

Update on the Statistical Analysis of Traffic Countings on Two-Lane Rural Highways

Raffaele Mauro¹ & Federico Branco²

¹ University of Trento, Faculty of Engineering, Trento, Italy

² Consulting Engineer, Trento, Italy

Correspondence: Raffaele Mauro, University of Trento, Faculty of Engineering, Trento, Italy. Tel: 39-461-282-588. E-mail: raffaele.mauro@ing.unitn.it

Received: April 7, 2013

Accepted: May 2, 2013

Online Published: May 21, 2013

doi:10.5539/mas.v7n6p67

URL: <http://dx.doi.org/10.5539/mas.v7n6p67>

Abstract

This paper deals with the problem of model identification, calibration and validation for traffic countings on two-lane rural highways. A criterion for preliminary selection of arrival laws as a function of appropriate sample statistics and a technique for deciding whether sample data sets of traffic counting are congruent with stationary time series behavior are suggested; besides arrival laws currently used in research and engineering practice, the Neyman distribution has been also applied although it is not frequently implemented in the field of traffic engineering. Moreover, this work aims at applying these methods to a set of empirical data derived from a recent survey on two two-lane rural highways; the arrival laws that best agree with the observations are found and the relations between the parameters identifying the arrival laws and the flow rates are worked out. Finally, the results have been compared to those achieved in similar observations, carried out by one of the authors in the past.

Keywords: two-lane rural highway, counting distribution, steady-state flow

1. Introduction

Traffic counting represents the first and, to this day, foremost empirical measurement designed to research on vehicular flows; counting distributions have direct relevance to discrete-time point processes applied to road traffic (Bertò, Schoen, & Speranza, 1996). Statistical models for traffic counting, also known as probabilistic arrival laws, and vehicular headway distributions (Mauro & Branco, 2012; Ha, Aron, & Cohen, 2012) are used in road and highway engineering, i.e. in traffic simulation procedures, in vehicle density estimations as well as in the study of the waiting phenomena at intersections and barriers. From the beginning of the 30s until the late 70s, special attention was devoted to theoretical and application aspects of these topics. Important studies have been made by Kinzer (1933), Adams (1936), Breiman (1962) and Gerlough and Barnes (1971). Helpful guidelines for traffic engineers about the study and application of counting distributions are also contained in Gerlough and Huber (1975) and in May (1990). After a few decades theoretical research has restarted and some works in the field have been produced by Jabari and Liu (2012), Clementi, Monti and Silvestri (2011) and Cao, Tai and Chan (2012) who have analysed some statistical models for counting distributions.

Since the 30s Poisson law has been proposed for theoretical distribution of arrivals, in that it is a peculiar flow model for discrete events under conditions of statistical regularity. Some interesting generalizations about this law have been subsequently made along with the progress of research on the issue of vehicular arrivals. Poisson law was introduced by Kinzer (1933) for the elaboration of certain aerial surveys on traffic, previously carried out by Johnson (1928). Afterwards, it was applied by Adams (1936) for further numerical exemplifications and by Greenshields, Shapiro and Ericksen (1947) in a work on intersections. A lot of applications of this traffic model have been later suggested by Gerlough, first in a book entirely on the topic (Gerlough, 1955) and more recently, in collaboration with others, in a paper that also deals with further theoretical distributions that may be used as arrival laws (Gerlough & Barnes, 1971). Finally, certain realistic models of vehicular flows leading to Poisson arrivals are described in Breiman's (1962, 1963) and Weiss and Herman's (1962) researches.

In the study of traffic conditions for which Poisson model is not useful, various scholars have suggested and sometimes verified through experimentation different counting laws. Specifically, among the most interesting

contributions to applications, Beckmann et al. have proposed a simple binomial distribution model (Beckmann, McGuire, & Winsten, 1956); Haight (1959) has introduced the generalized Poisson law and studied it with reference to real cases in collaboration with others (Haight, Whilser, & Mosher, 1961); Buckley (1965) and Drew (1965) have researched the possibility of applying the negative binomial distribution; and, once again, Buckley (1968) has further generalized the Poisson law. Theoretical works which propose counting models, barely or not at all verified through empirical validation, cannot be predominantly neglected. Oliver and Thibault (1962), Buckley (1965), Buckley (1967), and Serfling (1969) can be consulted on such a type of law.

Finally, Ha, Aron and Cohen (2012) can be referred to for most recent results; they have introduced innovative models for time headway and counting distributions, supported by empirical researches on French roads.

2. Probabilistic Models for Arrivals

Results from experiments have shown that according to the statistical hypothesis Poisson law is, at the roots of its deduction, correctly applicable as a model for arrivals, on one or multiple lanes, if in the observed interval:

- a) the phenomenon remains stationary, i.e. no external perturbation intervenes and affects flows;
- b) the gap between vehicles is such that they do not influence each other.

Under traffic conditions in which the vehicles are not farther apart than the distance at which they do not interact with each other, the circumstance b) involves that the lower the flow, the more consistent with empirical data is the model. Considering the results from this research, the authors suggest to apply the model to flow rates up to 400÷500 vph in ordinary road conditions (dry weather, daylight, pavement in good conditions). Thus, Poisson law cannot generally be used without stationarity and on high flow rates; if these two conditions are not met, other probability laws are to be adopted. The criterion for choosing alternative counting models, presented in the literature and exposed in Gerlough and Huber (1975), is described below.

If the mean $\bar{x}$ and variance s^2 of the sample turn out to be substantially equal, one can assume that the statistical distribution may be close to Poisson law. According to such a law, the mean μ and variance σ^2 turn out to be equal and this value completely defines the model. For the sake of simplicity, this paper does not deal with other statistical laws for which $\mu = \sigma^2$. If the mean of the sample turns out to be higher than the variance, the measure variability can be deduced as lower than that expected from purely poissonian arrivals of equal mean. Should that be the case, the empirical data can be checked to see how consistent they are with the positive binomial model or with the generalized Poisson model, according to which the mean is generally higher than the variance. Also the quantities μ and σ^2 completely identify both distributions. In terms of traffic, the circumstance $\bar{x} > s^2$ has been found on a frequent basis in flow conditions far from the free flow circulation, when flow rates usually present a very high volume. Finally, if the mean of the sample is lower than its variance—in other words, for a given mean value countings appear to be more dispersed than those derived from purely poissonian arrivals—the negative binomial distribution, also known as Pascal's law, should be generally determined, in that it presents, like the data, a lower expectation than the variance $\mu < \sigma^2$. Also in this case, mean and variance completely identify the model. Cases of traffic countings where $\bar{x} > s^2$ and in the presence of compliance with Pascal's law have been mainly observed in flow conditions developed under traffic-light regulation, although Buckley (1967) has used this distribution as a model for arrivals on roads with two or more lanes.

Table 1 summarizes the criterion for choosing a model according to the mean $\bar{x}$ and variance s^2 values.

Table 1. Selection criteria for traffic counting models

Value of $s^2/\bar{x}$	Suggested distribution
> 1	negative binomial
$\equiv 1$	Poisson
< 1	binomial or generalized Poisson

Table 2 shows the probability distribution of the binomial, Poisson and negative binomial random variables with their parameter expressions as a function of the sample statistics $\bar{x}$ and s^2 .

It is worth remembering that the choice for the length of the subdivision intervals Δt of the observation period T is known to influence the identification of the model, as well explained by Gerlough and Barnes (1971). For example, this means that, in the same period of observation, a flow can be Poissonian if the data are recorded on

subintervals Δt_1 , but on the other hand it can conform to other models if $\Delta t_2 < \Delta t_1$ are used. As for the estimates shown in this paper, the width of the interval was predetermined by the assumptions about the minimum extension of the periods of stationarity. Moreover, the duration T of the observation period, at a constant length of the subinterval of subdivision t , has been proved to affect the ratio $s^2/\bar{x}$: the bigger the interval T , the bigger the ratio $s^2/\bar{x}$ (Miller, 1970).

Table 2. Theoretical distributions of arrival: probability distribution, mean, variance, parameters estimate

Distribution	binomial	Poisson	negative binomial
probability distribution	$\binom{n}{x} p^x (1-p)^{n-x}$	$\frac{\mu^x}{x!} e^{-\mu}$	$\binom{x+k-1}{k-1} p^k (1-p)^x$
mean μ	np	μ	$\frac{k(1-p)}{p}$
variance σ^2	$np(1-p)$	μ	$\frac{k(1-p)}{p^2}$
$\frac{\mu}{\sigma^2}$	$(1-p)^{-1} > 1$	1	$p < 1$
parameters estimate	$p = (\bar{x} - s^2)/\bar{x}$ $n = \bar{x}^2/(\bar{x} - s^2)$	$\mu = \bar{x}$	$p = \bar{x}/s^2$ $k = \bar{x}^2/(s^2 - \bar{x})$

In addition to the selection criterion indicated in Table 1, a further criterion for the preliminary choice of the model, as a function of appropriate statistical data, is shown. In some research areas (e.g. Biometrics) (Gore & Paranipe, 2001), this criterion is more efficient than the simple comparison between the values of mean and variance: in order to have a greater amount of information than that gained only from the comparison of $\bar{I} = s^2/\bar{x}$ with $I = \sigma^2/\mu$, the parameter $\bar{L} = m_3/s^2$ can also be compared with $L = \mu_3/\sigma^2$, where μ_3 is the theoretical third central moment and m_3 is the corresponding frequency moment. On the plane (I,O,L) the point (I,L) is placed on a different position for each different theoretical distribution (Figure 1), lying on the segment AC if it is representative of a binomial distribution, on the half-line with origin in C if it is representative of a negative binomial distribution, and coinciding with the point C if it is representative of the Poisson law (Ord, 1972).

The positioning of the sample points $(\bar{I}, \bar{L})$ on the plane (I,O,L) near the Loci previously defined can be helpful for the choice of the theoretical distribution that best fits the sample data; however, the statistical hypothesis can be verified by applying a test of hypothesis to the chosen model.

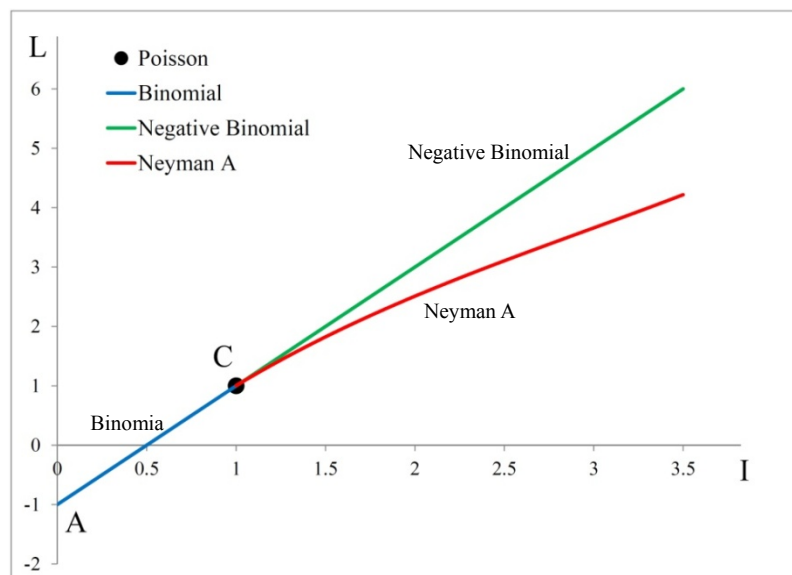


Figure 1. Probability distributions on the plane (I,O,L)

This method is very useful and easy to implement, but on the other hand it is not always univocally discriminative, as later shown in this work. In Figure 1 the curve $L = f(I)$ of the Neyman type A biparametric distribution is also illustrated. The probability distribution of the Neyman type A is:

$$p(0) = e^{-m_1} (1 - e^{-m_2}) \quad (1)$$

$$p(n+1) = \frac{m_1 m_2 e^{-m_2}}{n+1} \sum_{k=0}^n \frac{m_2^k}{k!} p(n-k), \quad n = 0, 1, 2, \dots \quad (2)$$

where m_1 and m_2 are:

$$m_1 = \bar{x}^2 / (s^2 - \bar{x}) \quad (3)$$

$$m_2 = (s^2 - \bar{x}) / \bar{x} \quad (4)$$

Relations $L = f(I)$ for the models just mentioned and deduced from the expressions of μ , σ^2 and μ_3 , are shown in Table 3.

Table 3. Relations $L = f(I)$ for the probability models

Distribution	Relation $L = f(I)$	
Poisson	$L = I = 1$	
binomial	$L = 2I - 1$	$I \in [0, 1]$
negative binomial	$L = 2I - 1$	$I \in [1, \infty[$
Neyman type A	$L = I + 1 - 1/I$	$I \in [1, \infty[$

The negative binomial and Neyman type A biparametric distributions belong to the larger family of probabilistic models known as “aggregate or contagious distributions”. They are of great importance in Biometrics when the elements of the surveyed populations are generally found in groups or clusters. As for traffic, the equivalent situation can be platoons. Therefore, in such a circumstance (running in platoons) traffic countings should be consistent with the negative binomial or Neyman type A biparametric distributions. This side of the problem will be dealt with in the following analysis of the empirical results.

The next section firstly shows a statistical procedure for the identification of steady-state periods; in fact, model identification, calibration and validation of counting laws have to be made on these stationarity periods. These periods need to be defined so as to proceed with the search for links between the statistical parameters identifying models and flow rates, or more in general, relations between flow rates and traffic processes parameters.

3. Flow Stationarity Tests

T is defined as the observation period of the traffic flow in a section of a lane. If T is divided into n smaller intervals t_i , $i = 1, 2, \dots, n$ of equal length Δt and the mean of the process, represented by the sequence $\{X_i\}$ of the random variables defined as “number of vehicles crossing the road section during the interval Δt_i ”, is constant, the flow is defined as stationary during the period T . In order to check that the realization $\{x_i\} = x_1, x_2, \dots, x_n$ of the sequence $\{X_i\}$ is extracted by a process of traffic countings with a constant mean, a “distribution free” test is used in this paper in that, unlike other types (for instance, the sequential probability ratio test by Wald (1947)), it has the advantage of not requiring any statistical hypothesis on the arrival law. Through this test the constancy of the process mean is controlled by verifying the independence hypothesis of the sequence $\{X_i\}$ compared to the independence of the first “ n ” natural numbers (Kendall & Stuart, 1967).

The parameter used in this test is the correlation coefficient r between the two sequences:

$$r = \frac{\frac{1}{n} \sum_{i=1}^n i x_i - \bar{x} \frac{n+1}{2}}{\left[\frac{1}{12} s^2 (n^2 - 1) \right]^{1/2}} \quad (5)$$

where $\bar{x}$ and s^2 are the mean and the variance of $\{X_i\}$ elements. The distribution of equation (5) is obtained by

considering that the n values of r calculated in n permutations of $\{X_i\}$ elements are equiprobable. The extreme values of “ r ” are evidently -1 and $+1$, where $r = +1$ if the sequence of traffic countings increases linearly with the natural numbers, $r = -1$ if the $\{X_i\}$ decrease when the natural numbers increase. Choosing suitably large n , typically $n > 8$, it is possible to limit the length of the acceptance zone of the test and to replace (5) with the parameter obtained from (5) that follows:

$$t = \left[\frac{(n-2)r^2}{1-r^2} \right]^{1/2} \quad (6)$$

distributed according to a Student's t with $(n-2)$ degrees of freedom.

Through these parameters, the hypothesis of flow stationarity at level α is refused when

$$|t| \geq t_{\alpha/2} \quad (7)$$

One has to choose n so that the length Δt is greater than a minimum value; in this way it is possible to consider the elements $\{X_i\}$ as independent elements and to apply to them a “distribution free” test, such as the one just mentioned.

Finally, the same length Δt of the observation interval must be kept sufficiently small to avoid non-linear flow variations during the interval. Such fluctuations cannot be pointed out by the test adopted which is rather very useful only for the linear trend hypothesis. Consequently, the present study focuses the quantities $\Delta t = 20$ sec and $n = 30$; so the observation period is equal to $T = 10$ min. With the foregoing assumptions, for definition the flow is considered stationary for intervals of less than 10 min.

In order to verify that the length $\Delta t = 20$ sec is appropriate to consider the observed traffic countings as independent elements, a lag 1 serial correlation test is carried out. This verification is usually used by applying, for the elements of $\{X_i\}$, the autocorrelation coefficient given by:

$$\rho = \left(\frac{\mu_3}{6\mu^3} - \frac{(\sigma^2)^2}{4\mu^4} - \frac{1}{12} \right) \left(\frac{\sigma^2 \Delta t}{\mu^3} + \frac{1}{6} + \frac{(\sigma^2)^2}{2\mu^4} - \frac{\mu_3}{3\mu^3} \right)^{-1} \quad (8)$$

where μ , σ^2 and μ_3 are the mean, variance and third central moment of the time headway distribution (Cox & Lewis, 1966). Using Equation (8) it is possible to calculate ρ and to evaluate if it is small enough to accept the hypothesis of absence of statistical relationships between the elements of the traffic counting sequence.

Indeed, in this research work no information about time headways is available; therefore, the lag 1 serial correlation test just above recalled is used through the quantity

$$r = \sum_{i=1}^n x_i x_{i+1} \quad (9)$$

Equation (9) for sufficiently large n ($n > 15$) is normally distributed with mean and variance given by:

$$\mu_r = \frac{s_1^2 - s_2}{n-1} \quad (10)$$

$$\sigma_R^2 = \frac{s_2^2 - s_4}{n-1} + \frac{s_1^4 - 4s_1^2 s_2 + 4s_1 s_3 + s_2^2 - 2s_4}{(n-1)(n-2)} - \mu_r \quad (11)$$

where:

$$s_k = \sum_{i=1}^n x_i^k \quad (12)$$

The knowledge of the theoretical distribution of r allows performing a common two-sided test at the chosen significance level α . Through this test, the hypothesis of independence is rejected if:

$$|r| > r_{\alpha/2} \quad (13)$$

Returning now to the correlation ratio test for the verification of the flow stationarity, it has been applied in two ways to the data surveyed and with the values of n , Δt and T previously specified.

The former way made the assumption of a test dimension equal to $\alpha = 0.05$ and treated the countings of the first thirty Δt_i starting from the observation instant. If the stationarity occurred, three additional consecutive subintervals were added and the first three subintervals were excluded; such a test was then repeated on the new

sequence. The iteration stopped when the non stationarity of $\{X_i\}$ was obtained. The test was then performed on arrivals per 10-minute interval, where the first minute corresponded to the last minute of the previous interval. The range of acceptance of this test, applied to the empirical data, was not small, even if the sample size n was not very large. The procedure for data sets creation was therefore modified in order to obtain realizations of gradually size increasing $\{X_i\}$ as well as increasingly narrow acceptance intervals: starting from the first instant, the countings relating to the following Δt_i were added one by one and the test was carried out, step by step, on larger and larger samples; the process is stopped when the non-stationary flow occurred. The procedure was iterated starting from the traffic counting value which corresponded to the last minute of the previous sequence and identifying further periods at a constant flow, and so forth up to the complete analysis of the available data. The test presented a considerable inertia and revealed the presence of a trend only in very few cases.

For those reasons, the stationarity intervals identified through the first technique were then employed in the following analyses.

The methodology used here has been described by Esposito and Mauro (1994) who applied the technique for the identification of stationarity periods to empirical measures of traffic countings. After nearly two decades the authors used the same technique on newly-acquired empirical data, as specified below; the results from newly-collected data and their comparison with those obtained by Esposito and Mauro (1994) are presented in the following paragraph.

4. Empirical Data Analysis

The methods described in paragraphs 2 and 3 were used in traffic countings data carried out in the spring of 2012; these data refer to two sections of two-lane rural highways (one lane for each direction). In particular, the data of 1800 intervals with a width $\Delta t = 20$ sec, equal to a total of 10 hours and 3053 vehicular passages, were analysed. During these periods, the observed road section was filmed by a video camera and the countings were later carried out by means of the recordings.

Empirical data have been surveyed on two road sections of the following roads in the province of Trento, Italy:

- 1) Provincial Road No. 36 “delle Grazie”, in the municipality of Arco;
- 2) National Road No. 421 “dei Laghi di Molveno e Tenno”, in the municipality of Tenno.

The first section is set on a long, flat and straight road where overtaking is disallowed. The roadway is about 7.00 m wide, with lane width of 3.25 m and paved shoulders of 0.25 m. Table 4 shows some information gained from the countings on section 1.

Table 4. Information about the countings on the Provincial Road No. 36 “delle Grazie”

Set	Date	Time	Direction
A	April 26 th , 2012	10:00-12:00	North
B	April 26 th , 2012	13:00-14:00	South
C	April 27 th , 2012	16:00-17:00	North
D	April 27 th , 2012	17:00-18:00	South

Instead, section 2 is set on a short, straight road, with a 6% grade, where overtaking is disallowed. The carriageway is about 6.50 m wide and the paved shoulders are about 0.50 m. Table 5 summarizes some information on the survey.

Table 5. Information about the countings on the National Road No. 421 “dei Laghi di Molveno e Tenno”

Set	Date	Time	Direction
E	May 10 th , 2012	15:00-18:00	South
F	May 10 th , 2012	17:00-19:00	North

By applying the methodology above described to all the collected data, the observed intervals were divided into 23 stationarity periods, summarized in Table 6 and in Table 7, respectively for the former and the latter road section.

Table 6. Data on Provincial Road No. 36 “delle Grazie”

Direction and date	Hour and stationarity period length	S e t	Traffic counting data			I and L estimate		Parameters					
			$\bar{X}$ (veh/20 sec)	s^2	Flow rate (veh/h)	I	L	negative binomial		binomial		Neyman A	
								p	k	p	n	m_1	m_2
North 04/26/'12	10.00-10.33 33 min	A ₁	2.389	5.830	430	2.44	3.67	0.410	1.658	-	-	-	-
	10.34-10.52 18 min	A ₂	2.567	8.637	462	3.37	6.24	0.297	1.085	-	-	-	-
	10.53-11.22 29 min	A ₃	1.694	2.434	305	1.44	1.47	-	-	-	-	3.882	0.436
	11.23-11.49 26 min	A ₄	1.583	4.051	285	2.56	4.48	0.391	1.016	-	-	-	-
South 04/26/'12	13.00-13.21 21 min	B ₁	2.694	6.829	485	2.53	4.10	0.395	1.756	-	-	-	-
	13.22-13.34 12 min	B ₂	2.617	3.614	471	1.38	1.44	-	-	-	-	6.865	0.381
	13.35-13.57 22 min	B ₃	2.878	7.899	518	2.74	4.35	0.364	1.649	-	-	-	-
North 04/27/'12	16.00-16.41 41 min	C ₁	1.783	3.700	321	2.07	3.00	0.482	1.659	-	-	-	-
	16.42-17.00 18 min	C ₂	2.111	3.600	380	1.71	1.78	-	-	-	-	2.993	0.705
South 04/27/'12	17.00-17.37 37 min	D ₁	2.739	10.430	493	3.81	5.98	0.263	0.975	-	-	-	-
	17.38-18.00 22 min	D ₂	3.178	14.518	572	4.57	7.84	0.219	0.890	-	-	-	-

Table 7. Data on National Road No. 421 “dei Laghi di Molveno e Tenno”

Direction and date	Hour and stationarity period length	S e t	Traffic counting data			I and L estimate		Parameters					
			$\bar{X}$ (veh/20 sec)	s^2	Flow rate (veh/h)	I	L	negative binomial		binomial		Neyman A	
								p	k	p	n	m_1	m_2
South 05/10/'12	15.00-15.26 26 min	E ₁	0.961	0.876	173	0.91	0.89	-	-	0.089	11	-	-
	15.27-16.03 36 min	E ₂	1.117	1.702	201	1.52	2.24	0.656	2.130	-	-	-	-
	16.04-16.29 25 min	E ₃	0.939	0.721	169	0.77	0.81	-	-	0.232	4	-	-
	16.30-16.58 28 min	E ₄	1.083	0.762	195	0.70	0.54	-	-	0.297	4	-	-
	16.59-17.21 22 min	E ₅	1.211	1.875	218	1.55	1.71	-	-	-	-	2.209	0.548
	17.22-17.48 26 min	E ₆	1.272	2.214	229	1.74	1.92	-	-	-	-	1.719	0.740
North 05/11/'12	17.00-17.18 18 min	F ₁	1.133	1.540	204	1.36	2.03	0.736	3.158	-	-	-	-
	17.19-17.29 10 min	F ₂	1.250	2.238	225	1.79	2.08	-	-	-	-	1.581	0.790
	17.30-17.44 13 min	F ₃	1.494	2.980	269	1.99	3.00	0.501	1.503	-	-	-	-
	17.45-18.12 27 min	F ₄	1.217	2.052	219	1.69	2.82	0.593	1.772	-	-	-	-
	18.13-18.31 18 min	F ₅	1.550	3.110	279	2.01	3.18	0.498	1.540	-	-	-	-
	18.32-19.00 28 min	F ₆	1.606	3.870	289	2.41	3.72	0.415	1.138	-	-	-	-

For each stationarity period such tables contain the mean (for vehicles per 20 sec and vehicles per hour), the variance and the parameters I and L. By adopting the criteria laid down in paragraph 2, the theoretical distributions that best approximate the observed data for each period with a constant flow rate were analysed; the values of the parameters obtained by the calculations are also indicated in the tables.

No empirical series is well approximated by a Poisson distribution; the observed data, instead, seem to be approximated by the negative binomial (in most cases), binomial or Neyman type A distributions. The graphical representation on the plane (I,O,L) of the sample points $(\bar{I}, \bar{L})$ has helped to identify the most suitable theoretical model; in that regard see Figure 2 and Figure 3. These models were calibrated through the criteria indicated in paragraph 2 and their parameters are shown in Table 6 and Table 7.

The validation of the chosen models to the theoretical predictions was then verified by performing a χ^2 test and by calculating the relative squared index, which confirmed the above choices. More specifically, the relative squared index I_{rs} is defined as follows:

$$I_{rs} = \frac{\sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n}}}{\sum \hat{y}_i}; \quad 1 < i < n \quad (14)$$

where y_i is the observed value i ; $\hat{y}_i$ is the value i of the theoretical distribution.

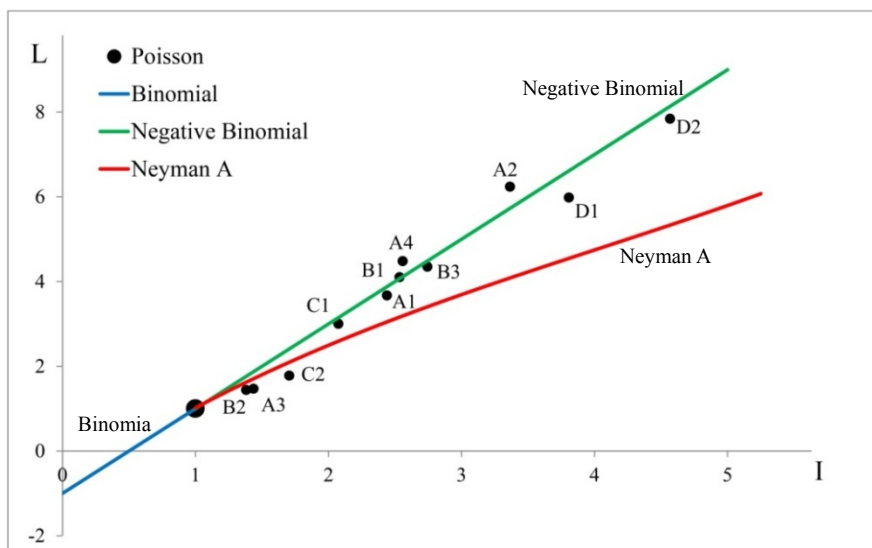


Figure 2. Plane (I,O,L) for countings on Provincial Road No. 36 “delle Grazie”

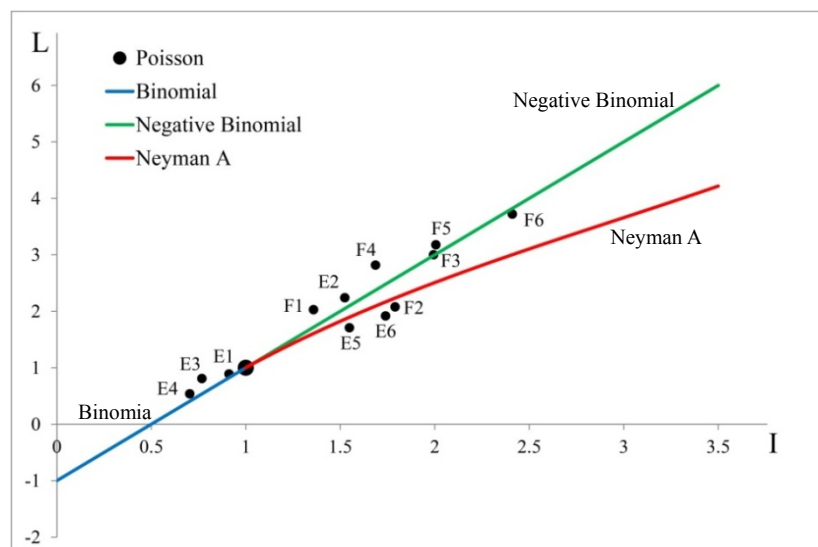


Figure 3. Plane (I,O,L) for countings on National Road No. 421 “dei Laghi di Molveno e Tenno”

By way of example, Figure 4, Figure 5 and Figure 6 show the frequency distributions of three sets of empirical traffic countings, together with the corresponding theoretical distributions that best approximate the data; the figures also show the value of the probability p_χ obtained by applying a χ^2 test to assess the conformity between the sample statistic and the law chosen as its theoretical representation (note that values p_χ tending to 1 support the hypothesis verification) and the relative squared index I_{rs} . An example for each analysed model (negative binomial, binomial and Neyman type A) is shown.

As for the binomial and negative binomial frequency distributions, in order to avoid approximations, the binomial coefficient was calculated by generalizing the factorial operator and using the Gamma function $\Gamma(x)$ through the following definition of general validity:

$$\binom{x}{y} = \frac{\Gamma(x+1)}{\Gamma(y+1)\Gamma(x-y+1)} \quad (15)$$

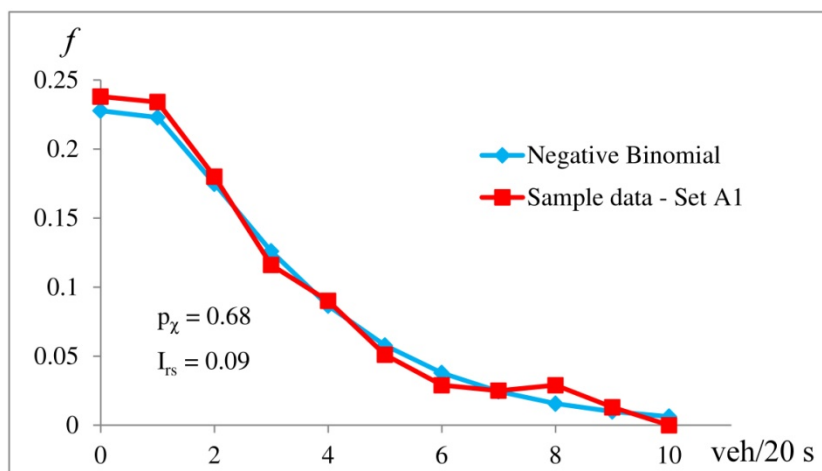


Figure 4. Frequency distributions for set A1

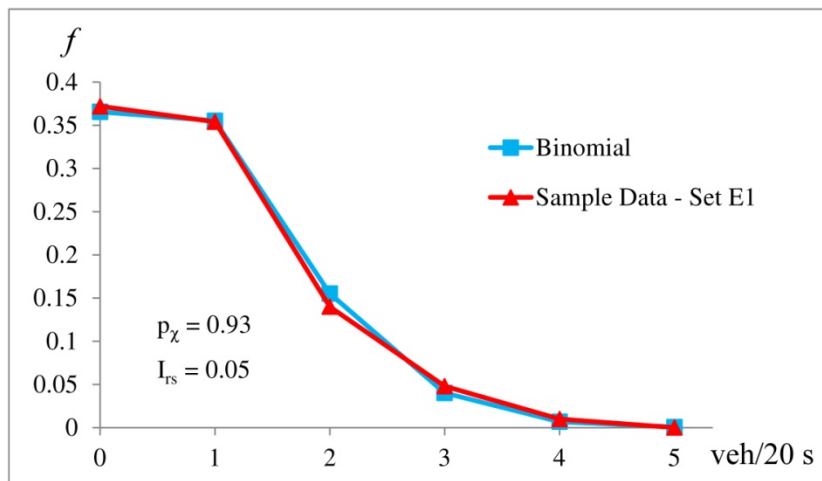


Figure 5. Frequency distributions for set E1

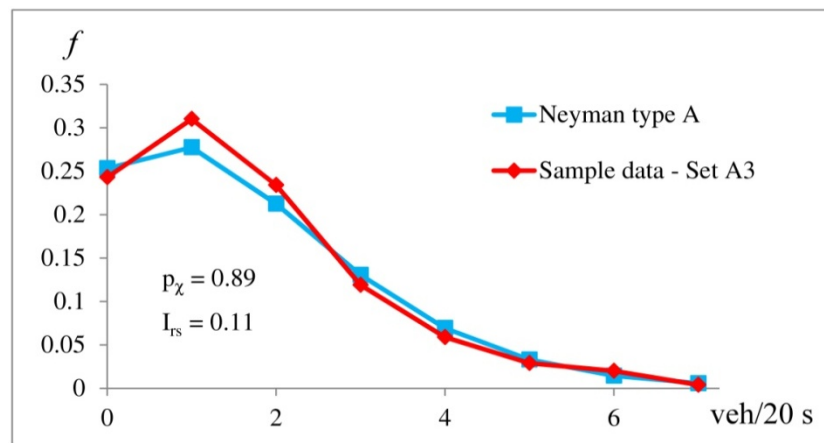
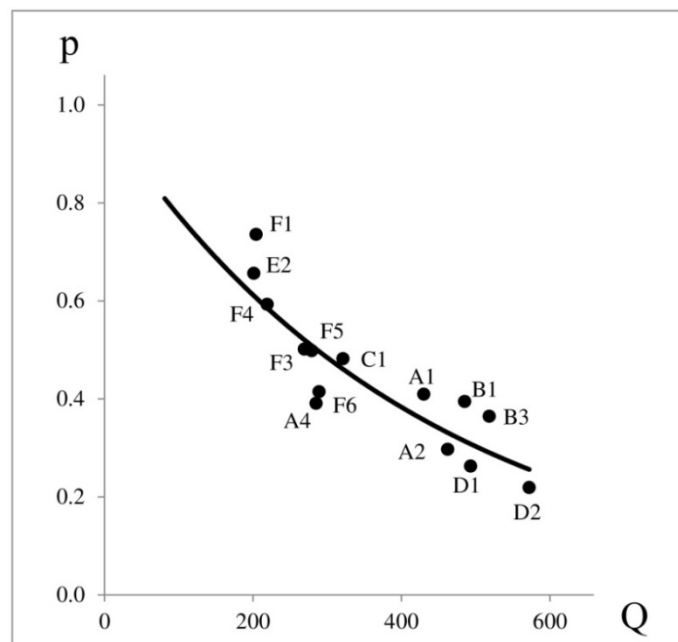


Figure 6. Frequency distributions for set A3

Finally, links between the parameters p and k , which define the negative binomial model, and the flow rate Q have been investigated by means of the least-squares method; the following equations were obtained (see also Figure 7 and Figure 8):

$$p = 0.9784e^{-0.0024Q}, Q > 50 \text{vph}; r = 0.885 \quad (16)$$

$$k = 34.93Q^{-0.543}, r = 0.557 \quad (17)$$

Figure 7. Regression curve between the negative binomial parameter p and the flow rate

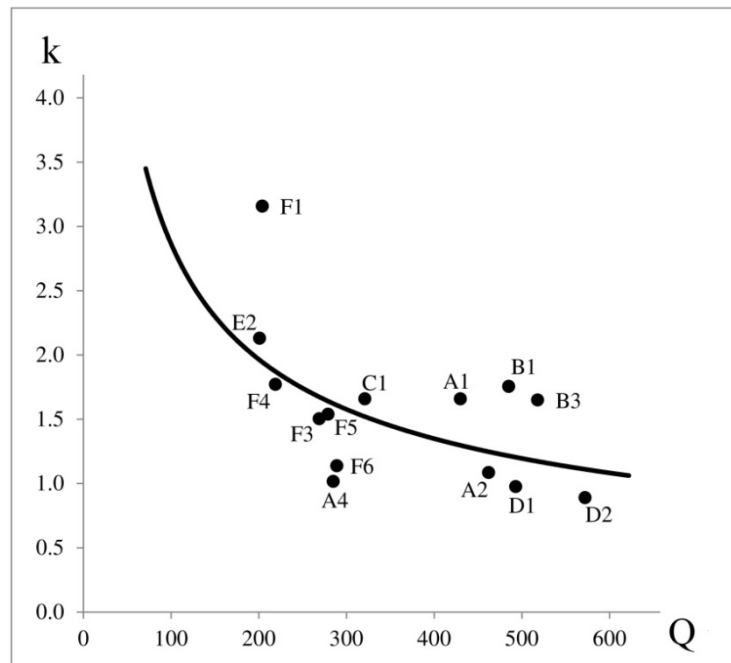


Figure 8. Regression curve between the negative binomial parameter k and the flow rate

As mentioned above, a similar analysis had been carried out by Esposito and Mauro (1994); in that paper traffic countings were carried out in 2892 intervals with a width $\Delta t = 20$ sec, for a total of 4745 vehicular passages. These data were collected on three two-lane rural highways between 1989 and 1992; each road had different geometric alignment features, as well as different overtaking rules.

Esposito and Mauro (1994) identified 33 steady-state periods; their data (minimum and maximum values) are summarised in Table 8.

Table 8. Summary of data collected by Esposito and Mauro (1994)

Number of stationarity period	Length of stationarity period (minutes)	Flow rate (veh/h)	I and L estimate		Parameters			
			I	L	negative binomial		binomial	
					28 periods		5 periods	
					p	k	p	n
33	min. 11	min. 113	min. 0.53	min. 0.35	min. 0.204	min. 0.807	min. 0.032	min. 3
	max. 60	max. 630	max. 4.91	max. 8.49	max. 0.828	max. 4.808	max. 0.475	max. 20

The regression curve equations, obtained by Esposito and Mauro (1994), are as follows:

$$p = 1.154e^{-0.0028Q}, Q > 50 \text{ vph}; r = 0.876 \quad (18)$$

$$k = 227Q^{-0.857}; r = 0.626 \quad (19)$$

These curves are shown in Figure 9 and Figure 10, together with the regression curves of the parameters p and k obtained by jointly examining data in the paper by Esposito and Mauro (1994) and data coming from the above said surveys carried out in the spring of 2012; the following equations were obtained:

$$p = 1.1022e^{-0.0027Q}, Q > 50 \text{ vph}; r = 0.876 \quad (20)$$

$$k = 137.34Q^{-0.773}; r = 0.610 \quad (21)$$

By comparing the relations at different times, the curves appear to be substantially corresponding to flow rate values which are located approximately in the central part of the examined range ($300 < Q < 500$); instead, the curves move away in correspondence with the extreme values, but not in a pronounced way. The curves obtained from the complete data analysis (data collected during the years 1989-1992 and 2012) are very close to those

found in the paper by Esposito and Mauro (1994) (data collected during the years 1989-1992); this can be easily explained by considering that the sample points $(\bar{I}, \bar{L})$ obtained by traffic countings made in 2012 are the half than those resulting from traffic countings carried out by Esposito and Mauro (1994).

Although carried out at a distance of about two decades, the results obtained from the two traffic counting surveys are essentially consistent with one another; they can be used in simulations which do not allow performing a complete empirical analysis and in parameter estimation for arrival statistical distributions, on condition that there is compliance with the negative binomial theoretical distribution.

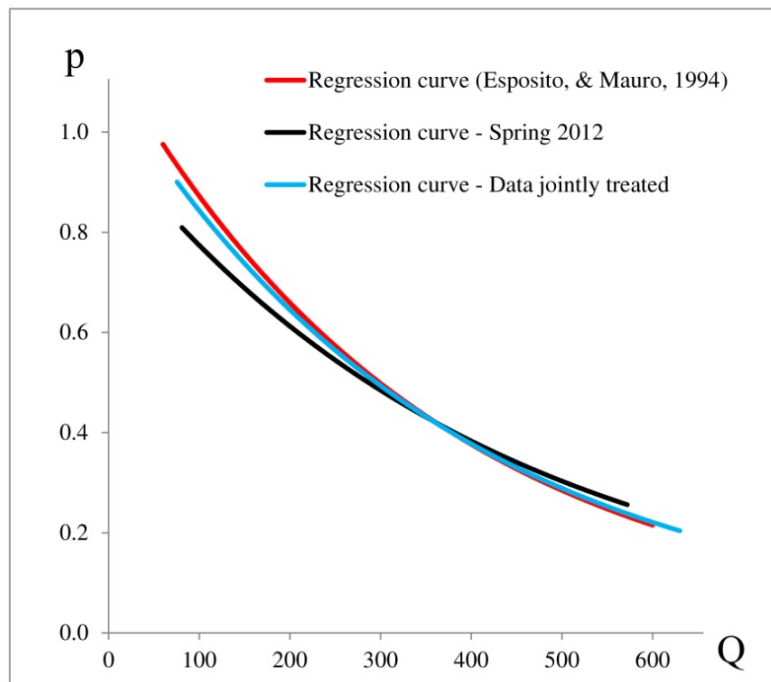


Figure 9. Regression curve between the negative binomial parameter p and the flow rate

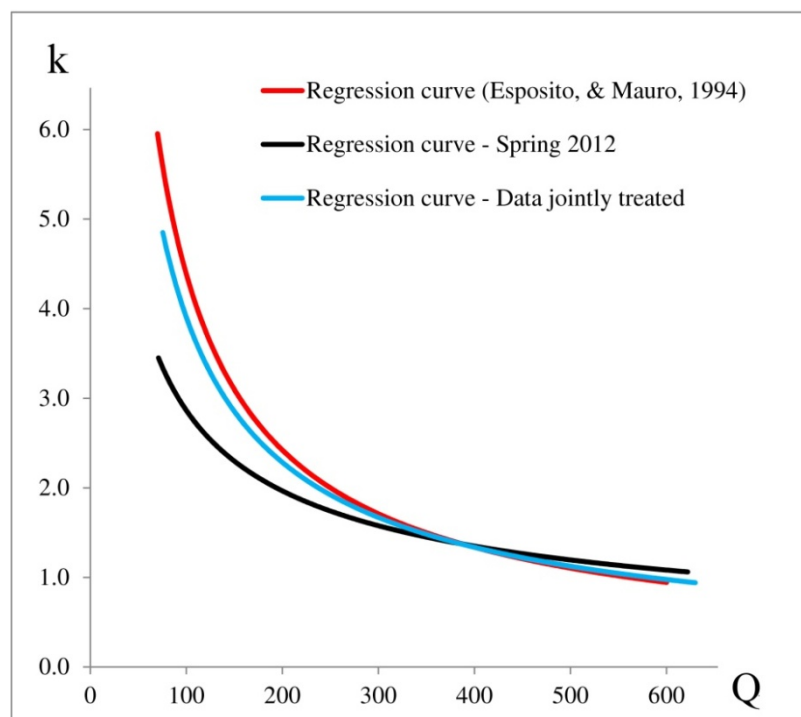


Figure 10. Regression curve between the negative binomial parameter k and the flow rate

5. Conclusions

The paper deals with the identification, calibration and validation of statistical models for countings of two-lane rural highways, especially with the definition and determination of the statistical stationarity periods of the flow rate, on the basis of the observed data analysis.

After a brief review of probabilistic arrival models usually used in the study of traffic phenomena, a complete statistical analysis methodology has been developed to define stationarity periods, to verify the independence of the events (the sequence of arrivals) as well as to identify the probability laws for traffic countings. The procedure was applied to some samples of traffic counting, empirically surveyed on two two-lane rural highways; then, links were established between flow rate and parameters of the negative binomial distribution (found to comply with most of the data). However, the same analysis was not performed in the binomial, Poisson and Neyman type A distributions because of the very limited availability of data. Finally, a comparison was made between the results from the 2012 data collection and those reported in the previous paper by Esposito and Mauro (1994); the latter applied the same procedure to other data sets.

The analysis of the data collected specifically for this study confirms what already mentioned by Esposito and Mauro (1994); indeed, it shows that the arrivals analysed in the two roads are mostly well modeled by the negative binomial counting distribution (which is an “aggregate” or “contagious” distribution), and other data (though insufficient for in-depth analysis) is well modeled by the Neyman type A distribution, which is defined as an “aggregate” or “contagious” distribution as well. These results are consistent with the fact that the flow mainly recorded on the two examined infrastructures is a commonly called “platoon” flow. So, in general and even for low flow values, the Poisson law seems to be unsuitable for representing the arrivals on two-lane rural highways; instead, other models are recommended in these cases. Such models have to take into account the vehicle group formation influenced by a leader.

Finally, as regards the relationships between the parameters of the negative binomial model and the flow rate, the comparison between the relations from this study and those by Esposito and Mauro (1994) has shown that the curves are substantially corresponding each other for flow rate values which are located approximately in the central part of the examined range ($300 < Q < 500$); instead, the curves move away in correspondence with the extreme values, but not in a pronounced way.

The results obtained from the two traffic measurement surveys, although carried out at a distance of about two decades, are essentially in good agreement, and they may be useful comparison tools for parameter estimates of arrival distributions, on condition that there is compliance with the negative binomial distribution. Such results can be useful in simulations when a complete empirical analysis cannot be performed. However, the limited size of the sample data analysed does not allow to extend the relationships found to other two-lane rural highways without taking proper precautions or performing adequate sensitivity analyses.

References

- Adams, W. F. (1936). Road traffic considered as a random series. *J. Inst. Civ. Engrs*, 4, 121-130. <http://dx.doi.org/10.1680/ijoti.1936.14802>
- Beckmann, M. J., McGuire, C. B., & Winsten, C. B. (1956). *Studies in the Economics of Transportation*. Yale University Press.
- Bertò, B., Schoen, F., & Speranza, M. G. (1996). Discrete time point processes: application to road traffic. In M. Papageorgiou (Eds.), *Concise Encyclopedia of traffic & transportation systems*. Pergamon Press.
- Breiman, L. (1962). On some probability distributions occurring in traffic flow. *Bull. Inst. Statist. Inst.*, 155-161.
- Breiman, L. (1963). The Poisson tendency in traffic distribution. *Ann. Math. Statist.*, 34, 308-311. <http://dx.doi.org/10.1214/aoms/1177704267>
- Buckley, D. J. (1965). *Inter-vehicle spacing and counting distributions*. Ph.D. Thesis, School of Traffic Engineering of New South Wales, Sydney.
- Buckley, D. J. (1967). Road traffic counting distributions. *Transp. Res.*, 1(2), 105-116. [http://dx.doi.org/10.1016/0041-1647\(67\)90166-9](http://dx.doi.org/10.1016/0041-1647(67)90166-9)
- Buckley, D. J. (1968). A semi-Poisson model for traffic flow. *Transp. Sci.*, 2(2), 107-133. <http://dx.doi.org/10.1287/trsc.2.2.107>
- Cao, M., Tai, A. H., & Chan, L. Y. (2012). New statistical distributions for group counting in Bernoulli and Poisson processes. *Test*, 21, 29-53. <http://dx.doi.org/10.1007/s11749-010-0232-0>

- Clementi, A., Monti, A., & Silvestri, R. (2011). Modelling mobility: A discrete revolution. *Ad Hoc Networks*, 9, 998-1014. <http://dx.doi.org/10.1016/j.adhoc.2010.09.002>
- Cox, D. R., & Lewis, P. A. (1966). *The statistical analysis of series of events*. New York, NY: J. Wiley & S.
- Drew, D. R. (1965). Application of discrete distributions to traffic. *Traff. Eng. Control*, 7(8), 24-25.
- Esposito, T., & Mauro, R. (1994). Analysis of counting distributions for two-lane two-way roads (in italian). *Le Strade*, 5, 433-440.
- Gerlough, D. L. (1955). *Use of Poisson distribution in highway traffic*. Saugatuck, CT: The Eno Foundation for Highway traffic control.
- Gerlough, D. L., & Barnes, F. C. (1971). *The Poisson and other probability distributions in highway traffic*. Saugatuck, CT: The Eno Foundation for Highway traffic control.
- Gerlough, D. L., & Huber, M. J. (1975). Traffic flow theory, a monograph. *TRB Special Report*, 165, Washington, D.C.
- Gore, A., & Paranipe, S. (2001). *A course in mathematical and statistical ecology*. Dordrecht, The Netherlands: Kluwer Academic Publishers. <http://dx.doi.org/10.1007/978-94-015-9811-8>
- Greenshields, B. D., Shapiro, D., & Ericksen, E. L. (1947). *Traffic Performance at Urban Street Intersections*. New Haven, CT: Tech. Rep. No. 1, Yale Bureau of Highway Traffic.
- Ha, D. H., Aron, M., & Cohen, S. (2012). Time Headway variable and probabilistic modeling. *Transportation Research Part C*, 25, 181-201. <http://dx.doi.org/10.1016/j.trc.2012.06.002>
- Haight, F. A. (1959). The generalized Poisson distribution. *Ann. Inst. Statist. Math*, 11, 101-105. <http://dx.doi.org/10.1007/BF01737397>
- Haight, F. A., Whilser, B. F., & Mosher, W. W. (1961). New statistical method for describing highway distribution of cars. *Proc. Highw. Res. Board*, 40, 557-563.
- Jabari, E. E., & Liu, H. X. (2012). A stochastic model of traffic flow: Theoretical foundations. *Transportation Research Part B*, 46, 156-174. <http://dx.doi.org/10.1016/j.trb.2011.09.006>
- Johnson, A. N. (1928). Maryland aerial survey of highway traffic between Baltimore and Washington. *Highw. Res. Bd.*, 8, 106-115.
- Kendall, H. G., & Stuart, A. (1967). *The Advanced Theory of Statistics-Vol. II*. London, UK: Griffin.
- Kinzer, J. P. (1933). Application of the Theory of Probability to Problems of Highway Traffic. *Inst. Traff. Engrs.*, 5, 118-124.
- May, A. D. (1990). *Traffic flow fundamentals*. New Jersey, NY: Prentice Hall.
- Mauro, R., & Branco, F. (2012). Two Vehicular Headways Time Dichotomic Models. *Modern Applied Science*, 6(12), 1-12. <http://dx.doi.org/10.5539/mas.v6n12p1>
- Miller, A. M. (1970). An empirical model for multilane road traffic. *Transp. Sci.*, 4(2), 164-186. <http://dx.doi.org/10.1287/trsc.4.2.164>
- Oliver, R. M., & Thibault, B. (1962). A highflow traffic counting distribution. *Highw. Res. Board Bull.*, 356.
- Ord, J. K. (1972). *Families of frequency distribution*. London, UK: Griffin.
- Serfling, R. J. (1969). Non-Poisson models for traffic flow. *Transp. Res.*, 3(3), 299-306. [http://dx.doi.org/10.1016/S0041-1647\(69\)80015-8](http://dx.doi.org/10.1016/S0041-1647(69)80015-8)
- Wald, A. (1947). *Sequential Analysis*. New York, NY: J. Wiley.
- Weiss, G. H., & Hermann, R. (1962). Statistical properties of low-density traffic. *Q. Appl. Math.*, 20, 121-130.