ENGINEER - Vol. XXXXIII, No. 01, pp. [16-21], 2010

Selection of an Appropriate 'Statistical Model' for Baseline Road for Low Volume Traffic Flow

K. S. Weerasekera

Abstract: It was attempted to select an appropriate 'statistical model' for a set of observed traffic data on Baseline road, Colombo. A traffic counting survey was conducted to collect traffic data along Baseline road on a weekend. A Sunday afternoon was selected to conduct the traffic survey to ensure an uninterrupted flow condition without congestion at the site. Data obtained from the survey was analysed and appropriate counting interval and time period for the development of the model were selected.

A counting interval of 10 seconds was conducted, and later a spectrum of 10, 20 & 30 second intervals were computed and plotted. It was observed that 10 second counting interval gave the best shape graph with a maximum frequency at one vehicle per 10 second interval and a positive skew. The observed frequency distribution for 10 second interval, showed a reasonably well distribution with a single mode, and a positive skew tailing to the right. The 20 and 30 second intervals showed a deviation from the regular pattern further and further. Hence the pattern seems to deteriorate with the increasing time interval.

The analysis proved that the sample could be identified as Negative Binomially distributed. A goodness-of-fit test was used to check the difference between field data and expected frequency. Negative Binomial Model (with k=2) gave the smallest Chi-square value with parameters of p = 0.38, which best fitted to the obtained sample data. Hence it could be concluded that Baseline road traffic was 'randomly' distributed under uninterrupted flow conditions, and Negative Binomial Model was the statistical model that best described the traffic flow conditions at the site at the time.

Keywords: Traffic Flow Modelling, Negative Binomially Distribution

1. Introduction

The aim of this study is to select an appropriate 'statistical model' for a set of observed traffic data on Baseline road, Colombo. A traffic counting survey was conducted to collect traffic data along Baseline road on a weekend. Sunday the 18th January 2009 afternoon was selected to conduct the traffic survey to ensure an uninterrupted flow condition without congestion at the site. Data obtained from the survey was analysed and appropriate counting interval and time period for the development of the model were selected.

2. Data Collection

The counting site (Figure 1) was selected on Baseline road at the Southern end of the flyover (across the main railway line). This section of the road is a divided road with uninterrupted traffic flow without any side friction from roadside activities. The traffic survey was carried out on a Sunday afternoon (2:00pm to 3:00pm) for duration of 60 minutes. The count was performed on north-bound middle lane traffic which seemed to be fairly steady.

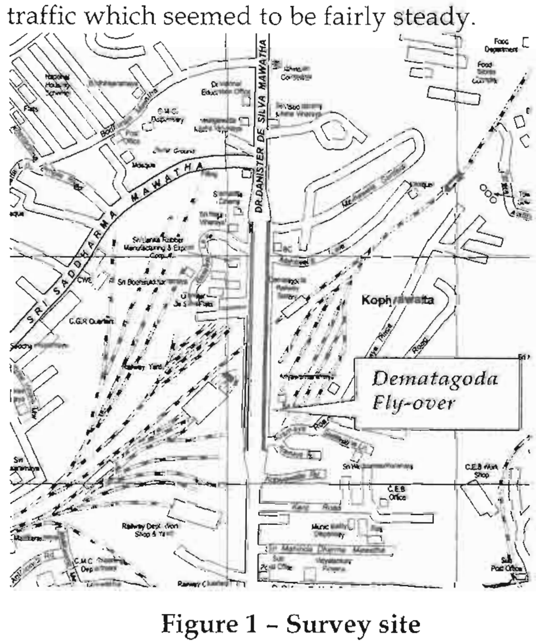

Figure 1 - Survey site

Eng. (Prof.) K. S. Weerasekera, BSc Eng (Moratuwa), MEngSc (UNSW), PhD (UNSW), FIE (Sri Lanka), CEng, MIE (Aust), CPEng, MIHT (UK), MASCE, Professor in Civil Engineering, The Open University of Sri Lanka Lanka.

Number of vehicles travelling in the middle lane towards north (i.e., from Borella to Urugodawatta direction) was recorded at 10 second intervals (Appendix 1). When deciding on the counting interval, it was ensured that the interval was such that it was humanly possible to record them at the field, and also able to manipulate the data to obtain a spectrum of intervals (10 seconds, 20 seconds, 30 seconds etc.).

3. Data Analysis

The frequency distributions were plotted for a spectrum of counting intervals of 10 seconds, 20 seconds, and 30 seconds as illustrated in Figures, 2, 3, and 4.

The observed frequency distribution for 10 second interval shows reasonably well a distribution with a single mode, and a positive skew tailing to the right (Figure 2). It can be seen from Figures 3 and 4, for 20 second interval and 30 second interval, shapes deviate from the regular pattern further and further. Hence the pattern seems to deteriorate with the increasing counting interval.

Out of the three frequency distribution graphs, the 10 second interval graph (Figure 2) which had the best shape with a positive skew and a maximum frequency, was selected for the development of a counting model.

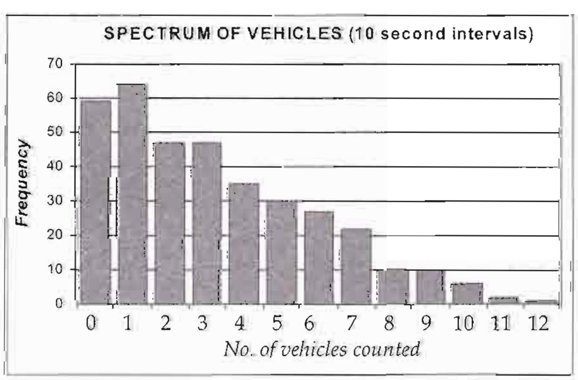

Figure 2 - Spectrum of vehicles at 10 second intervals

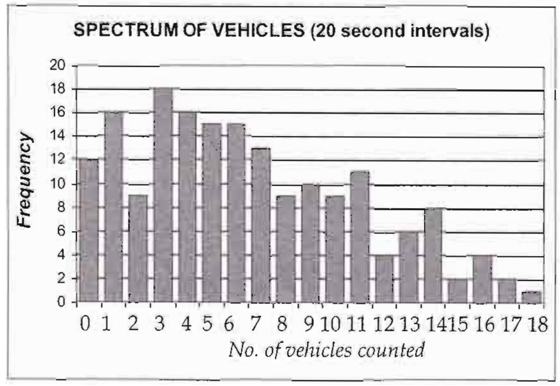

Figure 3 - Spectrum of vehicles at 20 second intervals

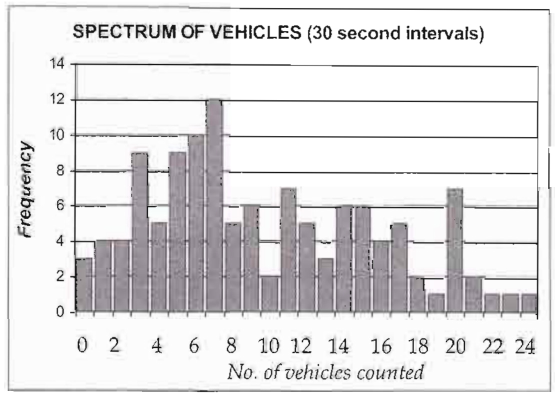

Figure 4 - Spectrum of vehicles at 30 second intervals

Statistical analysis for the selected time period (2:00pm to 3:00pm) at 10 second counting interval is indicated in Table 1.

Table 1 - Statistical Analysis

Count (x)	Frequency (f_x)	$x.f_x$	$x^2.f_x$
0	59	0	0
1	64	64	64
2	47	94	188
3	47	141	423
4	35	140	560
5	30	150	750
6	27	162	972
7	22	154	1078
8	10	80	640
9	10	90	810
10	6	60	600
11	2	22	242
12	1	12	144
> 12	0	0	0
TOTAL	**360**	**1169**	**6471**

3.1 Statistics:

Sample mean $= \bar{x} = \dfrac{\sum x.f_x}{\sum f_x}$

$= 1169 / 360 = 3.247$

Mean of square $= \overline{x^2} = \dfrac{\sum x^2.f_x}{\sum f_x}$

$= 6471 / 360 = 17.975$

Sample variance $= s^2 = \overline{x^2} - (\bar{x})^2$
$= 17.975 - (3.247)^2 = 7.432$

Standard deviation $= s = (7.432)^{1/2} = 2.726$

Index of dispersion $= I = \dfrac{s^2}{\bar{x}} = 7.432 / 3.247$
$= 2.289$

Since, $\bar{x} = 3.247$ veh/10 seconds;

traffic volume per hour = 3.247 x 6 x 60
= 1169 veh/hr.

3.2 Choice of Model

Index of dispersion (I) value was used for the selection of the model. If, I = 1 or approximately equal to 1, Poisson model may sometimes be suitable, but not always (Adams, [1]). If, I < 1, Binomial model may be suitable. This has been further justified by Mannering *et al.* [3].

As for the obtained data, I = 2.289 which is greater than 1, Negative Binomial distribution is analysed for selection, as suggested by Gosling[2] and Walpole and Myers [4].

Negative Binomial Distribution

$$p(X = x) = \binom{x+k-1}{k-1} p^k (1-p)^x \qquad x = 0, 1, 2, \ldots\ldots$$

$$p_{x+1} = \left(\frac{x+k}{x+1}\right)(1-p)\, p_x \qquad x \geq 0$$

Estimated the following parameters,

$\hat{p} = I^{-1} = 1/2.289 = 0.4369$

$$\hat{k} = \frac{\bar{x}}{I-1} = \frac{3.247}{(2.289-1)} = 2.5190$$

Nearest integers are $\hat{k} = 2$ and $\hat{k} = 3$

Let's try, $\hat{k} = 2$ and $\hat{k} = 3$

(1) <u>when, k = 2</u>

Hence,

$$\hat{p} = \frac{k}{\bar{x}+k} = \frac{2}{(3.247+2)} = 0.3812$$

$$p_{x+1} = \left(\frac{x+k}{x+1}\right)(1-p)\, p_x \qquad x \geq 0$$

$p_0 = p^k = (0.3812)^2 = 0.1453$

$e_0 = 360 \times p_0 = (360) \times 0.1453$

e_x - expected frequency
f_x - observed frequency

Calculation and results of e_x and χ_x^2 is given in Table 2.

(1) <u>when, k = 3</u>

Hence,

$$\hat{p} = \frac{k}{\bar{x}+k} = \frac{3}{(3.247+3)} = 0.4802$$

$$p_{x+1} = \left(\frac{x+k}{x+1}\right)(1-p)\, p_x \qquad x \geq 0$$

$p_0 = p^k = (0.4802)^3 = 0.1107$

$e_0 = 360 \times p_0 = (360) \times 0.1107$

e_x - expected frequency
f_x - observed frequency

Calculation and results of e_x and χ_x^2 is given in Table 2.

4. Results

Table 2 - Chi-square Test Analysis for k = 2 and k = 3

Count (x)	Freq. (f_x)	k=2		k=3	
		e_x	$(e_x - f_x)^2/e_x$	e_x	$(e_x - f_x)^2/e_x$
0	59	52.31	0.856	39.86	9.191
1	64	64.74	0.008	62.15	0.055
2	47	60.08	2.848	64.60	4.795
3	47	49.57	0.133	55.95	1.432
4	35	38.34	0.291	43.62	1.703
5	30	28.47	0.082	31.75	0.096
6	27	20.56	2.017	22.01	1.131
7	22	14.54	3.827	14.70	3.625
8	10	10.12	0.001	9.54	0.022
9	10	6.96	1.328	6.06	2.562
10	6	4.73	0.341	3.77	1.319
11	2	3.18	0.438	2.32	0.045
12	1	2.17	0.631	1.41	0.118
> 12	0	4.23	4.230	2.26	2.265
Σ	**360**		**17.032**		**28.360**

From Table 2 results it is evident that when k = 2, $\sum (e_x - f_x)^2 / e_x \;=\; 17.032$

and,

when k = 3, $\sum (e_x - f_x)^2 / e_x \;=\; 28.360$

Since when k = 2 $\sum (e_x - f_x)^2 / e_x$ has a lower value than when k = 3.

Since Negative Binomial Model (with k = 2) gives the smaller value of $\sum \chi_x^2 \;=\; 17.032$

Therefore the curve when k = 2 and p = 0.3812 is a better fit among the two curves.

5. Conclusion

From the results it can be clearly seen that road traffic is 'randomly' distributed under uninterrupted conditions. The particular counting time interval which demonstrates this randomness is 10 second intervals, and the vehicular flow is around 1170 veh/hr.

Therefore the chosen Negative Binomial Model is as follows:

$$p(X = x) \;=\; \binom{x+k-1}{k-1} p^k \, (1-p)^x$$

x = 0, 1, 2, ...

$$p(X = x) \;=\; \binom{x+1}{1} (0.3812)^2 \, (1-0.3812)^x$$

$$p(X = x) \;=\; \mathbf{0.1453\,(x+1) \times 0.6188^x}$$

It was seen that the frequency distribution for counting interval of 10 seconds represented a much improved shape of a graph and the 20 second and 30 second intervals deviated further and further from the regular pattern. It also gave the impression that 30 second interval sketch as developing towards a bi-modal shape. Therefore the regular pattern seemed to deviate from the expected shape, with increasing counting interval.

References

1. Adams, W. F., 'Road Traffic Considered as a Random Series', *Journal of the Institution of Civil Engineers*, Vol. 4, November, 1936, pp. 121-130.

2. Gosling, J. *'Introductory Statistics'*, Pascal Press, Glebe, NSW, Australia, 1995.

3. Mannering, F. L. , Kilareski, W. P. and Washburn, S. S. *Principles of Highway Engineering and Traffic Analysis*, 3rd ed., John Wiley & Sons, Inc. USA, 2005, pp. 146-147.

4. Walpole, R.E. and Myers, R.H. *'Probability and Statistics for Engineers and Scientists'*, Maxwell Macmillan International Editions, 1990.

APPENDIX A - Field Data Sheet

Time	10s interval	No. of veh.	Time	10s interval	No. of veh.	Time	10s interval	No. of veh.	Time	10s interval	No. of veh.
2:00pm	10	3	2:09pm	10	5	2:18pm	10	3	2:27pm	10	7
	20	4		20	1		20	2		20	2
	30	6		30	2		30	3		30	3
	40	3		40	2		40	0		40	3
	50	0		50	2		50	2		50	2
	60	0		60	3		60	3		60	1
2:01pm	10	2	2:10pm	10	7	2:19pm	10	5	2:28pm	10	3
	20	7		20	2		20	8		20	1
	30	9		30	6		30	7		30	3
	40	2		40	8		40	7		40	4
	50	4		50	5		50	5		50	2
	60	3		60	3		60	1		60	11
2:02pm	10	5	2:11pm	10	4	2:20pm	10	0	2:29pm	10	8
	20	6		20	1		20	1		20	5
	30	6		30	1		30	2		30	4
	40	4		40	2		40	1		40	2
	50	4		50	0		50	4		50	0
	60	7		60	2		60	3		60	1
2:03pm	10	2	2:12pm	10	3	2:21pm	10	4	2:30pm	10	3
	20	1		20	2		20	3		20	3
	30	2		30	2		30	4		30	0
	40	6		40	6		40	0		40	1
	50	4		50	5		50	0		50	4
	60	5		60	1		60	0		60	6
2:04pm	10	2	2:13pm	10	10	2:22pm	10	0	2:31pm	10	7
	20	1		20	10		20	1		20	9
	30	3		30	0		30	0		30	4
	40	5		40	2		40	0		40	1
	50	8		50	3		50	5		50	1
	60	7		60	1		60	11		60	1
2:05pm	10	1	2:14pm	10	0	2:23pm	10	5	2:32pm	10	0
	20	1		20	1		20	5		20	1
	30	1		30	1		30	4		30	1
	40	4		40	4		40	1		40	1
	50	4		50	0		50	1		50	7
	60	3		60	9		60	1		60	4
2:06pm	10	3	2:15pm	10	7	2:24pm	10	3	2:33pm	10	6
	20	0		20	4		20	3		20	2
	30	2		30	4		30	3		30	9
	40	3		40	0		40	2		40	6
	50	3		50	1		50	5		50	0
	60	5		60	0		60	2		60	0
2:07pm	10	4	2:16pm	10	2	2:25pm	10	4	2:34pm	10	0
	20	3		20	2		20	10		20	2
	30	0		30	3		30	3		30	2
	40	5		40	0		40	1		40	1
	50	3		50	9		50	1		50	5
	60	8		60	5		60	3		60	3
2:08pm	10	2	2:17pm	10	4	2:26pm	10	3	2:35pm	10	0
	20	1		20	0		20	1		20	1
	30	0		30	1		30	10		30	8
	40	3		40	2		40	4		40	6
	50	2		50	0		50	7		50	1
	60	7		60	0		60	9		60	0

APPENDIX A - Field Data Sheet (Contd.)

Time	10s interval	No. of veh.
2:36pm	10	1
	20	0
	30	0
	40	0
	50	2
	60	5
2:37pm	10	12
	20	6
	30	7
	40	3
	50	5
	60	3
2:38pm	10	1
	20	1
	30	2
	40	4
	50	2
	60	2
2:39pm	10	6
	20	6
	30	6
	40	4
	50	5
	60	5
2:40pm	10	1
	20	1
	30	0
	40	0
	50	1
	60	2
2:41pm	10	0
	20	0
	30	0
	40	7
	50	6
	60	8
2:42pm	10	10
	20	3
	30	1
	40	0
	50	0
	60	1
2:43pm	10	3
	20	7
	30	5
	40	6
	50	2
	60	3
2:44pm	10	6
	20	0
	30	0
	40	0
	50	4
	60	2
2:45pm	10	4
	20	7
	30	3
	40	5
	50	5
	60	6
2:46pm	10	6
	20	7
	30	9
	40	3
	50	0
	60	1
2:47pm	10	1
	20	0
	30	4
	40	5
	50	1
	60	5
2:48pm	10	2
	20	5
	30	8
	40	4
	50	1
	60	2
2:49pm	10	1
	20	3
	30	2
	40	1
	50	4
	60	0
2:50pm	10	1
	20	2
	30	2
	40	8
	50	6
	60	5
2:51pm	10	5
	20	1
	30	0
	40	0
	50	0
	60	0
2:52pm	10	10
	20	1
	30	1
	40	6
	50	9
	60	9
2:53pm	10	6
	20	0
	30	1
	40	2
	50	1
	60	4
2:54pm	10	0
	20	4
	30	5
	40	7
	50	7
	60	6
2:55pm	10	9
	20	7
	30	4
	40	0
	50	1
	60	6
2:56pm	10	6
	20	0
	30	2
	40	1
	50	2
	60	7
2:57pm	10	6
	20	8
	30	7
	40	3
	50	0
	60	0
2:58pm	10	3
	20	6
	30	1
	40	0
	50	1
	60	4
2:59pm	10	3
	20	1
	30	0
	40	1
	50	3
	60	3
3:00pm		