

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Programa de Pós-Graduação em Estatística

Raquel Ude Braz

**INTERVALO DE CREDIBILIDADE PARA INTENSIDADE DE TRÁFICO EM FILAS
M/M/S UTILIZANDO O ALGORITMO AMOSTRAGEM-REAMOSTRAGEM POR
IMPORTÂNCIA (SIR)**

Belo Horizonte

2022

Raquel Ude Braz

**Intervalo de Credibilidade para Intensidade de Tráfego em filas M/M/s Utilizando
o Algoritmo Amostragem-Reamostragem por Importância (SIR)**

Monografia apresentada ao curso de Especialização em Estatística do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do título de Especialista em Estatística.

Orientador: Prof. Roberto da Costa Quinino

Belo Horizonte

2022

Braz, Raquel Ude.

B827i Intervalo de credibilidade para intensidade de tráfico em filas M/M/s utilizando o algoritmo amostragem reamostragem por importância (SIR) [manuscrito] / Raquel Ude Braz.— 2022.
24.f. il.

Orientador: Roberto da Costa Quinino.

Monografia (especialização) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística.

Referências: f. 20-21.

1. Estatística. 2. Teoria das filas. 3. Levantamentos de trânsito. 4. Análise de intervalos (Matemática). I. Quinino, Roberto da Costa . II. Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística .III. Título.

CDU 519.2 (043)

Ficha catalográfica elaborada pela bibliotecária Belkiz Inez Rezende Costa CRB

6/1510 – Instituto de Ciências Exatas da UFMG



Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística
Programa de Pós-Graduação / Especialização
Av. Pres. Antônio Carlos, 6627 - Pampulha
31270-901 – Belo Horizonte – MG

E-mail: pgest@ufmg.br
Tel: 3409-5923 – FAX: 3409-5924

ATA DO 240º. TRABALHO DE FIM DE CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA DE RAQUEL UDE BRAZ.

Aos onze dias do mês de fevereiro de 2022, às 09:00 horas, com utilização de recursos de videoconferência a distância, reuniram-se os professores abaixo relacionados, formando a Comissão Examinadora homologada pela Comissão do Curso de Especialização em Estatística, para julgar a apresentação do trabalho de fim de curso da aluna **Raquel Ude Braz**, intitulado: *“Intervalo de Credibilidade para Intensidade de Tráfego em filas M/M/s Utilizando o Algoritmo Amostragem-Reamostragem por Importância (SIR)”*, como requisito para obtenção do Grau de Especialista em Estatística. Abrindo a sessão, o Presidente da Comissão, Professor Roberto da Costa Quinino – Orientador, após dar conhecimento aos presentes do teor das normas regulamentares, passou a palavra ao candidato para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa da candidata. Após a defesa, os membros da banca examinadora reuniram-se sem a presença da candidata e do público, para julgamento e expedição do resultado final. Foi atribuída a seguinte indicação: a candidata foi considerada Aprovada condicional às modificações sugeridas pela banca examinadora no prazo de 30 dias a partir da data de hoje por unanimidade. O resultado final foi comunicado publicamente ao candidato pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente Ata, que será assinada por todos os membros participantes da banca examinadora. Belo Horizonte, 11 de fevereiro de 2022.

Roberto da
Costa
Quinino808712
91720

Assinatura de forma digital
por Roberto da Costa
Quinino80871291720
Criado em 2022.02.11
16:16:42 -0500

Prof. Roberto da Costa Quinino (Orientador)
Departamento de Estatística / ISEX / UFMG

Prof. Frederico Rodrigues Borges da Cruz
Departamento de Estatística / ISEX / UFMG

AGRADECIMENTOS

Primeiramente agradeço a Deus por estar sempre presente em minha vida.

Agradeço aos meus pais, Marcos e Denise, e ao meu irmão, Guilherme, por me inspirarem e incentivarem constantemente.

Agradeço ao meu marido, Ricardo, e aos meus filhos, Luiza e Arthur, pelo apoio incondicional e por me proporcionarem concluir mais esta etapa dos meus estudos.

Agradeço aos meus colegas, professores e em especial ao Prof. Roberto da Costa Quinino por sua dedicação, orientação e sabedoria.

Resumo

Em teoria de filas, um dos principais interesses dos pesquisadores é estudar o seu comportamento, seu processo de formação e analisar algumas características de desempenho, tais como, por exemplo, a intensidade do tráfico. Esta monografia visa detalhar a metodologia para obtenção do intervalo de credibilidade para a intensidade de tráfico em filas markovianas finitas com s servidores, denominadas M/M/s, na notação de Kendall. Para tanto utilizaremos o método de Sampling Importance Resampling (SIR) em conjunto com o intervalo de credibilidade de mínima amplitude.

Palavras-chave: Filas markovianas, intensidade de tráfico, Sampling Importance Resampling (SIR), Intervalo de mínima amplitude.

Abstract

One of the major research interests in queuing theory is to study the performance characteristics of the queues, such as the traffic intensity. This paper aims to detail the methodology for obtaining the credibility interval for the traffic intensity in finite Markovian queues with s servers, called M/M/s, in Kendall's notation. For this, we will use the Sampling Importance Resampling (SIR) method together with the minimum amplitude credibility interval.

Key Words: Markovian rows, traffic intensity, minimum amplitude credibility interval

LISTA DE FIGURAS

Figura 1 – Distribuição a posteriori para ρ	17
---	----

LISTA DE TABELAS

Tabela 1 – Valores observados e suas frequências.....	16
--	----

Sumário

1	Introdução.....	10
2	Modelo Probabilístico e Função de Máxima Verossimilhança.....	12
3	Inferência Bayesiana	13
3.1	Sampling Importance Resampling (SIR)	14
4	Exemplo Numérico	16
5	Conclusões e Sugestão de Estudos Futuros	18
	Referências Bibliográficas	19
	Anexo A: Programa R para obtenção do Estimador de Máxima Verossimilhança.....	21
	Anexo B: Programa R para obtenção da distribuição a Posteriori e Intervalo de Credibilidade para ρ	22

1 Introdução

Um dos maiores interesses na pesquisa da teoria das filas é o estudo da performance como uma função de seus parâmetros, como a taxa de chegada (λ), a taxa de serviço (μ), ou ainda a intensidade do tráfico (ρ). O problema matemático em questão é como estimar esses parâmetros usando métodos de inferência estatística. Esta monografia visa detalhar a metodologia para obtenção do intervalo de credibilidade para a intensidade de tráfico em filas markovianas com s servidores, denominadas M/M/s, na notação de Kendall.

A intensidade do tráfico (ρ) é definida pela razão $\rho = \lambda/s\mu$.

Com base em ρ , importantes medidas de performances podem ser derivadas desta razão, como a probabilidade do sistema estar vazio (sem clientes) em um determinado momento (P_0), a estimativa da quantidade de clientes no sistema (L), a estimativa da quantidade de clientes esperando na fila (L_q), a estimativa do tempo de atendimento (W) e a estimativa do tempo na fila (W_q). O modelo de filas M/M/s tem muitas aplicações e conseqüentemente muita necessidade de inferências. Muitas situações cotidianas são descritas como modelos de filas [1], como redes de computação e telecomunicação, linhas de produção na indústria, serviços de saúde, supermercados entre outros.

Os resultados de inferência estatística através de métodos clássicos – baseados na verossimilhança – foram estudados por Clark [2], que forneceu estimativas de verossimilhança para λ e μ em filas markovianas de atendimento único, ou seja, M/M/1.

Mais recentemente, Schurben e Kullkarni [3] provaram que populares medidas de performance infelizmente não apresentam valores esperados e erros padrão em forma

fechada. Basawa e Prahbu [4] relataram algumas propriedades assintóticas dos estimadores para a avaliação de desempenho das filas M/M/1. Zhenga e Seila [5] mostraram que abaixo de um certo limite superior de intensidade de tráfico ($\rho < 1$), os estimadores existem e apresentam as propriedades necessárias, e Almeida et al. [6] propôs métodos de correção de viés de estimadores clássicos para filas M/M/1.

Entretanto, provavelmente devido às inerentes dificuldades computacionais, a inferência Bayesiana não havia alcançado destaque e interesse suficientes até que os resultados pioneiros divulgados por Armero e Bayarri [7-9], que em seguida foram aprimorados por Armero e Conesa [10-13], McGrath e Singpurwalla [14], Choudhury e Borthakur [15] e mais recentemente, Almeida e Cruz [16], Cruz et al [17], Quinino e Cruz [18], Choudhury e Basak [19], entre outros.

Nesse estudo o objetivo é a derivação do estimador Bayesiano e intervalo de credibilidade para ρ em filas M/M/s, a partir de amostras de quantidade de clientes no sistema em momentos suficientemente espaçados.

O trabalho está organizado como se segue. Na seção 2, detalhamos o modelo probabilístico para filas M/M/s e a função de verossimilhança. Na seção 3 explicamos o procedimento da inferência Bayesiana aplicado em filas M/M/s incluindo o intervalo de credibilidade de mínima amplitude. Na seção 4 apresentamos um exemplo numérico. Conclusões e sugestões de trabalhos futuros estão na seção 5.

2 Modelo Probabilístico e Função de Máxima Verossimilhança

O modelo M/M/s parte do pressuposto de que tanto os tempos entre atendimentos quanto os tempos de atendimento possuem uma distribuição exponencial e que o número de atendentes é s (qualquer inteiro positivo).

Em filas M/M/s as chegadas dos clientes pressupõem uma distribuição de Poisson com taxa λ , o que significa que o tempo entre as chegadas sucessivas são independentes e identicamente distribuídas exponencialmente com média $1/\lambda$. Os tempos de atendimento sucessivos são independentes e identicamente distribuídos exponencialmente e aleatoriamente com média $1/\mu$.

Considerando $\rho = \lambda/s\mu$ como a intensidade do tráfego, ou taxa de utilização do sistema, é necessário que $\rho < 1$ para que a fila seja considerada estável, o que é uma consideração comum na teoria das filas. A intensidade do tráfego (ρ) pode também ser interpretada como a média de tempo em que cada atendente está ocupado. Se a intensidade do tráfego é tal que $\rho < 1$, então o sistema está em equilíbrio e apresenta uma distribuição estacionária. Da teoria desenvolvida dos processos de nascimento e morte, a distribuição estacionária da quantidade de clientes N no sistema no momento das saídas de clientes é dada pela Equação 2.1 abaixo:

$$P_n = P(N \equiv n) = \begin{cases} \frac{(s\rho)^n}{n!} P_0, & 0 < n \leq s \\ \frac{s^s \rho^n}{s!} P_0, & n > s \end{cases} \quad (2.1)$$

Sendo $P_0 = P(N = 0)$ é dada pela usual condição de contorno que a soma das probabilidades deve ser 1; ou seja

$$P_0 = \left(\sum_{j=0}^{s-1} \frac{(s\rho)^j}{j!} + \frac{(s\rho)^s}{s!} \frac{1}{(1-\rho)} \right)^{-1} \quad (2.2)$$

Considerando a quantidade de clientes observadas em instantes aleatórios denotada por $x = \{x_1, x_2, \dots, x_n\}$, em que n é o tamanho de instantes aleatórios observados, temos a amostra de tamanho n , então a função de verossimilhança correspondente é dada pela equação 2.3 abaixo

$$L(\rho|x) = \prod_{i=1}^n \left[\frac{(s\rho)^{x_i}}{x_i!} P_0 I_{\{0 \leq x_i \leq s\}} + \frac{s^s \rho^{x_i}}{s!} P_0 I_{\{x_i > s\}} \right] \quad (2.3)$$

Onde I é a função indicadora. O processo de geração de dados garante a independência das observações das amostras, desde que estas estejam suficientemente espaçadas.

3 Inferência Bayesiana

Nesta seção descreveremos brevemente o processo da inferência Bayesiana a ser utilizada nesta monografia. Como $0 < \rho < 1$ e de acordo Gupta e Nadarajah [20], uma priori natural para ρ seria a expressão (3.1), ou seja, proporcional a uma distribuição Beta com parâmetros a e b , sendo indicada por Beta (a, b).

$$\pi(\rho) \propto \rho^{(a-1)}(1-\rho)^{(b-1)}, 0 < \rho < 1, a > 0, b > 0 \quad (3.1)$$

Esse resultado é vantajoso pois a distribuição Beta é conhecida pela sua flexibilidade em modelar diferentes tipos de opiniões a priori sobre proporções. A determinação dos parâmetros a e b para a distribuição a priori é muitas vezes feita indiretamente.

Pode-se determinar os parâmetros por meio dos percentis (por exemplo 5%, 50% e 95%) ou pela média e desvios padrão da distribuição priori da intensidade de tráfico (ρ). Depois, estima-se a e b a partir de expressões matemáticas conhecidas que relacionam a e b com a média e desvio padrão da distribuição priori. Assim, a distribuição a Posteriori é dada pela equação (3.2).

$$\pi(\rho|dados) \propto \rho^{(a-1)}(1-\rho)^{(b-1)} L(\rho|dados)$$

$$\pi(\rho|dados) \propto \rho^{(a-1)}(1-\rho)^{(b-1)} \prod_{i=1}^n \left[\frac{(s\rho)^{x_i}}{x_i!} P_{0I_{\{0 \leq x_i \leq s\}}} + \frac{s^s \rho^{x_i}}{s!} P_{0I_{\{x_i > s\}}} \right] \quad (3.2)$$

A expressão (3.2) é básica para realização do processo inferencial para ρ . Entretanto é necessário multiplicá-la por uma constante de normalização (K) de tal forma que a integral em relação a ρ e variando de 0 a 1 seja igual a 1. Ou seja, a distribuição a posteriori é uma função densidade de probabilidade. A média dessa distribuição a posteriori constituirá em uma estimativa Bayesiana para ρ . Também com esta distribuição a posteriori podemos obter um intervalo de credibilidade para ρ de tal modo que a área entre o limite inferior e superior seja uma probabilidade pré-estabelecida, por exemplo, 0.95. Infelizmente as integrais envolvendo (3.2) não são triviais e assim usaremos nesta monografia resultados aproximados por meio do algoritmo Sampling Importance Resampling (SIR).

3.1 Sampling Importance Resampling (SIR)

Considerando a dificuldade em obter a distribuição probabilística $\pi(\rho|dados)$ dada pela Equação 3.2, é usado o popular método de Monte Carlo chamado Sampling Importance Resampling (SIR). Esse método foi amplamente discutido por Rubin [21] e Smith [22] e Gelfand [23], e seu algoritmo será brevemente descrito a seguir.

Uma amostra aleatória de tamanho k é retirada da distribuição a priori de ρ , sendo que é recomendado que k seja igual a 5.000 ou maior. Os dados são denominados ρ_i , $i=1, 2, \dots, k$. A função de normalização $w_i = 1, 2, \dots, k$ é associada a cada ponto da amostra a priori, onde w_i é proporcional ao lado direito da Equação 3.2. Em seguida, uma amostra aleatória de tamanho k é retirada da densidade a priori de ρ com os pesos w_i . Os pontos são então denominados ρ_i^* , $i=1, 2, \dots, k$ e são aproximadamente distribuídos como uma observação da Equação 3.3. Nesse sentido, pode-se aproximar as medidas usuais da inferência Bayesianas. Por exemplo, o estimador Bayesiano é a média de ρ_i^* . O intervalo de credibilidade com probabilidade $1-\alpha$, de limite superior L_s e limite inferior L_i , é expressado pelos percentis $100(\alpha/2) \%$ e $100(1-\alpha/2) \%$.

O intervalo de credibilidade pode ser aperfeiçoado de tal forma que sua amplitude seja mínima. Para tal usaremos nesta monografia o algoritmo de Chen-Shao.

Passo 1: Obtenha uma amostra SIR, $\{\rho_i, i=1,2, \dots, n\}$ da Equação 3.3.

Passo 2: Ordene $\{\rho_i, i=1,2, \dots, n\}$ de forma crescente:

$$\rho^{(1)} \leq \rho^{(2)} \leq \dots \leq \rho^{(n)}.$$

Passo 3: Registre o intervalo de credibilidade $100(1-\alpha) \%$

$$R_{j(n)} = (\rho_{(j)}, \rho_{(j+[(1-\alpha)n])})$$

para $j=1,2, \dots, n - [(1-\alpha)n]$.

Passo 4: O intervalo de credibilidade $100(1-\alpha) \%$ HPD é denotado por $R_{j(n)}$, com a menor amplitude entre todos os intervalos de credibilidade.

4 Exemplo Numérico

Nosso exemplo numérico se baseia em dados coletados em um supermercado para dar um melhor entendimento da metodologia apresentada no trabalho. Experiências passadas indicam que a intensidade do tráfego não deve ser igual a 90% para não prejudicar a experiência do cliente e conseqüentemente as vendas do supermercado. O objetivo é avaliar a intensidade de tráfego (ρ) de uma fila M/M/s com 6 atendentes (M/M/6).

Foram coletados 100 dados aleatórios da quantidade dos clientes no sistema (em atendimento e em fila). Os valores observados (O) e respectivas frequências (F) estão apresentados na Tabela 1. Por exemplo, em 3 vezes apenas 1 cliente foi encontrado no sistema e em 10 vezes, 5 clientes estavam no sistema.

Tabela 1: Valores observados e suas frequências									
O	F	O	F	O	F	O	F	O	F
1	3	6	9	11	3	16	3	21	1
2	1	7	6	12	3	17	1	22	4
3	8	8	6	13	1	18	2	23	1
4	13	9	6	14	5	19	3		
5	10	10	3	15	4	20	4		

Considerando os dados observado na Tabela 1 e utilizando o programa contido no Anexo B com a implementação do SIR obtemos a distribuição a posteriori de ρ descrita na Figura 1. Neste exemplo utilizamos a distribuição a priori com ausência de informação, assim $a=1$ e $b=1$, na distribuição a priori Beta, o que equivale a uma distribuição a priori uniforme para ρ .

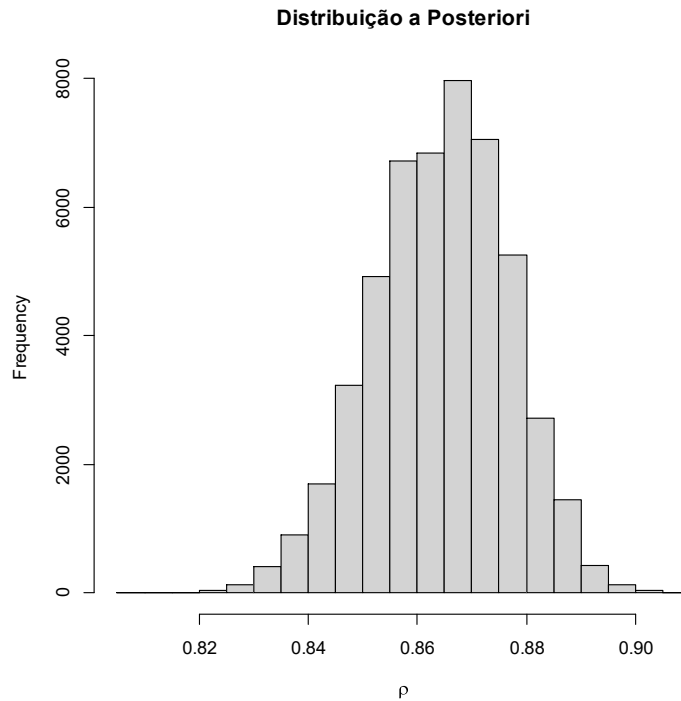


Figura 1: Distribuição a posteriori para ρ

A média a posteriori ($\hat{\rho}$) para ρ foi de 0,8643356 com um intervalo percentílico de credibilidade de 95% dado por [0,8387211; 0,8875731]. O intervalo de credibilidade 95% com amplitude mínima foi [0,8403779; 0,8887925]. Como $\rho = 0,9$ não está contido no intervalo de credibilidade de 95% então podemos rejeitar a hipótese $H_0: \rho = 0,9$. Além disso, como $\hat{\rho} = 0,8643356$ é menor do que 0,9 então não existe evidência para uma preocupação imediata do supermercado. Observe que quando a distribuição a priori de ρ é uma uniforme então a moda a posteriori (outro possível estimador Bayesiano de ρ) coincide com o clássico estimador de máximo verossimilhança que maximiza a expressão (3.1). Utilizando o Anexo A o estimador de máxima verossimilhança para ρ foi 0,8650847 sendo muito próximo da média a posteriori.

5 Conclusões e Sugestão de Estudos Futuros

Método inferencial para intensidade de tráfego utilizando o procedimento de inferência Bayesiana e respectivo intervalo de credibilidade foi utilizado em filas M/M/s com uso do algoritmo SIR. O procedimento apresentado mostrou-se eficiente gerando resultados que ajudam no processo de decisão em sistemas de filas. Em particular aplicamos a metodologia em um exemplo de filas em supermercados.

Como trabalho futuro recomendamos um novo estudo usando filas M/M/s/h para avaliar o impacto do tamanho máximo que o sistema comporta (h) na estimação de ρ .

Referências Bibliográficas

- [1] D. Gross, J. F. Shortle, J. M. Thompson, and C. M. Harris, *Fundamentals of Queueing Theory*, Wiley-Interscience, New York, NY, 4th edition, 2008.
- [2] Clark, A.B. (1957) "Maximum likelihood estimates in a simple queue". *The Annals of Mathematical Statistics*, vol. 28, 1036-1040.
- [3] L. Schruben and R. Kulkarni, "Some consequences of estimating parameters for the M/M/1 queue," *Operations Research Letters*, vol. 1, no. 2, pp. 75–78, 1982
- [4] I. V. Basawa and N. U. Prabhu, "Large sample inference from single server queues," *Queueing Systems*, vol. 3, no. 4, pp. 289– 304, 1988
- [5] S. Zheng and A. F. Seila, "Some well-behaved estimators for the M/M/1 queue," *Operations Research Letters*, vol. 26, no. 5, pp. 231–235, 2000.
- [6] M. A. C. Almeida, F. R. B. Cruz, F. L. P. Oliveira, and G. de Souza, "Bias correction for estimation of performance measures of a Markovian queue," *Operational Research*, Article ID 10.1007/s12351-017-0351-4, pp. 1–20, 2017.
- [7] C. Armero and M. J. Bayarri, "Bayesian prediction in M/M/1 queues," *Queueing Systems*, vol. 15, no. 1-4, pp. 401–417, 1994.
- [8] C. Armero and M. J. Bayarri, "A Bayesian analysis of a queueing system with unlimited service," *Journal of Statistical Planning and Inference*, vol. 58, no. 2, pp. 241–261, 1997.
- [9] C. Armero and M. J. Bayarri, "Dealing with uncertainties in queues and networks of queues: A Bayesian approach," in *Multivariate, Design and Sampling*, S. Ghosh, Ed., pp. 579–608, Marcel Dekker, Inc, New York, 1999.
- [10] C. Armero and D. Conesa, "Inference and prediction in bulk arrival queues and queues with service in stages," *Applied Stochastic Models in Business and Industry*, vol. 14, no. 1, pp. 35– 46, 1998.
- [11] C. Armero and D. Conesa, "Prediction in Markovian bulk arrival queues," *Queueing Systems*, vol. 34, no. 1-4, pp. 327–350, 2000.
- [12] C. Armero and D. Conesa, "Statistical performance of a multiclass bulk production queueing system," *European Journal of Operational Research*, vol. 158, no. 3, pp. 649–661, 2004.
- [13] C. Armero and D. Conesa, "Bayesian hierarchical models in manufacturing bulk service queues," *Journal of Statistical Planning and Inference*, vol. 136, no. 2, pp. 335–354, 2006.
- [14] M. F. McGrath and N. D. Singpurwalla, "A subjective Bayesian approach to the theory of queues II - Inference and information in M/M/1 queues," *Queueing Systems*, vol. 1, no. 4, pp. 335–353, 1987.

- [15] A. Choudhury and A. C. Borthakur, "Bayesian inference and prediction in the single server Markovian queue," *Metrika*, vol. 67, no. 3, pp. 371–383, 2008.
- [16] M. A. C. Almeida and F. R. B. Cruz, "A note on Bayesian estimation of trafrc intensity in single server Markovian queues," *Communications in Statistics - Simulation Computation*, vol. 47, no. 9, pp. 2577–2586, 2017.
- [17] F. R. Cruz, R. C. Quinino, and L. L. Ho, "Bayesian estimation of trafrc intensity based on queue length in a multi-server M/M/s queue," *Communications in Statistics - Simulation and Computation*, vol. 46, no. 9, pp. 7319–7331, 2017.
- [18] R. C. Quinino and F. R. B. Cruz, "Bayesian sample sizes in an M/M/1 queueing systems," *The International Journal of Advanced Manufacturing Technology*, vol. 88, no. 1-4, pp. 995–1002, 2017.
- [19] A. Choudhury and A. Basak, "Statistical inference on trafrc intensity in an M/M/1 queueing system," *International Journal of Management Science and Engineering Management*, vol. 13, no. 4, pp. 274–279, 2018.
- [20] Gupta, A. K., Nadarajah, S. (2004). *Handbook of beta distribution and its applications*, Marcel Dekker, New York, NY.
- [21] Rubin, D. B. (1988). "Using the SIR algorithm to simulate posterior distributions", In Bernardo, J. M.; DeGroot, H. M.; Lindley, D. V. & Smith, A. F. M., eds., *Bayesian Statistics 3*, Oxford University Press, Cambridge, MA, pp. 395-402.
- [22] Ross, S. M. (1996). *Stochastic Processes*, 2th ed., John Wiley & Sons, Inc.
- [23] Smith, A. F. M. & Gelfand, A. E. (1992). "Bayesian statistics without tears: A sampling–resampling perspective", *The American Statistician*, vol. 46, n. 2, pp. 84-88.

Anexo A: Programa R para obtenção do Estimador de Máxima Verossimilhança

```
## Obtenção do Estimador de Máxima Verossimilhança ##

# Amostra de tamanho 100 - entrada de valores observados e respectivas
# frequências
R<-rep(c(1:23),c(3,1,8,13,10,9,6,6,6,3,3,3,1,5,4,3,1,2,3,4,1,4,1))
Ta=size(R)
Ta=Ta[2]
# 6 caixas
s=6
# Estimando via EMV - para encontrar o valor de rho que maximiza a
# expressão de EMV
##1a parte: cálculo de P0 em função de rho - equação 2.2
rf <- function(rhoe)
{
  P0=0
  for(j in 0:(s-1))
  {
    P1=((s*rhoe)^j)/factorial(j)
    P0=P1+P0
  }
  P0=P0+((s^s)*(rhoe^s))/(factorial(s)*(1-rhoe))
  P0=1/P0

##2a parte: função de verossimilhança - produtório da equação 2.3
#obs: E1=1 para não zerar o produtório
  E1=1
  for (i in 1:Ta)
  {
    i1=R[i]
    if (i1<=s) {
      E11=(((s^i1)*((rhoe^i1))/factorial(i1))*P0
      E1=E1*E11
    }else{
      E11=(((rhoe^i1)*(s^s))/factorial(s))*P0
      E1=E1*E11
    }
  }
  VeroA=E1

  return(-log(VeroA))
}
## o estimador de rho é o valor que maximiza a função de verossimilhança e
# é o mesmo valor que maximiza o logaritmo neperiano dessa função. Essa
# propriedade permite simplificar os cálculos, pois ao invés de trabalhar com
# produtório pode-se fazer soma.
##Como no R a função de maximização (optim) faz o cálculo pelo mínimo,
# multiplica-se o resultado por -1 para ter o resultado da maximização de
# rho.
##Na função optim é necessário informar um valor inicial de rho em par=0.5

EMV<- optim(par = 0.5, fn = rf, method = "Brent", lower = 0, upper = 1)

cat('rho estimado via MV oti=',EMV$par,"\n")
```

Anexo B: Programa R para obtenção da distribuição a Posteriori e Intervalo de Credibilidade para ρ

```

#install.packages('HDInterval')
library(pracma)
library(HDInterval)

clear()
## definindo o tamanho da amostra a priori
corridas=50000
alfa=0.05 #A ser usado no Intervalo de Credibilidade
# Amostra de tamanho 100 - entrada de valores observados e respectivas
frequências
R<-rep(c(1:23),c(3,1,8,13,10,9,6,6,6,3,3,3,1,5,4,3,1,2,3,4,1,4,1))
Ta=size(R)
Ta=Ta[2]
s=6
##Função de Verossimilhança
rf <- function(rhoe){
  P0=0
  for(j in 0:(s-1)){
    P1=((s*rhoe)^j)/factorial(j)
    P0=P1+P0
  }
  P0=P0+((s^s)*(rhoe^s))/(factorial(s)*(1-rhoe))
  P0=1/P0
  E1=1
  for (i in 1:Ta) {
    i1=R[i]
    if (i1<=s) {
      E11=(((s^i1)*((rhoe^i1))/factorial(i1))*P0
      E1=E1*E11
    }else{
      E11=(((rhoe^i1)*(s^s))/factorial(s))*P0
      E1=E1*E11
    }
  }
  VeroA=E1
  return(VeroA)
}

##Passo 1 do SIR: Geração da amostra a priori, números aleatórios com
distribuição Beta uniforme(a=b=1)
R1<-rbeta(corridas,1,1)
windows(record = T)
hist(R1,main="Distribuição a Priori - Beta uniforme",xlab=expression(rho))

##Passo 2: Substituir o valor da priori na função de verossimilhança
##Matriz 3 colunas para salvar os dados [Verossimilhança,Pesos, Posteriori]

R2=matrix(0,corridas,3)
for (i in 1:corridas){
  k=R1[i]
  P1=rf(k)
  R2[i,1]=P1
}

##Passo 3: Calcular os pesos
R2[,2]=R2[,1]/sum(R2[,1])
##Passo 4: Cálculo da Posteriori

```

```

R2[,3]=sample(R1,corridas,replace=T,prob=R2[,2])
Posteriori=R2[,3]
##Análise Gráfica

windows(record = T)
hist(Posteriori,main="Distribuição a Posteriori",xlab=expression(rho))

#Determinação do intervalo de credibilidade
LimitesA<-hdi(Posteriori)
LimitesB=quantile(Posteriori,c(0.025,0.975))
cat('Limites HPI - pacote do R',"\n")
cat('Limite Inferior de Credibilidade - 95% =',LimitesA[1],"\n")
cat('Limite Superior de Credibilidade - 95% =',LimitesA[2],"\n")
cat('Limites Percentílico',"\n")
cat('Limite Inferior de Credibilidade - 95% =',LimitesB[1],"\n")
cat('Limite Superior de Credibilidade - 95% =',LimitesB[2],"\n")

#Chen-Shao Estimation Algorithm - HDI
##Encontrando o intervalo de credibilidade com menor amplitude possível
Posteriori=sort(Posteriori)
jmax=corridas-ceil((1-alfa)*corridas)
Intervalos<-matrix(0,jmax,3)
for (i in 1:jmax){
  j1=i
  j2=j1+ceil((1-alfa)*corridas)
  Intervalos[i,1]=Posteriori[j1]
  Intervalos[i,2]=Posteriori[j2]
}
Intervalos[,3]=(Intervalos[,2]-Intervalos[,1])
##ordenar os intervalos
Intervalos=sortrows(Intervalos,3)
cat('Limites Chen_Shao',"\n")
cat('Limite Inferior de Credibilidade - 95% =',Intervalos[1,1],"\n")
cat('Limite Superior de Credibilidade - 95% =',Intervalos[1,2],"\n")

cat('Estimativa de rho a posteriori=',mean(Posteriori),"\n")

```