# VEHICULAR HEADWAY DISTRIBUTIONS: TESTING AND RESULTS

J. E. Tolle, Transportation Research Institute, Carnegie-Mellon University

Numerous investigations and studies have been performed concerning the spacings and headways of vehicles in the traffic stream. This paper briefly reviews, examines, and tests some of the mathematical models with real world data collected for both two-lane roadways and urban freeways. These testing methods include both graphical fits and statistical fits, and the results are presented, along with conclusions on the appropriateness of the model representations. The results indicate the composite exponential, Pearson Type III, and log-normal distributions best represent conditions; generally the log-normal distributions are the best for a wide range of traffic volumes.

•THE SPACINGS and headways of vehicles in the traffic stream have been the concern of many investigations and studies. Normally spacing refers to the physical distance between vehicles in feet (meters), and headway is the time between successive vehicles in seconds. There have been numerous attempts to mathematically describe vehicle spacing and headways in the traffic stream. This paper examines these models and presents the results obtained by using the log-normal distribution.

Before attempting to develop a model, one should ask why a mathematical description of headways is desired. One reason is for input to a simulation model of traffic flow on a digital computer such as an intersection, car-following, or other simulation. If such a simulation model exists, then the problem of generating traffic data as input arises. At least two methods of solving this problem exist. The first is the reading in of actual data, which has two disadvantages:

1. The reading in of data consumes a lot of computer and data collection time, and
2. Because only actual observed volumes are used, one benefit of simulation, that of investigating situations that are extremely difficult to observe in the real world, is lost.

The second method eliminates these disadvantages because it allows the computer to generate its own data. The problem of time is then solved, since computer generation of data requires only a fraction of the required read-in time. Also, situations that are difficult to observe in the real world can be investigated. The one problem with internal generation of data is that a mathematical model is needed with which to accurately generate data that agree with the real world situation.

## PREVIOUSLY PROPOSED DISTRIBUTIONS

The Poisson distribution has been successfully used to obtain the arriving rate of vehicles. Under conditions of free flow, i.e., most vehicles are able to choose their desired speed, this gives satisfactory results.

Assume that free-flow conditions exist, and consider the probability of x vehicles

Publication of this paper sponsored by Committee on Traffic Flow Theory and Characteristics.

arriving per t time interval. This is known as the arriving rate of vehicles. If the hourly volume Q is known, then the Poisson probability function is

$$P(x) = \begin{cases} \exp(-m_m x)/x! & \text{for } x = 0, 1, \ldots \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

This is a discrete distribution, i.e., it takes on different values only at the integer points. It is appropriate at this time to recall the differences between counting distributions and time headway distributions. The counting distribution is always discrete, and the time headway distribution is continuous. This is clear from the fact that only a whole number of vehicles may arrive in a given time interval, but a time gap may be a fractional value.

If we write the Poisson function in the form

$$P(x) = \frac{1}{x!}\left(\frac{Qt}{3{,}600}\right)^x \exp(QT/3{,}600) \tag{2}$$

where t is the time interval to be considered in seconds and Qt/3,600 is the average number of vehicles per t seconds, this corresponds to the m-value that represented the mean value in equation 1. Then P(x) is the probability of x vehicles arriving during the given time interval t.

One of the first distributions tested for headways was the negative exponential, which may be derived from the Poisson by considering the probability of zero arrivals in the time interval. If $\lambda > 0$,

$$P(t) = \begin{cases} \lambda \exp(-\lambda t) & \text{for } t > 0 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

Then the probability of a gap greater than or equal to T seconds is the integral from T to infinity of the negative exponential; i.e., for $T > 0$,

$$P(t > T) = \exp(-\lambda t) = \exp(-QT/3{,}600) = e \tag{4}$$

The negative exponential cannot be satisfactory because it assigns a higher probability to very low headways, and as the time approaches zero the probability increases. From observed results this is not true; however, there is normally a peak number of headways around 0.5 to 2 sec, and the probability of a zero headway is by necessity zero.

Gerlough (5) suggested a method of possible correction for the negative exponential that consists of shifting the negative exponential away from the origin. The difficulty with that method is that small gaps are impossible, which does not satisfy the known conditions. Equation 5 represents the shifted exponential (shift of a to the right).

$$f(t) = \begin{cases} \exp[-\lambda(t - a)] & \text{for } t > a \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

Other distributions that may be evaluated are the composite exponential and the Pearson Type III.

## COMPOSITE EXPONENTIAL

Schuhl (13) proposed a composite exponential in which vehicles are classified as constrained or free flowing. A constrained vehicle is one that is prevented from passing, and a free-flowing vehicle is able to pass. We then have the following definitions:

$$n = \frac{\text{volume of constrained vehicles}}{\text{total volume}},$$

$$1 - m = \frac{\text{volume of free-flowing vehicles}}{\text{total volume}},$$

$b_1$ = average headway of constrained vehicles = 3,600/number of constrained vehicles,

$b_2$ = average headway of free-flowing vehicles = 3,600/number of free-flowing vehicles,

$\Delta$ = minimum headway required by constrained vehicle, and

$t$ = headway in seconds.

The composite exponential probability distribution function is as follows:

$$P(t) = P_r(g < t) = \begin{cases} (1 - n)(1 - e^{-t/b_2}) + n[-e^{-(t-\Delta)/(b_1-\Delta)}] & \text{for } t > 0.5 \\ 0 & \text{for } t < 0.5 \end{cases} \tag{6}$$

That is, the probability of a headway less than t seconds is $P(t)$.

Intervals of a half second were used except for the last one to test the composite exponential and to apply a chi-square test for goodness of fit. Figures 1 through 6 show the plots of the theoretical and observed distribution functions. Only one statistically good fit was obtained with a volume of 339 vehicles per hour, but graphically the fits for some volumes were fairly close, indicating that there is some merit for considering the composite exponential.

## PEARSON TYPE III

The Pearson Type III distribution may be stated as follows for $x > 0$:

$$f(x) = \frac{\lambda^k e^{-\lambda} x^{k-1}}{\Gamma(k)} \tag{7}$$

If k is an integer, $\Gamma(k) = (k - 1)!$

Equation 7 is the probability density function. If $k = 1$, equation 7 simply reduces to the previously discussed negative exponential distribution. Figure 7 shows the effect of varying k and holding $\lambda$ constant. From equation 7 it is seen that the distribution function is

$$F(x) = P(X < x) = \int_0^x \frac{\lambda^k e^{-\lambda x} x^{k-1}}{\Gamma(k)} dx \tag{8}$$

**Figure 1. Observed and theoretical cumulative headway distributions for volumes of 339 and 442 vph.**

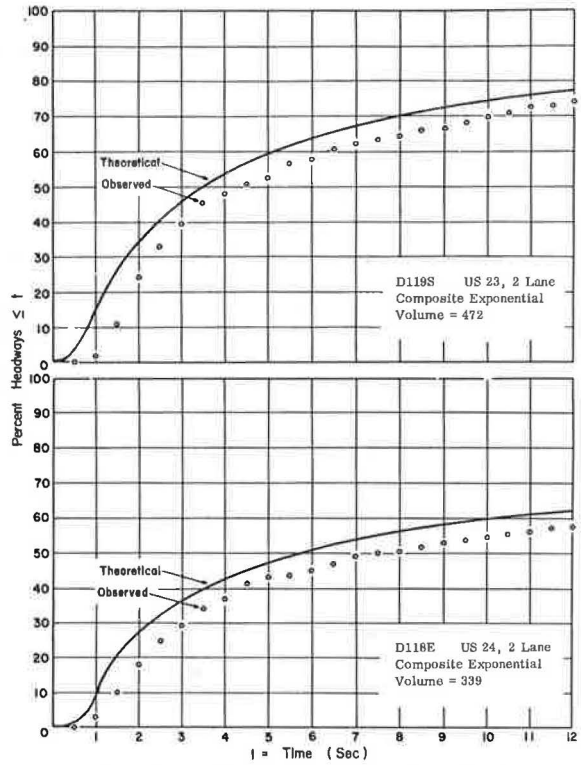



**Figure 2. Observed and theoretical cumulative headway distributions for volumes of 445 and 515 vph.**

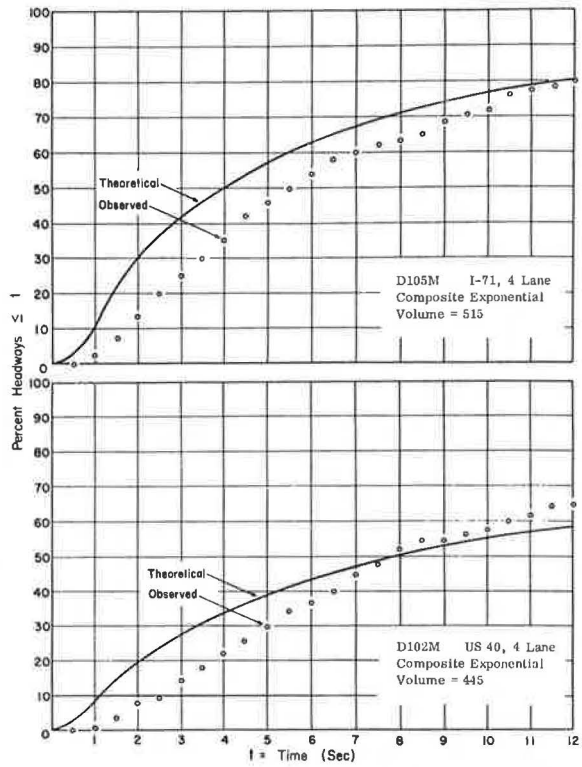


Figure 3. Observed and theoretical cumulative headway distributions for volumes of 551 and 659 vph.

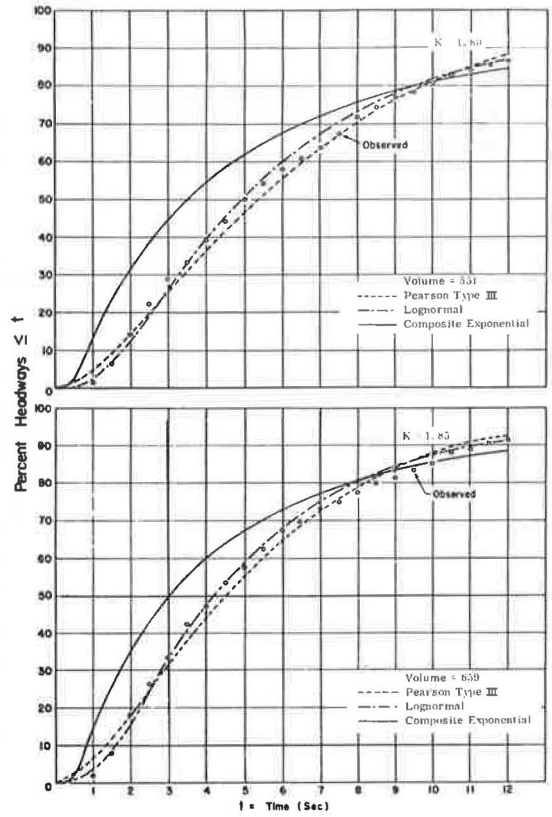



Figure 4. Observed and theoretical cumulative headway distributions for volumes of 781 and 902 vph.

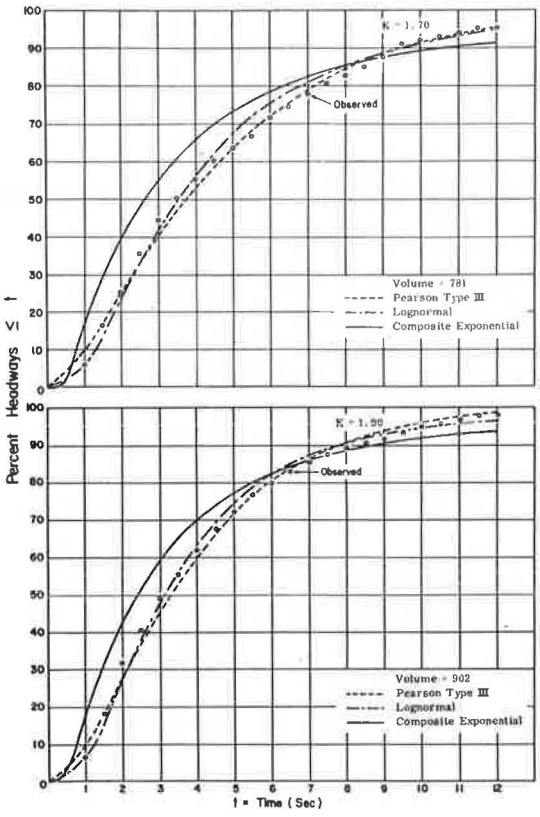
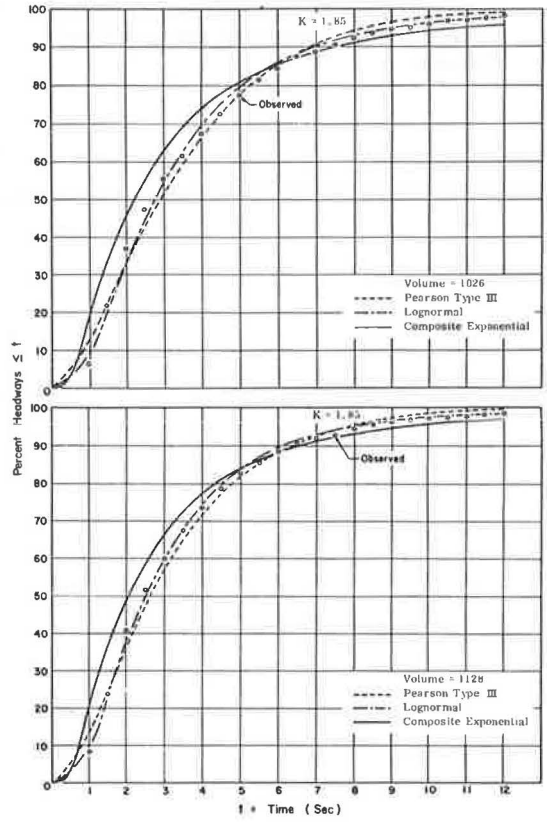



Figure 5. Observed and theoretical cumulative headway distributions for volumes of 1,026 and 1,128 vph.



The reduction of F to a simple formula is easily done for integer values of k. For noninteger k-values, it is necessary to evaluate the gamma function with a table or algorithm and use an approximation for the integral.

In the evaluation of the data collected, the parameter $\lambda$ was set at k divided by the average gap length in seconds. Again a chi-square test was used, but no good statistical fits were found. Graphs of the theoretical and observed distribution functions are very close and an improvement over the composite exponential (Figures 3 through 6). The use of noninteger k did not improve the statistical fits, and these graphs were not plotted.

## LOG-NORMAL DISTRIBUTION

Consider now the log-normal distribution. The log-normal distribution is the distribution of a variate whose logarithm obeys the normal law of probability. A number of names have been given the log-normal distribution such as the Galton-McAlister, Kapteyn, and Gibrat distribution. Although this is a relatively new application of the log normal, its origin dates back to 1879, and it has been used extensively in economic theory.

Consider a random variable X that ranges from zero to infinity, i.e., $0 < X < \infty$. By making the transformation $Y = \lg X$, where lg is the natural logarithm, then Y is again a random variable having an infinite range; that is, Y takes on values between minus and plus infinity, $-\infty < Y < \infty$.

If Y is normally distributed, then by introducing the following notation:

$$L(x|\mu, \sigma^2) = P(X < x) \text{ and } N(y|\mu, \sigma^2) = P(Y < y)$$

the relation becomes

$$N(y|\mu, \sigma^2) = N(\lg x|\mu, \sigma^2)$$

That is, L and N are the distribution functions of X and Y respectively. Because X and Y are related by $L(x) = N(\lg x)$ for $x > 0$,

$$L(x) = \int \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-(\lg t - \mu)^2/2\sigma^2\right] dt \qquad \text{for } x > 0$$

By differentiating L with respect to x

$$\frac{dL}{dx} = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-(\lg x - \mu)^2/2\sigma^2\right] dt \qquad \text{for } x > 0$$

it follows immediately that the probability density function of X is

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-(\lg x - \mu)^2/2\sigma^2\right] \qquad \text{for } x > 0$$

Evaluation of the distribution was again carried out on the computer by using a system subroutine to evaluate the normal distribution. Because Y is normally distributed

Figure 6. Observed and theoretical cumulative headway distributions for volume of 1,369 vph.

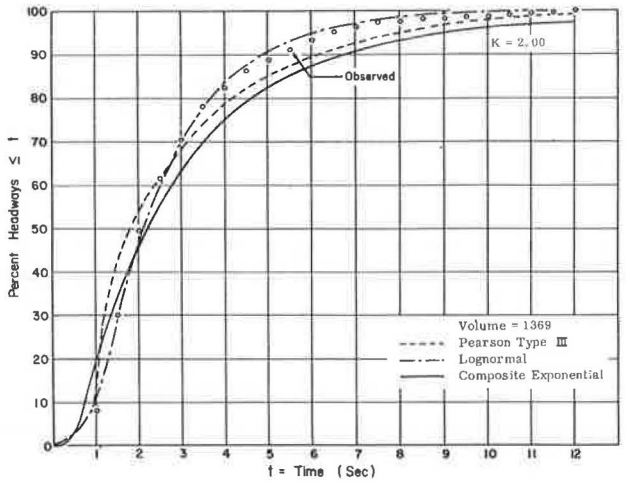



Figure 7. Frequency curves for (a) Type III distribution for various values of K, (b) normal and log-normal distributions, (c) log-normal distribution for different $\mu$, and (d) log-normal distribution for different $\sigma^2$.

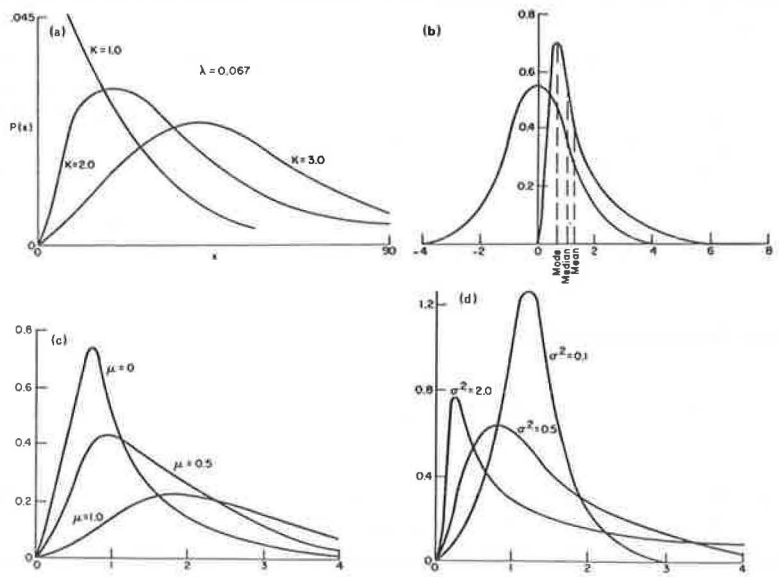



Table 1. Summary of distributions.

| Distribution | Type | Comments | Equation |
|---|---|---|---|
| Poisson | Counting | | $f(x) = \begin{cases} \frac{e^{-m}m^x}{x!} & \text{for } x = 0, 1, \ldots \\ 0 & \text{otherwise} \end{cases}$ |
| Generalized Poisson | Counting | | $P_n = \sum_{i=1}^{k} \frac{e^{-\lambda t}(\lambda t)^{nk+i-1}}{(nk+i-1)!} \quad \text{for } t = 0, 1, \ldots$ |
| Negative exponential | Gap | Unsatisfactory | $f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$ |
| Shifted exponential | Gap | Unsatisfactory | $f(x) = \begin{cases} \lambda e^{-\lambda(x-a)} & \text{for } x \geq a \\ 0 & \text{otherwise} \end{cases}$ |
| Composite exponential | Gap | One out of 11 good fits. Good fit at 515 vph on 4-lane highway using 38 deg of freedom. | $P(g < t) = \begin{cases} (1-n)(1-e^{-t/b_1}) + n\left[1 - e^{-\frac{(t-\Delta)}{b_2-\Delta}}\right] & \text{for } t \geq \Delta \\ 0 & \text{for } t < \Delta \end{cases}$ |
| Pearson Type III | Gap | No good fits. Actual graphic prediction better than implied by no good fits. | $f(x) = \begin{cases} \frac{\lambda^k e^{-\lambda x} x^{k-1}}{(k-1)!} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$ |
| Log normal | Gap | Four out of 7 good fits with selected 5-min data. Two out of 4 good fits with 1 hour of data. Zero out of 2 good fits with 515 vph on 4 lanes and 339 vph on 2 lanes. Graphically predictions all seem close. | $f(x) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases}$ |

with mean $\mu$ and variance $\sigma^2$, $z = (Y - \mu)/\sigma$ has a normal distribution with mean 0 and variance 1. As estimates for the parameters

$$\hat{\mu} = \frac{\sum^{n} \lg x_i}{n}$$

$$\hat{\sigma}^2 = \frac{\sum^{n} (\lg x_i - \hat{\mu})^2}{n - 1}$$

Figure 7 shows a comparison of the normal probability function and the corresponding log-normal probability function and also the effect of varying $\mu$ and $\sigma^2$ on the log normal. Figures 3 through 6 show the theoretical and observed frequencies in graphical form. Four out of seven good statistical fits were obtained. Graphically all of the plots were close to the actual data.

## DATA COLLECTION

Data were collected on I-71, a four-lane Interstate highway in Ohio, in order to evaluate these distributions. The distributions have been tested in two ways. First, 22 hours of data were collected and in turn were divided into hourly data, resulting in representative volumes of 882, 930, 1,040, and 1,278 vehicles per hour per lane.

Next the 22 hours of data were separated into 5-min intervals. Based on these intervals, hourly volumes were constructed from representative 5-min volumes. For example, 12 intervals with volumes ranging from 55 to 60 vehicles were chosen to obtain an hourly volume of between 660 and 720 vehicles. Using this method gave representative volumes of 551, 659, 781, 902, 1,026, 1,128, and 1,369 vehicles per hour per lane.

Finally, 1 hour of data was collected from each of the following sites in Ohio:

| Highway | Number of Lanes | Vehicles per Hour |
|---|---|---|
| US-40 | 4 | 445 |
| I-71 | 4 | 515 |
| US-24 | 2 | 339 |
| US-33 | 2 | 472 |

## CONCLUSION

Table 1 gives a summary of the distributions. Table 2 gives a comparison of the three distributions tested.

Of the distributions previously used for headway distributions, two of these, the composite exponential and the Pearson Type III, were chosen to be tested. One good statistical fit was obtained for the composite exponential at a low volume of 339 vph on a two-lane highway. The Pearson Type III was more representative at higher volumes. Although no other good statistical fits were obtained with these distributions, this may be considered more the rule than the exception, since there is such a variation in traffic flow due to factors that cannot be taken into consideration when vehicles are counted. To partially correct for this, we chose representative intervals to obtain an hour's data of specific volume. The closest fits for the Pearson Type III occurred, as expected,

**Table 2. Comparison of results of the three distributions.**

| Volume (vehicles per hour per lane) | Composite Exponential | | Pearson Type III | | Log Normal | |
|---|---|---|---|---|---|---|
| | Chi Square | Good Fit | Chi Square | Good Fit | Chi Square | Good Fit |
| 549 | 91.38 | No | 60.31 | No | 36.40 | Yes |
| 659 | 86.96 | No | 64.32 | No | 32.76 | Yes |
| 781 | 237.75 | No | 76.37 | No | 71.27 | No |
| 902 | 289.92 | No | 61.07 | No | 42.12 | Yes |
| 1,026 | 446.58 | No | 108.83 | No | 52.18 | No |
| 1,128 | 1,078.95 | No | 128.02 | No | 42.11 | Yes |
| 1,369 | 2,247.21 | No | 231.62 | No | 75.39 | No |

when noninteger values for the parameter k were used, although this does not improve the fit enough to be statistically good.

The resulting statistically good fits with the log-normal distribution were actually better than expected. They indicate as the plots tended to suggest that the log normal is a better prediction of headways than the other two distributions.

## ACKNOWLEDGMENTS

The research reported in this paper was sponsored by the Ohio Department of Highways in cooperation with the Federal Highway Administration while the author was associated with Transportation Research Center, Ohio State University.

## REFERENCES

1. W. F. Adams. Road Traffic Considered as a Random Series. Journal of Institute of Civil Engineers, 1936.
2. J. Aitchison and J. A. C. Brown. The Lognormal Distribution. Cambridge Univ. Press, New York, 1957.
3. J. M. Aitken. Simulation of Traffic Conditions at an Uncontrolled T-Junction. Traffic Engineering and Control, Oct. 1963.
4. J. Evidente. Vehicular Headway Distribution on Upgrades of 2-Lane Rural Highways. Dept. of Civil Engineering, Ohio State Univ., MS thesis, 1964.
5. D. L. Gerlough. Traffic Inputs for Simulation on a Digital Computer. HRB Proc., Vol. 38, 1959, pp. 480-492.
6. B. D. Greenshields and F. M. Weida. Statistics With Applications to Highway Traffic Analyses. Eno Foundation, 1952.
7. F. A. Haight. The Generalized Poisson Distribution. Annals of Institute of Statistical Mathematics, Tokyo, Vol. 11, No. 2, 1959, pp. 101-105.
8. F. A. Haight, B. F. Whisler, and W. W. Mosher, Jr. New Statistical Methods for Describing Headway Distribution of Cars. HRB Proc., Vol. 40, 1961, pp. 557-564.
9. Investigation and Measurement of Traffic Dynamics. Ohio State Univ., Rept. EES202C-1, 1965.
10. J. P. Kinger. Application of the Theory of Probability to Problems of Highway Traffic. Proc., Institute of Traffic Engineers, 1934.
11. A. D. May and F. A. Wagner, Jr. Headway Characteristics and Interrelationships of Fundamental Characteristics of Traffic Flow. HRB Proc., Vol. 39, 1960, pp. 524-547.
12. K. Pearson. Tables of the Incomplete Gamma Function. Cambridge Univ. Press, 1922.
13. A. Schuhl. The Probability Theory Applied to Distribution of Vehicles on Two-Lane Highways. In Poisson and Traffic, Eno Foundation, 1955.
14. J. E. Tolle. The Lognormal Headway Distribution. Traffic Engineering and Control, Vol. 13, No. 11, 1971.
15. J. E. Tolle and J. Treiterer. The Distribution of Vehicular Headways—A Stochastic Model. In Investigation of Traffic Dynamics by Aerial Photogrammetry Techniques, Ohio State Univ., Rept. EES278-2, 1969.