$$= \int_0^{\lambda t} \frac{y^{nk-1}e^{-y}}{(nk)!}dy - \int_0^{\lambda t} \frac{y^{(n+1)k-1}e^{-y}}{[(n+1)k]!}$$

Now integrate by parts the second integral k times. Poisson terms will arise from each reduction, finally leaving an integral which cancels with the first integral. The remaining terms give exactly Eq. 6.

APPENDIX C

ALTERNATE FORM OF GENERALIZED POISSON DISTRIBUTION

The incomplete gamma function is defined by

$$\Gamma(n, z) = \int_z^\infty e^{-t} t^{n-1}\,dt.$$

Therefore, from Appendix B,

$$p_t(n) = \frac{\Gamma[(n+1)k, \lambda t]}{\Gamma[(n+1)k]} - \frac{\Gamma(nk, \lambda t)}{\Gamma(nk)}.$$

But the incomplete gamma function is defined for all values of n whether integral or not, and therefore, values of $p(n)$ can be found from tables of this function.