NT 206

Distribuições Estatísticas aplicadas ao tráfego

Engº Sun Hsien Ming

1. Introdução

Durante os trabalhos para desenvolver o Manual de Critérios de Implantação de Semáforos, houve a necessidade de realização de várias pesquisas de campo com o intuito de verificar se a chegada de veículos em uma seção da via obedecia a alguma distribuição teórica.

Uma distribuição teórica é uma função $f(x)$ que apresenta as seguintes propriedades:

a) $f(x) \geq 0$ (1)

b) $\int_{\alpha}^{\beta} f(x)dx = 1$ (2)

c) $\int_{a}^{b} f(x)dx = P(a < x < b)$ (3)

onde α e β são os limites de domínio da função $f(x)$, $\alpha \leq a < b \leq \beta$ e $P(a < x < b)$ é a probabilidade de x estar entre a e b.

As distribuições testadas foram: Poisson, Binomial, Binomial-Negativa (ou Distribuição de Pascal) e a Poisson Generalizada.

Se o comportamento de tráfego fosse aproximadamente igual ao preconizado por alguma distribuição teórica, seria possível elaborar critérios para a implantação de semáforos utilizando-se das expressões matemáticas fornecidas pela distribuição teórica, reduzindo-se assim o volume de pesquisas de campo para a obtenção de dados que expressassem o comportamento real do tráfego. Em outras palavras, a ideia era verificar a possibilidade de se fazer uma simplificação substituindo o comportamento real do tráfego pelo comportamento dado pela distribuição teórica. Assim, o objetivo do estudo proposto é o de verificar se havia condições de se estabelecer alguma relação entre as características de tráfego e distribuições teóricas, como por exemplo, fluxo aleatório com a distribuição de Poisson, flutuações cíclicas com a distribuição Binomial Negativa, fluxo forçado com a distribuição de Poisson Generalizada, etc.

O objetivo deste trabalho é apresentar a metodologia utilizada nesse estudo e os seus principais resultados.

2. Pesquisas de campo realizadas

Foram realizadas pesquisas de chegadas de veículos em 6 locais, os quais estão relacionados no quadro abaixo.

Local	Mão de direção	Aproximação pesquisada	Data da pesquisa	Horário da pesquisa
Teodoro Sampaio x Cristiano Viana	Única	1	12/04/2000	14:00 – 15:00
Almirante Pereira Guimarães	Dupla	2	13/04/2000	13:00 – 14:00
Pedroso de Morais x Pça. dos Omaguás	Única	1	15/04/2000	13:10 – 14:10
Cerro corá x Pça. Silvestre Rabelo	Dupla	2	17/04/2000	13:45 – 14:45
Carlos Lacerda x Nainpur	Dupla	2	18/04/2000	17:15 – 18:15
Ida Kolb x Horácio Vergueiro Rudge	Dupla	2	20/04/2000	17:15 – 18:15

Os locais foram escolhidos com o intuito de contemplar uma diversidade de situações tais como intensidades de fluxo, existência de semáforos próximos, número de faixas de tráfego, etc. que pudessem representar características de tráfego como tráfego aleatório, flutuações cíclicas, fluxo forçado, etc.

A pesquisa consistiu em uma contagem de veículos a cada intervalo de 5 segundos, feita numa seção da via. Essa contagem foi realizada por 2 pesquisadores em cada aproximação: 1 contando em voz alta o número de veículos que passa pela seção e outro anotando o resultado e conferindo o cronômetro.

Com base nos dados obtidos, foram feitos vários agrupamentos resultando em distribuições de intervalos de várias durações: 5, 10, 15 e 30 segundos. Foram testadas cada uma das 4 distribuições teóricas (Poisson, Binomial, Binomial-Negativa e a Poisson Generalizada), para cada local, para cada aproximação e para cada duração de intervalo. Além disso, para cada local de mão dupla testou-se também a distribuição correspondente à soma dos dois sentidos, o que resultou em um total de 4 distribuições x 4 durações de intervalo x (10 aproximações + 4 somas dos 2 sentidos) = 224 verificações. Os resultados estão mostrados na Seção 9.

3. Teste de aderência

O teste de aderência utilizado para determinar se os dados pesquisados em campo seguem alguma das distribuições aqui citadas foi a distribuição χ^2**,** com nível de confiança de 95%.

Os valores de χ^2 são obtidos pela expressão (4).

$$\chi^2 = \sum_{i=1}^{n} \frac{(f_i)^2}{F_i} \qquad (4)$$

A expressão (4) também pode ser escrita como:

$$\chi^2 = \sum_{i=1}^{n} \frac{(f_i - F_i)^2}{F_i} - n \qquad (5)$$

onde n é o tamanho da mostra, f_i é a frequência observada na amostra e F_i é a frequência teórica esperada na distribuição, dada pela expressão (6):

$$F_i = nf(x_i) \qquad (6)$$

Com uma probabilidade $P = 0{,}05$ (95% de confiança) obtém-se em tabelas estatísticas (com grau de liberdade υ), um valor $\chi_P^2 = \chi_{0,05}^2$. A amostra segue a distribuição $f\ (x)$ se:

$$\chi^2 < \chi_P^2 \qquad (7)$$

O grau de liberdade υ é dado pela expressão (8).

$$\upsilon = (g - 1) - A \qquad (8)$$

onde g é o número de grupos e A é o número de parâmetros estimados no processo.

Os valores de A e de υ podem ser sintetizados no quadro abaixo:

Distribuição	$\boldsymbol{A}$	$\boldsymbol{\upsilon}$
Poisson	1	$g - 2$
Binomial Negativa	2	$g - 3$
Binomial	2	$g - 3$
Poisson Generalizada	2	$g - 3$

Os valores de F_i devem ser maiores que 5. Para valores menores que 5, os mesmos devem ser agrupados, diminuindo o valor de g.

4. Distribuição de Poisson

Dada uma amostra, sejam m e s^2 a sua média e a sua variância, respectivamente. Na distribuição de Poisson, tem-se:

$$s^2 = m \tag{9}$$

A distribuição de Poisson é dada pela expressão (10):

$$f(x) = \frac{m^x e^{-m}}{x!} \tag{10}$$

5. Distribuição Binomial Negativa

Na distribuição de Binomial Negativa, tem-se:

$$s^2 > m \tag{11}$$

A distribuição Binomial Negativa é calculada pela expressão (12):

$$f(x) = C_{k-1}^{x+k-1} p^k q^x \tag{12}$$

onde:

$$C_{k-1}^{x+k-1} = \begin{bmatrix} x+k-1 \\ k-1 \end{bmatrix} = \frac{(x+k-1)!}{(k-1)!x!} \tag{13}$$

$$p = \frac{m}{s^2} \tag{14}$$

$$k = \frac{m^2}{s^2 - m} \tag{15}$$

$$q = (1 - p) \tag{16}$$

Como o valor de k dado pela expressão (15) não é um número inteiro, não é possível calcular o valor de $f(x)$ utilizando diretamente a expressão (12) por envolver fatorial de números não inteiros. Para contornar esse problema foram utilizadas as seguintes expressões recorrentes:

$$f(0) = p^k \tag{17}$$

$$f(x+1) = \frac{x+k}{x-1} q f(x) \tag{18}$$

6. Distribuição Binomial

Na distribuição Binomial tem-se:

$$s^2 < m \tag{19}$$

A distribuição Binomial é definida pela expressão (20):

$$f(x) = C_x^n p^x q^{n-x} \tag{20}$$

onde:

$$C_x^n = \begin{bmatrix} n \\ x \end{bmatrix} = \frac{n!}{(n-x)!x!} \tag{21}$$

$$p = \frac{m-s^2}{m} \tag{22}$$

$$q = (1-p) \tag{23}$$

Para facilitar o cálculo, é possível usar as seguintes expressões recorrentes:

$$f(0) = q^n \tag{24}$$

$$f(x+1) = \frac{pn-x}{qx+1} f(x) \tag{25}$$

7. Distribuição de Poisson Generalizada

Na distribuição de Poisson Generalizada, tem-se:

$$s^2 < m \tag{26}$$

Cada termo da distribuição de Poisson Generalizada consiste da soma de k termos de uma distribuição de Poisson simples, isto é:

$$f(x) = \sum_{i=1}^{k} \frac{e^{-\lambda}\lambda^{xk-1+i}}{(xk-1+i)!} \tag{27}$$

Por exemplo:

Para $k = 2$:

$$f(0) = e^{-\lambda} + e^{-\lambda}\lambda$$

$$f(1) = \frac{e^{-\lambda}\lambda^2}{2!} + \frac{e^{-\lambda}\lambda^3}{3!}$$

$$f(2) = \frac{e^{-\lambda}\lambda^4}{4!} + \frac{e^{-\lambda}\lambda^5}{5!}$$

e assim por diante.

Para $k = 3$:

$$f(0) = e^{-\lambda} + e^{-\lambda}\lambda + \frac{e^{-\lambda}\lambda^2}{2!}$$

$$f(1) = \frac{e^{-\lambda}\lambda^3}{3!} + \frac{e^{-\lambda}\lambda^4}{4!} + \frac{e^{-\lambda}\lambda^5}{5!}$$

e assim por diante.

Na expressão (27), o valor de λ é dado por:

$$\lambda = km + \frac{k-1}{2} \tag{28}$$

É claro que para $k = 1$, a expressão (27) resulta na distribuição de Poisson simples.

Apesar de k ser o número de parcelas que compõem cada valor de $f(x)$, isto é, $f(x)$ é uma soma de k parcelas, na expressão generalizada de Poisson k não é um número inteiro.

Dessa forma, a determinação do valor de k não é tarefa simples. Uma das maneiras de se encontrar o valor de k, identificadas na bibliografia pesquisada, é através de nomogramas.

Entretanto, esse método é difícil de ser incorporado numa planilha Excel para que os cálculos sejam feitos de forma automática. Para contornar esse problema, foi solicitado ao consultor Pedro Szasz um método aproximado para calcular $f(x)$ e que pudesse ser feito de forma automática em Excel. A solução encontrada foi definir uma aproximação linear para k diferente de inteiro, conforme exemplo abaixo.

Se denotarmos por:

$$P_0 = e^{-\lambda}$$

$$P_1 = \lambda e^{-\lambda}$$

$$P_2 = \frac{\lambda^2 e^{-\lambda}}{2!}$$

$$P_3 = \frac{\lambda^3 e^{-\lambda}}{3!}$$

e assim por diante, tem-se que:

para $k = 2$:

$$f(0) = P_0 + P_1$$

$$f(1) = P_2 + P_3$$

$$f(2) = P_4 + P_5$$

e assim por diante;

para $k = 2{,}2$:

$$f(0) = P_0 + P_1 + 0{,}2P_2$$

$f(1) = 0{,}8P_2 + P_3 + 0{,}4P_4$

$f(2) = 0{,}6P_4 + P_5 + 0{,}6P_6$

e assim por diante.

Com esta simplificação foi criada uma planilha para calcular os valores de s^2. Os valores de $f(x)$ são calculados de forma que o valor de s^2 calculado fique o mais próximo possível do valor de s^2 da amostra. Isto é feito através da ferramenta "Solver" do Excel, minimizando o quadrado da diferença e fazendo variar os valores de k e m.

8. Exemplos das pesquisas realizadas

A seguir apresentam-se dois exemplos, um onde a amostra não obedeceu à distribuição teórica e outro onde a amostra moldou-se a uma distribuição.

Exemplo 1:

Distribuição de Poisson – Rua Almirante Pereira Guimarães – aproximação Sumaré/Pacaembu – Intervalos de 5 segundos

x_i	f_i	$f_i x_i$	x_i^2	$f_i x_i^2$	$f(x_i)$	F_i	f_i^2/F_i
0	155	0	0	0	0,1920	138,27	173,74
1	173	173	1	173	0,3168	228,15	131,17
2	222	444	4	888	0,2614	188,22	261,83
3	119	357	9	1071	0,1437	103,52	136,78
4	44	176	16	704	0,0593	42,704	45,335
5	4	20	25	100	0,0195	14,092	1,1353
> 6	3	18	36	108	0,0053	5,0175	1,7937
Soma	720				1,0000	719,99	751,80

Os valores de x_i representam número de veículos, f_i é a frequência observada na amostra, enquanto F_i é a frequência teórica esperada pela distribuição. Assim, em 155 intervalos não veio nenhum veículo, em 173 intervalos veio um veículo, em 222 intervalos vieram 2 veículos e assim por diante. Pela distribuição teórica, deveria haver 138 intervalos sem nenhum veículo (em vez de 155 intervalos observados), 228 intervalos com 1 veículo (contra 173 observados), etc.

Efetuando-se os cálculos, onde $f(x_i)$ é a probabilidade de chegar x_i veículos pela distribuição de Poisson, tem-se que:

$m = 1{,}65$
$s^2 = 1{,}505528$
$g = 7$
$v = 5$
$\chi^2 = 31{,}808$
$\chi^2_{0,05} = 11{,}1$

Como $\chi^2 > \chi^2_{0,05}$, os dados não obedecem à distribuição de Poisson.

Exemplo 2:

Distribuição Binomial Negativa – Rua Carlos Lacerda – aproximação Campo Limpo/Interlagos – intervalos de 30 segundos

x_i	f_i	$f_i x_i$	x_i^2	$f_i x_i^2$	$f(x_i)$	F_i	f_i^2/F_i
0	55	0	0	0	0,1857	44,569	67,871
1	48	48	1	48	0,2437	58,507	39,379
2	45	90	4	180	0,2094	50,266	40,285
3	34	102	9	306	0,1482	35,586	32,484
4	25	100	16	400	0,0937	22,502	27,774
5	18	90	25	450	0,0550	13,209	24,528
6	8	48	36	288	0,0306	7,3541	8,7026
> 7	5	35	49	245	0,0164	8,0037	6,1221
Soma	240				1,0000	240	247,14

Sendo $f(x_i)$ a probabilidade de chegar x_i veículos pela distribuição Binomial Negativa, os resultados dos cálculos são:

$m = 2{,}20833$
$s^2 = 3{,}71493$
$p = 0{,}59445$
$k = 3{,}23692$
$q = 0{,}40555$
$g = 8$
$v = 5$
$\chi^2 = 7{,}1483$
$\chi^2_{0,05} = 11{,}1$

Como $\chi^2 < \chi^2_{0,05}$, os dados obedecem à distribuição Binomial Negativa.

9. Resultados Obtidos

Os quadros abaixo sintetizam os resultados obtidos para intervalos de 5, 10, 15 e 30 segundos, onde:

m = média
s^2 = variância
P = distribuição de Poisson
BN = distribuição Binomial Negativa
B = distribuição Binomial
PG = distribuição de Poisson Generalizada
N = dados de pesquisa não se amoldaram ao comportamento previsto pela distribuição
S = dados de pesquisa se amoldaram ao comportamento previsto pela distribuição

Intervalos de 5 segundos

LOCAL	m	s^2	P	BN	B	PG
Almirante A: PA-SU	1,6500	1,5053	N	N	N	N
Almirante: Soma	3,2986	2,9261	N	N	N	N
Almirante B: SU-PA	1,6486	1,5140	N	N	N	N
Carlos Lacerda A: CL-IT	0,7361	0,9887	N	N	N	N
Carlos Lacerda B: IT-CL	0,6792	0,8873	N	N	N	N
Carlos Lacerda: Soma	1,4153	1,8928	N	N	N	N
Cerro Corá A: B-C	1,6690	1,5344	N	N	N	N
Cerro Corá B: C-B	1,3583	1,2383	N	N	N	N
Cerro Corá: Soma	3,0292	2,7210	S	N	S	S
Ida Kolb A: B-C	1,0167	0,7108	N	N	S	N
Ida Kolb B: C-B	1,1153	0,7909	N	N	S	N
Ida Kolb: Soma	2,1319	1,5229	N	N	S	S
Pedroso de Morais	2,6528	5,4628	N	N	N	N
Teodoro Sampaio	2,5708	2,7228	N	N	N	N

Intervalos de 10 segundos

LOCAL	m	s^2	P	BN	B	PG
Almirante A: PA-SU	3,3000	3,5711	N	N	N	N
Almirante: Soma	6,5972	6,8683	S	S	N	N
Almirante B: SU-PA	3,2972	3,9033	N	N	N	N
Carlos Lacerda A: CL-IT	1,4722	2,3159	N	N	N	N
Carlos Lacerda B: IT-CL	1,3583	2,0522	N	N	N	N
Carlos Lacerda: Soma	2,8306	4,5518	N	N	N	N
Cerro Corá A: B-C	3,3667	4,5044	N	N	N	N
Cerro Corá B: C-B	2,7167	3,0642	N	N	N	N
Cerro Corá: Soma	6,0833	7,0153	S	S	N	N
Ida Kolb A: B-C	2,0333	1,5378	N	N	N	N
Ida Kolb B: C-B	2,2306	1,8663	N	N	S	S
Ida Kolb: Soma	4,2639	3,3887	S	N	S	S
Pedroso de Morais	5,3056	15,6844	N	N	N	N
Teodoro Sampaio	5,1417	5,9938	N	N	N	N

Intervalos de 15 segundos

LOCAL	m	s^2	P	BN	B	PG
Almirante A: PA-SU	4,9500	6,2975	N	N	N	N
Almirante: Soma	9,8958	12,518	N	S	N	N
Almirante B: SU-PA	4,9458	6,8512	N	N	N	N
Carlos Lacerda A: CL-IT	2,2083	3,7149	N	S	N	N
Carlos Lacerda B: IT-CL	2,0375	3,3944	N	S	N	N
Carlos Lacerda: Soma	4,2458	7,9771	N	S	N	N
Cerro Corá A: B-C	5,0500	7,6058	N	S	N	N
Cerro Corá B: C-B	4,0750	5,0110	N	N	N	N
Cerro Corá: Soma	9,1250	12,017	N	S	N	N
Ida Kolb A: B-C	3,0500	2,2892	N	N	S	S
Ida Kolb B: C-B	3,3458	3,1346	N	N	N	N
Ida Kolb: Soma	6,3958	5,4225	N	N	N	N
Pedroso de Morais	7,9583	28,414	N	N	N	N
Teodoro Sampaio	7,7125	10,571	N	S	N	N

Intervalos de 30 segundos

LOCAL	m	s^2	P	BN	B	PG
Almirante A: PA-SU	9,9000	12,756	N	S	N	N
Almirante: Soma	19,697	25,874	N	N	N	N
Almirante B: SU-PA	9,8917	18,079	N	S	N	N
Carlos Lacerda A: CL-IT	4,4167	8,3264	N	S	N	N
Carlos Lacerda B: IT-CL	4,0750	6,7194	N	S	N	N
Carlos Lacerda: Soma	8,4917	16,599	N	S	N	N
Cerro Corá A: B-C	10,100	16,306	N	S	N	N
Cerro Corá B: C-B	8,1500	13,127	N	N	N	N
Cerro Corá: Soma	18,250	28,470	N	N	N	N
Ida Kolb A: B-C	6,1000	5,4400	S	N	S	S
Ida Kolb B: C-B	6,6917	8,3799	S	S	N	N
Ida Kolb: Soma	12,791	14,048	S	S	N	N
Pedroso de Morais	14,776	64,477	N	N	N	N
Teodoro Sampaio	15,425	21,194	N	S	N	N

10. Conclusão

Pelos resultados obtidos, verifica-se que parte das pesquisas apresentou o perfil de uma ou outra distribuição. Entretanto, não se conseguiu estabelecer nenhuma relação entre característica de tráfego com distribuição. Por exemplo, locais com tráfego aparentemente apresentando flutuações aleatórias não seguiram a distribuição de Poisson. Dessa forma, o que se pode concluir de fato é que dado um determinado local a ser estudado, é impossível afirmar, a priori, que o tráfego nesse local segue essa ou outra distribuição. Esse resultado derruba a ideia de assumir alguma distribuição teórica como hipótese, com o objetivo de eliminar a pesquisa de campo. Em outras palavras, chega-se à conclusão de que a pesquisa de campo, para a obtenção dos dados necessários não pode ser dispensada.

Engº: Sun Hsien Ming
CTA 5 – Gerência de Sistemas de Controle de Tráfego – GSC