

AMOSTRAGEM E REAMOSTRAGEM POR IMPORTÂNCIA PARA A ESTIMAÇÃO DE MEDIDAS DE DESEMPENHO EM FILAS $M/M/1$

Victor B. Quinino

Departamento de Engenharia de Produção, Universidade Federal de Minas Gerais
31270-901 – Belo Horizonte – MG, Brasil
victorquinino@ufmg.br

F. R. B. Cruz

Departamento de Estatística, Universidade Federal de Minas Gerais
31270-901 – Belo Horizonte – MG, Brasil
fcruz@est.ufmg.br

RESUMO

Neste artigo utilizamos a amostragem e reamostragem por importância (SIR) para obtenção de uma amostra da distribuição a posteriori para a inferência da intensidade de tráfego em filas $M/M/1$. Esse parâmetro é importante porque, além de representar a fração de tempo em que o atendente fica ocupado, permite a obtenção de importantes medidas de desempenho em filas, tais como o tamanho médio da fila ou o número esperado de usuários no sistema. O desempenho do estimador proposto é analisado via simulações Monte Carlo e comparado com o estimador de máxima verossimilhança clássico. Observamos que os estimadores bayesianos são numericamente fáceis de ser obtidos e podem ter propriedades frequentistas superiores ao estimador de máxima verossimilhança, principalmente em filas sobrecarregadas, ou seja, filas submetidas a intensidades de tráfego elevadas. Todo o processo é exemplificado um exemplo de filas em um posto de saúde.

PALAVRAS CHAVE. Filas markovianas, Estimação bayesiana, SIR.

Tópicos: EST&MP - Estatística e Modelos Probabilísticos; SIM - Simulação

ABSTRACT

In this paper we apply sampling/importance resampling (SIR) to obtain a sample of the posteriori distribution in order to infer the traffic intensity on $M/M/1$ queues. This parameter is important because, besides representing the fraction of time in which a server is occupied, it allows an analyst to obtain important performance measures in queues, such as the average queue length or the expected number of users in the system. The performance of the proposed estimator is analysed by using Monte Carlo simulations and compared to the classical maximum likelihood estimator. The Bayesian estimators were observed to be numerically simple to obtain and can possess superior frequentist properties in relation to the maximum likelihood method, particularly in overloaded queues, that is, queues under high traffic intensities. The entire process is exemplified in a situation in a health center.

KEYWORDS. Markovian queues. Bayesian estimation. SIR.

Paper topics: EST&MP - Statistics and Probabilistic Models; SIM - Simulation

1. Introdução

Teoria da comunicação [Lakatos et al., 2013], projeto de computadores [Kleinrock, 1976; Koole e Mandelbaum, 2002], processos de fabricação [Papadopolous et al., 1993; Govil e Fu, 1999], sistema de saúde [Almehdawe et al., 2013; van Brummelen et al., 2015] e sistemas de transporte [Tas et al., 2013] são apenas alguns exemplos de aplicações da teoria das filas. Mesmo que os gerentes de sistemas possam não estar muito interessados em filas estacionárias, modelos básicos em regime permanente, como as filas markovianas de servidor único, que são representadas na notação de Kendall por $M/M/1$, são de interesse, pois podem ser vistos como um primeiro passo para um processo mais elaborado de análise, com modelos de filas mais sofisticados. Uma vez definido que o modelo $M/M/1$ é apropriado, a tarefa seguinte é a estimação estatística de seus parâmetros. De fato, uma vez que a observação simples e direta de suas características — tais como o tamanho da fila ou o número de usuário no sistema, em momentos aleatórios — está sujeita a variações estocásticas, algum método estatístico precisa ser utilizado na estimação dos parâmetros associados ao modelo, conforme será descrito neste artigo.

Apresenta-se aqui a utilização do método de amostragem e reamostragem por importância (ou SIR, do inglês *sampling/importance resampling*), para a obtenção da distribuição a posteriori, o que permite a estimação da intensidade de tráfego em filas $M/M/1$, ou seja, a fração de tempo em que o atendente fica ocupado. A intensidade de tráfego, ρ , é definida como a razão entre a taxa de chegada e a taxa de serviço. Também estimaremos o valor de ρ pelo método da máxima verossimilhança e compararemos seu resultado com as estimativas bayesiana em uma perspectiva frequentista. Com o valor de ρ definido, podemos calcular o tamanho médio da fila e o número esperado de usuários no sistema. Esse método de estimação bayesiana foi usado com sucesso em recente artigo sobre inferência em filas markovianas [Quinino et al., 2022].

O artigo está organizado da seguinte forma. A fundamentação matemática é apresentada na Seção 2 incluindo a discussão sobre os estimadores bayesianos e de máxima verossimilhança. Os resultados computacionais são apresentados e discutidos na Seção 3. Um exemplo numérico é detalhado na Seção 4. A Seção 5 apresenta as conclusões e considerações finais.

2. Metodologia

Considere uma fila markoviana de servidor único, $M/M/1$, em que o número de chegadas segue uma distribuição de Poisson com taxa $\lambda > 0$ e os tempos entre as partidas seguem uma distribuição exponencial com taxa $\mu > 0$. A razão $\rho = \lambda/\mu$ é a intensidade de tráfego.

Assumindo que a fila é estável ($\rho < 1$) e que o estado de equilíbrio foi alcançado, a quantidade de pessoas no sistema será observada em instantes aleatórios e utilizada para a estimação de ρ . Suponha que $X_1, X_2, \dots, X_i, \dots$, sejam variáveis aleatórias independentes e identicamente distribuídas, em que X_i denota o número de usuários (m) no sistema no i -ésimo momento aleatório. Assim, pode ser mostrado que a variável aleatória X_i segue a distribuição geométrica $X_i \sim \text{Geom}(1 - \rho)$, isto é, com a seguinte função de probabilidade [Medhi, 2003]:

$$P(X_i = m) = \rho^m(1 - \rho), \quad m = 0, 1, 2, \dots, \quad (1)$$

em que $\rho = \lambda/\mu$ é a intensidade de tráfego.

2.1. Estimativas pela Máxima Verossimilhança

Suponha agora que $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ seja uma amostra aleatória (isto é, uma amostra independente e identicamente distribuída), de tamanho n , da distribuição de probabilidade dada pela Eq. (1). Note-se que X_i deve ser observada em tempos aleatórios e suficientemente espaçados para evitar correlação entre as observações e, embora a determinação teórica do espaçamento correto não seja trivial, do ponto de vista prático, se cada observação X_i tem clientes que são todos diferentes

das observações anteriores X_{i-1} , então o espaçamento deve ser adequado. Assim, temos que a função de verossimilhança, baseada na amostra \mathbf{x} , é dada por:

$$\begin{aligned} L(\rho, \mathbf{x}) &= \rho^{x_1}(1 - \rho) \times \rho^{x_2}(1 - \rho) \times \cdots \times \rho^{x_n}(1 - \rho) \\ &= \rho^y(1 - \rho)^n, \end{aligned} \quad (2)$$

em que $y = \sum_{i=1}^n x_i$, a partir da qual o seguinte estimador de máxima verossimilhança (EMV) para ρ pode ser deduzido:

$$\begin{aligned} \hat{\rho}_{\text{EMV}} &= \arg \max_{\rho} L(\rho, \mathbf{x}) \\ &= \frac{y}{y + n}. \end{aligned} \quad (3)$$

Uma observação importante é que, sob diferentes formas de coletar a amostra, diferentes funções de probabilidade surgem, bem como diferentes funções de verossimilhança e estimadores de máxima verossimilhança. De fato, diferentes esquemas amostrais e os respectivos estimadores podem ser encontrados na literatura, como, por exemplo, em Almeida e Cruz [2018].

Por meio da intensidade de tráfego, importantes medidas de desempenho podem ser encontradas, como o tamanho médio da fila e o número esperado de pessoas no sistema, de acordo com as seguintes expressões, respectivamente [Gross et al., 2009]:

$$L_q = \frac{\rho^2}{1 - \rho}, \quad (4)$$

$$L = L_q + \rho = \frac{\rho}{1 - \rho}. \quad (5)$$

Pelo princípio da invariância dos estimadores de máxima verossimilhança [Tan e Drossos, 1975], os respectivos estimadores de máxima verossimilhança para L_q e L podem ser deduzidos:

$$\hat{L}_{q\text{EMV}} = \frac{\hat{\rho}_{\text{EMV}}^2}{1 - \hat{\rho}_{\text{EMV}}}, \quad (6)$$

$$\hat{L}_{\text{EMV}} = \hat{L}_{q\text{EMV}} + \hat{\rho}_{\text{EMV}} = \frac{\hat{\rho}_{\text{EMV}}}{1 - \hat{\rho}_{\text{EMV}}}. \quad (7)$$

2.2. Amostragem e Reamostragem por Importância

A abordagem bayesiana é bem conhecida em inferência estatística e também pode ser usada em inferência em filas. A inferência estatística do ponto de vista bayesiano trata o parâmetro de interesse ρ como uma variável aleatória e começa com a definição de uma função de densidade de probabilidade para modelar ρ no seu espaço paramétrico Ω , chamada de distribuição a priori $\pi(\rho|\alpha)$. Essa priori pode modelar convenientemente o parâmetro ρ , por meio de uma escolha apropriada de α que é o vetor de hiperparâmetros.

Após a coleta de uma amostra aleatória $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$, de tamanho n , a função de verossimilhança $L(\rho, \mathbf{x})$, dada pela Eq. (2), pode ser definida. Com uso do Teorema de Bayes obtemos a distribuição posteriori de ρ , $g(\rho|\mathbf{x}, \alpha)$ dada pela Eq. (8),

$$g(\rho|\mathbf{x}, \alpha) = \frac{L(\rho, \mathbf{x})\pi(\rho|\alpha)}{p(\mathbf{x}, \alpha)}, \quad \rho \in \Omega, \quad (8)$$

em que o denominador:

$$p(\mathbf{x}, \alpha) = \int_{\rho \in \Omega} L(\rho, \mathbf{x})\pi(\rho|\alpha)d\rho$$

é o inverso de uma constante de normalização, que assegura que $\int_{\Omega} g(\rho|\mathbf{x}, \alpha)d\rho = 1$. Assim, por meio da distribuição a posteriori dada pela Eq. (8), todas as inferências relacionadas ao parâmetro

de interesse ρ podem ser executadas. Neste trabalho usaremos a distribuição uniforme definida entre 0 e 1 quando não tivermos informação útil do especialista e a distribuição triangular quando houver uma opinião informativa do especialista uma vez que sua elicitación é simples. Para evitar o uso de integração numérica e facilitar o entendimento dos responsáveis pela gestão do sistema de saúde no processo de inferência bayesiana usaremos o método SIR que é fácil de ser implementado e entendido.

A amostragem e reamostragem por importância (ou SIR, do inglês, *sampling/importance resampling*), discutida em detalhes por Rubin, é aqui proposta para realização da inferência bayesiana [Rubin, 1988]. O método SIR tem como vantagem o fato de não demandar a resolução da integral da função $p(\mathbf{x}, \alpha)$, que pode não ser de solução trivial quando a distribuição a priori não é a conjugada natural para a função de verossimilhança adotada. Além disso, as medidas sintetizadoras da distribuição a posteriori podem ser obtidas similarmente a um estudo descritivo de um banco de dados. O SIR é descrito a seguir.

Uma amostra aleatória (independente e identicamente distribuída) de tamanho k é selecionada da função densidade de probabilidade a priori $\pi(\rho|\alpha)$, em que $k \geq 5.000$. Os valores amostrais são designados como $\rho_{\text{prior},i}$, $i = 1, \dots, k$. Uma função peso W_i , $i = 1, \dots, k$, é calculada para cada ponto $\rho_{\text{prior},i}$, em que W_i é proporcional à função de verossimilhança, Eq. (2). Por último, uma reamostra aleatória, também de tamanho k , é selecionada dos valores a priori de $\rho_{\text{prior},i}$, $i = 1, \dots, k$, com probabilidade proporcional a W_i . Essa nova amostra, denominada $\rho_{\text{post},i}$, $i = 1, \dots, k$, pode ser considerada como proveniente da distribuição a posteriori de ρ e pode ser utilizada para estimar ρ (média a posteriori), pela expressão:

$$\hat{\rho}_{\text{SIR}} = \frac{\sum_{i=1}^k \rho_{\text{post},i}}{k}. \quad (9)$$

Entretanto, Ross descreve uma forma alternativa e potencialmente mais eficiente para a implementação do SIR, em que não é realizada a reamostragem da distribuição a priori, para obtenção de uma amostra da posteriori e consequente estimação do parâmetro ρ [Ross, 1996]. Essencialmente o método SIR baseia-se na estimação da quantidade $\mathbb{E}(\rho|\mathbf{x})$ pela Eq. (9). Ross argumenta que, pelo Teorema de Rao-Blackwell, sobre esperanças condicionais, a expressão

$$\frac{\sum_{i=1}^k W_i \rho_{\text{prior},i}}{\sum_{i=1}^k W_i} \quad (10)$$

possui a mesma média que a Eq. (9) e uma variância inferior, e, portanto, possui um erro quadrático médio inferior a essa forma de estimar $\mathbb{E}(\rho|\mathbf{x})$. Assim, neste trabalho, a estimativa de $\mathbb{E}(\rho|\mathbf{x})$ é feita pela Eq. (10), em que W_i é proporcional à verossimilhança, Eq. (2), e os $\rho_{\text{prior},i}$ são amostrados da distribuição a priori.

Para as medidas de desempenho L_q e L , obtêm-se inicialmente as seguintes quantidades,

$$L_{q\text{prior},i} = \frac{\rho_{\text{prior},i}^2}{1 - \rho_{\text{prior},i}},$$

$$L_{\text{prior},i} = L_{q\text{prior},i} + \rho_{\text{prior},i} = \frac{\rho_{\text{prior},i}}{1 - \rho_{\text{prior},i}},$$

para $i = 1, \dots, k$, após o que obtêm-se as estimativas bayesianas para L_q e L , através das seguintes expressões, respectivamente:

$$\frac{\sum_{i=1}^k W_i L_{q\text{prior},i}}{\sum_{i=1}^k W_i}, \quad (11)$$

$$\frac{\sum_{i=1}^k W_i L_{\text{prior},i}}{\sum_{i=1}^k W_i}. \quad (12)$$

3. Resultados Computacionais

O comportamento dos estimadores apresentados é avaliado computacionalmente. Os experimentos computacionais a seguir objetivam comparar, por meio de simulações Monte Carlo, a efetividade do método SIR em relação ao EMV, quando utilizados para estimar ρ , L_q e L .

Para tal, foram utilizados tamanhos amostrais n iguais a 10, 20, 50, 80, 100 e 200, com 1.000 replicações Monte Carlo. As amostras aleatórias das quantidades de clientes no sistema foram geradas por uma distribuição geométrica com parâmetro $(1 - \rho)$, Eq. (1), sendo que a intensidade de tráfego ρ foi fixada nos valores 0,1, 0,2, 0,4, 0,5, 0,6, 0,8 e 0,9. No método SIR, o tamanho das amostras geradas da distribuição a priori de ρ foi fixado em $k = 5.000$.

Para uma comparação em situações diversas, foram considerados três tipos diferentes de distribuição a priori para o método SIR: (i) uma distribuição uniforme no espaço paramétrico de ρ , *i.e.*, no intervalo $(0, 1)$, representando uma situação em que o analista de dados não tem informação a priori sobre o valor do parâmetro; (ii) uma distribuição triangular com a moda, valor mais provável, localizada exatamente no valor fixado do parâmetro ρ (que é desconhecido na prática), e valores mínimo e máximo respectivamente iguais a 1 e 0, representando uma situação em que o analista tem a melhor informação possível a priori sobre a localização mais provável do parâmetro; e (iii) uma distribuição triangular com moda localizada em um ponto distante do valor fixado do parâmetro ρ (na prática, desconhecido), representando uma situação em que o analista tem uma opinião a priori equivocada sobre a localização do parâmetro ρ . Para os valores fixados de $\rho \leq 0.5$ usamos uma distribuição a priori triangular com mínimo=moda=0 e máximo igual a 1; se $0.5 < \rho \leq 1$ usamos uma distribuição a priori triangular com mínimo=0 e moda=máximo=1.

A Tabela 1 apresenta os resultados computacionais obtidos para os estimadores da intensidade de tráfego ρ , utilizando-se o EMV e o método SIR, para as diversas distribuições a priori, através de valores das médias e variâncias dos estimadores calculados sobre as 1.000 replicações Monte Carlo. Nota-se que, em geral, as estimativas aproximam-se do valor fixado (segunda coluna da tabela) e as variâncias decrescem, com o aumento do tamanho da amostra, conforme é desejável.

A Figura 1 sintetiza a tabela anterior, em termos do desempenho médio dos estimadores, em função da posição do parâmetro ρ no espaço paramétrico e em função do tamanho da amostra n . Nota-se que, embora o EMV apresente em geral o menor viés, Figura 1-(a), as suas variâncias são as maiores dentre todos, Figura 1-(b). Além disso, os estimadores bayesianos têm um desempenho equivalente, em termos do viés, para valores elevados da intensidade de tráfego ($\rho \geq 0,5$), aliados a uma variância menor, tanto para a priori uniforme entre 0 e 1, quanto para o melhor caso da priori triangular, indicando que o método bayesiano é preferível para essa situação. Sob uma priori ruim (pior caso da triangular), a variância do estimador é equivalente ao EMV, mas apresenta o pior viés dentre todos os estimadores, indicando que na elicitação da distribuição a priori o analista precisa apresentar uma boa informação, ou então optar por uma priori vaga, ou seja, optar por uma distribuição a priori uniforme entre 0 e 1. Em relação ao tamanho da amostra, o viés, Figura 1-(c), e as variâncias médias, Figura 1-(d), tendem para zero, com o seu aumento, o que sugere que os estimadores devem ser consistentes.

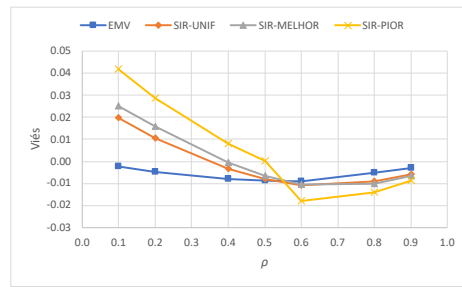
A Tabela 2 apresenta os resultados computacionais obtidos para os estimadores do tamanho médio da fila L_q , utilizando-se o EMV e o método SIR, que incluem a média e a variância, para as diversas distribuições a priori. Pelos resultados apresentados, pode-se observar que o EMV tende a apresentar o menor viés em relação ao método SIR, nos casos testados. Também é possível observar que os dois métodos apresentam baixo viés, nos valores simulados considerando o ρ fixado. Entretanto, a variância cresce bastante conforme o valor de L_q aumenta, embora diminua com o crescimento do tamanho da amostra n .

Tabela 1: Desempenho dos estimadores para ρ (média e variância)

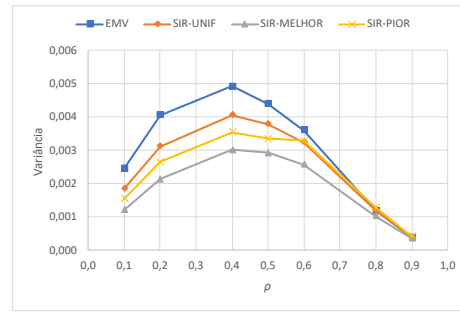
n	ρ fixado	Média				Variância			
		EMV	SIR-UNIF	SIR-MELHOR	SIR-PIOR	EMV	SIR-UNIF	SIR-MELHOR	SIR-PIOR
10	0,1	0,09609	0,15896	0,16885	0,21895	0,00718	0,00456	0,00262	0,00341
	0,2	0,18797	0,23340	0,24569	0,28374	0,01226	0,00819	0,00476	0,00628
	0,4	0,37948	0,39516	0,40319	0,42764	0,01533	0,01144	0,00750	0,00923
	0,5	0,47693	0,48133	0,48638	0,50564	0,01365	0,01090	0,00757	0,00901
	0,6	0,57527	0,57132	0,57362	0,55072	0,01145	0,00978	0,00715	0,00991
	0,8	0,78421	0,77322	0,77123	0,75791	0,00384	0,00385	0,00315	0,00420
	0,9	0,88981	0,88164	0,87976	0,87394	0,00128	0,00140	0,00121	0,00155
	0,1	0,09504	0,12900	0,13845	0,16413	0,00391	0,00307	0,00195	0,00260
	0,2	0,19011	0,21377	0,22270	0,24253	0,00619	0,00502	0,00343	0,00434
20	0,4	0,38576	0,39300	0,39743	0,41036	0,00748	0,00647	0,00491	0,00579
	0,5	0,48449	0,48590	0,48815	0,49841	0,00696	0,00623	0,00493	0,00563
	0,6	0,58434	0,58146	0,58200	0,57040	0,00564	0,00525	0,00429	0,00529
	0,8	0,79053	0,78474	0,78326	0,77693	0,00176	0,00178	0,00155	0,00188
	0,9	0,89467	0,89060	0,88937	0,88742	0,00057	0,00060	0,00054	0,00058
	0,1	0,09884	0,11286	0,11798	0,12846	0,00152	0,00138	0,00104	0,00131
	0,2	0,19773	0,20732	0,21130	0,21935	0,00238	0,00218	0,00174	0,00205
	0,4	0,39476	0,39736	0,39892	0,40447	0,00273	0,00258	0,00220	0,00247
	0,5	0,49441	0,49457	0,49511	0,49974	0,00238	0,00230	0,00200	0,00219
50	0,6	0,59485	0,59339	0,59308	0,58880	0,00182	0,00178	0,00158	0,00179
	0,8	0,79691	0,79452	0,79368	0,79202	0,00069	0,00070	0,00064	0,00069
	0,9	0,89869	0,89706	0,89644	0,89457	0,00019	0,00019	0,00018	0,00020
	0,1	0,09960	0,10841	0,11188	0,11701	0,00096	0,00091	0,00072	0,00089
	0,2	0,19930	0,20529	0,20783	0,21307	0,00152	0,00144	0,00122	0,00138
	0,4	0,39852	0,40002	0,40087	0,40461	0,00172	0,00167	0,00148	0,00162
	0,5	0,49826	0,49832	0,49847	0,50138	0,00147	0,00143	0,00129	0,00139
	0,6	0,59683	0,59591	0,59567	0,59294	0,00124	0,00121	0,00111	0,00122
	0,8	0,79912	0,79766	0,79702	0,79489	0,00040	0,00040	0,00038	0,00046
100	0,9	0,89974	0,89872	0,89828	0,89653	0,00012	0,00012	0,00011	0,00012
	0,1	0,09843	0,10558	0,10862	0,11464	0,00083	0,00079	0,00065	0,00071
	0,2	0,19735	0,20224	0,20448	0,20839	0,00134	0,00129	0,00111	0,00125
	0,4	0,39533	0,39660	0,39742	0,40029	0,00159	0,00155	0,00140	0,00151
	0,5	0,49479	0,49488	0,49517	0,49740	0,00135	0,00132	0,00121	0,00129
	0,6	0,59590	0,59518	0,59503	0,59278	0,00108	0,00106	0,00098	0,00107
	0,8	0,79842	0,79722	0,79676	0,79554	0,00034	0,00034	0,00032	0,00035
	0,9	0,89945	0,89865	0,89831	0,89697	0,00009	0,00009	0,00009	0,00010
	0,1	0,09946	0,10303	0,10462	0,10781	0,00039	0,00038	0,00033	0,00038
200	0,2	0,19892	0,20136	0,20253	0,20458	0,00065	0,00063	0,00058	0,00063
	0,4	0,39822	0,39884	0,39919	0,40071	0,00068	0,00068	0,00063	0,00067
	0,5	0,49833	0,49835	0,49840	0,49957	0,00060	0,00059	0,00055	0,00059
	0,6	0,59774	0,59735	0,59725	0,59619	0,00043	0,00043	0,00041	0,00044
	0,8	0,79950	0,79890	0,79863	0,79823	0,00015	0,00016	0,00015	0,00016
	0,9	0,89924	0,89884	0,89928	0,89833	0,00004	0,00004	0,00004	0,00005

Tabela 2: Desempenho dos estimadores para L_q (média e variância)

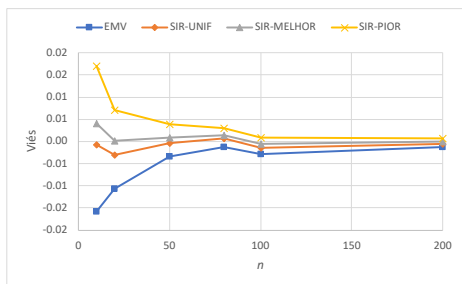
n	L_q fixado	Média				Variância			
		EMV	SIR-UNIF	SIR-MELHOR	SIR-PIOR	EMV	SIR-UNIF	SIR-MELHOR	SIR-PIOR
10	0,5	0,53307	0,62875	0,60192	0,70477	0,12280	0,13048	0,08965	0,13651
	1,0	1,02863	1,13373	1,08082	0,99753	0,32847	0,33667	0,23629	0,26817
	2,0	2,01692	2,12770	2,02455	1,88703	0,88727	0,88802	0,64275	0,72046
	4,0	3,95242	4,06398	3,86457	3,67800	2,53482	2,52763	1,87392	2,19788
	6,0	5,96312	6,07339	5,77127	5,48624	5,38940	5,35054	3,99050	4,09347
20	8,0	8,09942	8,20863	7,78042	7,30274	9,32346	9,39667	6,99312	7,49286
	0,5	0,50606	0,55473	0,54635	0,59190	0,05791	0,05995	0,04731	0,06151
	1,0	0,99127	1,04466	1,02437	0,97758	0,15529	0,15740	0,12608	0,13934
	2,0	2,04141	2,09745	2,04471	1,97390	0,46932	0,47117	0,38368	0,42262
	4,0	3,99739	4,05300	3,95177	3,82075	1,28357	1,28470	1,05622	1,17845
50	6,0	5,99851	6,05414	5,89954	5,83264	2,78309	2,78352	2,31477	2,38073
	8,0	8,06112	8,11241	7,90096	7,60353	4,13416	4,12131	3,42153	3,76557
	0,5	0,50197	0,52171	0,51967	0,53710	0,02101	0,02137	0,01861	0,02153
	1,0	1,00261	1,02438	1,01688	0,99696	0,05865	0,05889	0,05188	0,05619
	2,0	1,98193	2,00443	1,98650	1,95481	0,16800	0,16834	0,14966	0,16134
80	4,0	3,98997	4,01267	3,97338	3,91837	0,51487	0,51547	0,46116	0,48184
	6,0	5,98515	6,01010	5,94850	5,88296	1,06665	1,07242	0,96007	0,95595
	8,0	7,99770	8,01771	7,93799	7,80915	1,66336	1,66352	1,49714	1,74821
	0,5	0,50643	0,51892	0,51759	0,52815	0,01335	0,01349	0,01213	0,01355
	1,0	1,00314	1,01671	1,01242	0,99924	0,03890	0,03891	0,03549	0,03773
100	2,0	1,99272	2,00667	1,99592	1,97531	0,09939	0,09936	0,09126	0,09685
	4,0	4,02572	4,03888	4,01309	3,95044	0,32869	0,32921	0,30149	0,33004
	6,0	6,00373	6,01664	5,98021	5,91095	0,58325	0,58249	0,53780	0,56768
	8,0	8,08843	8,10116	8,04703	7,91278	1,08208	1,09674	1,00232	1,05624
	0,5	0,49506	0,50504	0,50460	0,51243	0,01187	0,01197	0,01092	0,01198
200	1,0	0,99386	1,00492	1,00169	0,99072	0,03241	0,03251	0,02994	0,03168
	2,0	2,00466	2,01637	2,00671	1,99069	0,09307	0,09341	0,08642	0,09143
	4,0	4,01924	4,03107	4,01017	3,97144	0,26027	0,26126	0,24205	0,24775
	6,0	5,97650	5,98601	5,95813	5,91038	0,47353	0,47565	0,44180	0,54774
	8,0	8,01517	8,02559	7,98616	7,92914	0,81120	0,81233	0,75601	0,80450
	0,5	0,49973	0,50473	0,50445	0,50842	0,00535	0,00536	0,00504	0,00541
	1,0	0,99407	0,99944	0,99801	0,99263	0,01305	0,01308	0,01235	0,01294
	2,0	2,00923	2,01463	2,01031	2,00174	0,04341	0,04363	0,04133	0,04318
	4,0	4,00766	4,01322	4,00337	3,99468	0,12100	0,12154	0,11527	0,12784
	6,0	6,01230	6,01742	6,00338	5,97813	0,26439	0,26555	0,25246	0,26778
	8,0	8,01011	8,01534	8,02133	7,95951	0,41233	0,41264	0,40392	0,42700



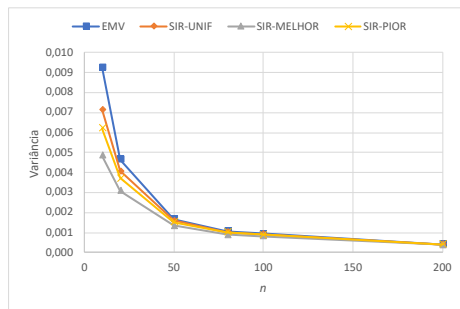
(a) Viés em função de ρ



(b) Variância média em função de ρ

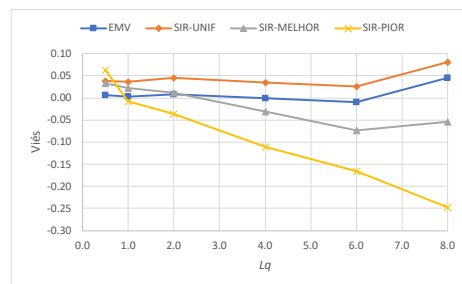


(c) Viés em função de n

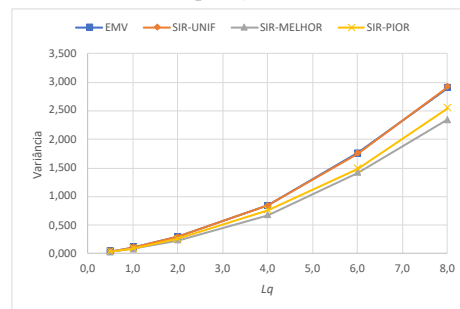


(d) Variância média em função de n

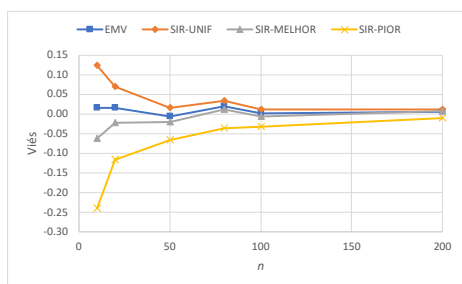
Figura 1: Desempenho médio dos estimadores para ρ



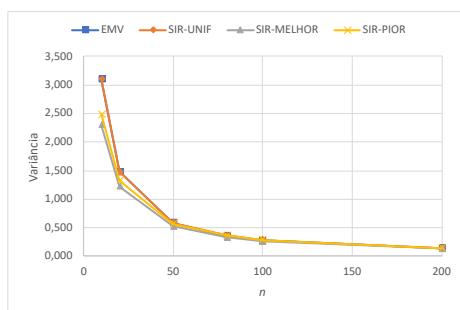
(a) Viés em função de L_q



(b) Variância média em função de L_q



(c) Viés em função de n



(d) Variância média em função de n

Figura 2: Desempenho médio dos estimadores para L_q

A Figura 2 sintetiza a Tabela 2, em termos do viés e das variâncias médias das estimativas, em função do valor do parâmetro estimado L_q e em função do tamanho da amostra n . Em geral, as figuras explicitam melhor os padrões observados na Tabela 2. Assim, evidencia-se o melhor desempenho do estimador de máxima verossimilhança em relação ao viés, seguido de perto pelo

método SIR sob distribuições a priori uniforme e melhor caso da triangular, ao mesmo tempo em que o desempenho médio dos dois métodos se mostrou próximo em relação à variabilidade. Também aqui observa-se que o viés e as variâncias das estimativas tendem para zero, com o aumento do tamanho da amostra n , o que é um comportamento bastante desejável para um estimador.

Resultados similares (não apresentados, por falta de espaço) foram obtidos para o número esperado de usuários no sistema L .

4. Exemplo Numérico

Nesta seção a metodologia descrita é exemplificada em uma situação prática, com o intuito de facilitar a compreensão do leitor. Um ponto já comum no mundo são as filas para vacinação contra a pandemia do COVID-19, que atualmente ainda representa uma grande preocupação de saúde pública global [Johns Hopkins University & Medicine, 2022]. Algo que as secretarias de saúde talvez queiram controlar é a quantidade de pessoas nos postos de vacinação, visto que um número alto de pessoas em mesmo local configura-se como aglomeração, o que pode aumentar os riscos de contaminação.

Buscamos evitar a alocação de um profissional para fazer o controle da quantidade de pessoas presentes em um local de vacinação, visto que isso onera o sistema de saúde. Para tal, primeiramente devemos obter um parecer de um profissional de saúde sobre quantas pessoas podem estar em cada ambiente de forma segura.

Nesse exemplo tratamos de um posto de saúde específico. Por meio de diversos estudos anteriores, podemos afirmar que pessoas chegam ao posto durante um dado período de tempo seguindo aproximadamente uma distribuição de Poisson e os tempos de serviço são aproximadamente distribuídos de forma exponencial. Segundo um experiente profissional de saúde consultado, esse posto de saúde pode conter em média 6 pessoas sem aglomerações.

Para concluir se será ou não necessário controlar a entrada de pessoas nesse sistema (que não apresenta controle de entrada), seguiremos os seguintes passos.

Passo 1: Utilizando a Eq. (5), podemos encontrar a intensidade de tráfego ρ que equivale à quantidade média de 6 pessoas:

$$L = \frac{\rho}{1 - \rho} = 6 \therefore \rho = 0,857143; \quad (13)$$

Passo 2: Definimos uma hipótese nula H_0 , em que $\rho \leq 0,857143$, e uma hipótese alternativa H_1 , em que $\rho > 0,857143$:

$$\begin{cases} H_0 : \rho \leq 0,857143, \\ H_1 : \rho > 0,857143; \end{cases} \quad (14)$$

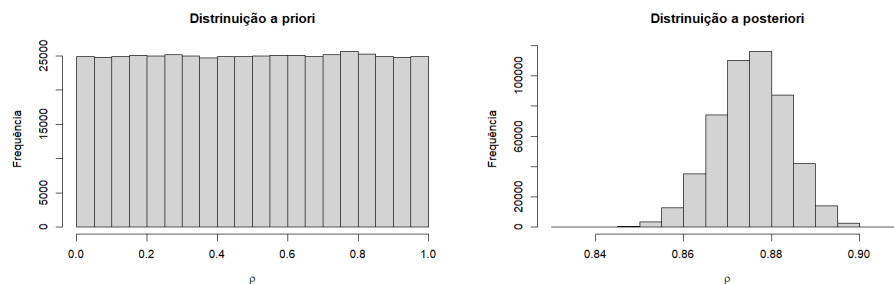
Passo 3: Coletamos 200 amostras da quantidade de pessoas no sistema em momentos suficientemente espaçados, para garantir independência dos dados, apresentados na Tabela 3 (para um melhor entendimento do leitor, na Tabela 3, o valor 0 (zero) foi observado com uma frequência de 32 vezes e o valor 1 foi observado com uma frequência de 25 vezes; os demais valores seguem o mesmo entendimento);

Passo 4: Uma vez que temos os dados coletados apresentados na Tabela 3, obtemos a distribuição a posteriori via método SIR, adotando uma priori uniforme entre 0 e 1, visto que não temos informação sobre o parâmetro ρ relativo vacinação de uma doença tal como a COVID-19; em

Tabela 3: Valores observados (O) e frequência (F) para uma amostra de tamanho $n = 200$

O	F	O	F	O	F	O	F
0	32	8	6	16	6	27	1
1	25	9	7	17	3	28	1
2	20	10	3	18	4	29	1
3	16	11	8	19	2	31	1
4	12	12	4	21	1	33	1
5	9	13	2	22	2	37	1
6	12	14	5	23	1	38	1
7	5	15	6	24	1	43	1

seguida, encontramos a distribuição a posteriori, que apresentou uma média $\hat{\rho} = 0,8753931$; as duas distribuições são descritas, respectivamente, em histogramas, na Figura 3;



(a) distribuição a priori

(b) distribuição a posteriori

Figura 3: Histogramas para as distribuições a priori e a posteriori

Passo 5: Com as amostras das distribuições a priori e a posteriori, podemos realizar o teste de hipótese de H_0 versus H_1 , definido no passo 3; segundo Choudhury e Borthakur [2008], podemos encontrar o fator de Bayes (FB),

$$\frac{P(H_1|\text{data})}{P(H_0|\text{data})} = \text{FB} \frac{P(H_1)}{P(H_2)} \Rightarrow \text{FB} = 379,6079, \quad (15)$$

em que:

- $P(H_1|\text{data})$ representa a porcentagem de valores na distribuição a posteriori $[g(\rho|\mathbf{x}, \alpha)]$ que se encaixam na hipótese H_1 , isto é, que são maiores que 0,857143;
- $P(H_0|\text{data})$ representa a porcentagem de valores na distribuição a posteriori $[g(\rho|\mathbf{x}, \alpha)]$ que se encaixam na hipótese H_0 , isto é, que são menores ou iguais a 0,857143;
- $P(H_1)$ representa a porcentagem de valores na distribuição a priori $[\pi(\rho|\alpha)]$ que se encaixam na hipótese H_1 , isto é, que são maiores que 0,857143;
- $P(H_0)$ representa a porcentagem de valores na distribuição a priori $[\pi(\rho|\alpha)]$ que se encaixam na hipótese H_0 , isto é, que são menores ou iguais a 0,857143;

Passo 6: Finalmente, utilizamos a regra definida por Kass e Raftery [1995], para a escolha entre H_0 e H_1 , de acordo com a Tabela 4, através do cálculo de $2 \ln(\text{FB})$, que resulta em 11,87828 e leva à conclusão de que há evidências decisivas contra H_0 , ou seja, há evidências decisivas contra a hipótese que não irão ocorrer aglomerações no posto, fazendo-se necessário, portanto, controlar a entrada de pessoas nesse posto.

Tabela 4: Regra para escolha entre H_0 e H_1

$2 \ln(B)$	Evidência contra H_0
0 a 2	Mínima
2 a 6	Positiva
6 a 10	Forte
>10	Decisiva

O fator de bayes (FB) discutido nesta seção pode ser difícil de interpretar, pois possui um aspecto técnico. Assim, o cálculo do valor esperado da função de probabilidade expressa na Eq. (1) em relação a distribuição a posteriori ρ_{post} , denominado probabilidade preditiva, pode ser útil e fácil de interpretar pois é expressa em probabilidade. Neste sentido, a probabilidade preditiva de termos uma quantidade de pessoas no sistema igual ou superior a 6 é 0,4506. Este resultado mostra que a probabilidade de acontecer aglomeração no local da vacinação é muito alta e ratifica a decisão do teste de hipóteses realizado em relação a necessidade de termos controle de entrada ou aumentar o número de atendentes.

5. Conclusões e Observações Finais

A partir dos resultados obtidos por meio do uso de simulações Monte Carlo, foi possível concluir que a amostragem e reamostragem por importância (ou SIR, *sampling/importance resampling*) mostra-se, de forma geral, superior ao estimador de máxima verossimilhança (EMV), para intensidades de tráfego elevadas ($\rho \geq 0,5$), com baixo viés aliado a baixas variâncias das estimativas. Além disso, conforme o tamanho amostral n aumenta, as estimativas pelos dois métodos aproximam-se umas das outras e convergem para o valor do parâmetro fixado na simulação.

Podemos observar também que caso o conhecimento a priori seja acurado, o SIR mostra-se superior ao EMV, o que é uma situação razoável na prática considerando que alguns especialistas da área de saúde possuem anos de experiência administrando postos de saúde. De fato, em vários sistemas, os trabalhadores ou gerentes da área têm uma ideia razoável de ρ , por meio do conhecimento da proporção do tempo que em média o atendente fica ocupado.

Futuros trabalhos na área podem ser feitos utilizando-se outros métodos de coleta amostral, por exemplo, contando o número de chegadas durante o serviço do i -ésimo usuário. Além disso, novos trabalhos podem ser desenvolvidos para estimativas em outros tipos de filas, como, por exemplo, em filas markovianas com servidores múltiplos, ou seja, em filas $M/M/s$, na notação de Kendall.

Agradecimentos

Os autores agradecem ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico, processos 305515/2018-7 e 160974/2020-8) e à FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais, processo CEX-PPM-00564-17) pelo auxílio financeiro.

Referências

- Almehdawe, E., Jewkes, B., e He, Q.-M. (2013). A Markovian queueing model for ambulance offload delays. *European Journal of Operational Research*, 226(3):602–614.
- Almeida, M. A. C. e Cruz, F. R. B. (2018). A note on Bayesian estimation of traffic intensity in single-server Markovian queues. *Communications in Statistics - Simulation & Computation*, 47(9):2577–2586.
- Armero, C. e Bayarri, M. J. (1994). Prior assessments for prediction in queues. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 43(1):139–153.
- Choudhury, A. e Borthakur, A. C. (2008). Bayesian inference and prediction in the single server markovian queue. *Metrika*, 67(3):371–383.

- Govil, M. K. e Fu, M. C. (1999). Queueing theory in manufacturing: A survey. *Journal of Manufacturing Systems*, 18(3):214.
- Gross, D., Shortle, J. F., Thompson, J. M., e Harris, C. M. (2009). *Fundamentals of queueing theory*. Wiley-Interscience, New York, NY, 4th edition.
- Johns Hopkins University & Medicine (2022). Coronavirus resource center. <https://coronavirus.jhu.edu/>. Accessed on 2022-09-10.
- Kass, R. E. e Raftery, A. E. (1995). Bayes factors. *Journal of the american statistical association*, 90(430):773–795.
- Kleinrock, L. (1976). *Queueing systems - Volume I: Theory*, p. 21–53. John Wiley & Sons.
- Koole, G. e Mandelbaum, A. (2002). Queueing models of call centers: An introduction. *Annals of Operations Research*, 113(1-4):41–59.
- Lakatos, L., Szeidl, L., e Telek, M. (2013). *Introduction to queueing systems with telecommunication applications*. Springer Science & Business Media, New York, NY.
- Medhi, J. (2003). *Stochastic Models in Queueing Theory*. Academic Press, 2 ed.
- Papadopolous, H. T., Heavey, C., e Browne, J. (1993). *Queueing Theory in Manufacturing Systems Analysis and Design*. Springer Science & Business Media.
- Quinino, V. B., Cruz, F. R. B., e Quinino, R. C. (2022). Sampling and importance resampling in traffic intensity estimates in markovian single-server queues. *Queueing Models and Service Management*, 5(2):29–44.
- Ross, S. M. (1996). Bayesians should not resample a prior sample to learn about the posterior. *The American Statistician*, 50(2):116–116.
- Rubin, D. B. (1988). Using the SIR algorithm to simulate posterior distributions. *Bayesian Statistics*, 3:395–402.
- Tan, P. e Drossos, C. (1975). Invariance properties of maximum likelihood estimators. *Mathematics Magazine*, 48(1):37–41.
- Tas, D., Dellaert, N., van Woensel, T., e de Kok, T. (2013). Vehicle routing problem with stochastic travel times including soft time windows and service costs. *Computers & Operations Research*, 40(1):214–224.
- Gupta, Arjun K and Nadarajah, Saralees (2004). *Handbook of beta distribution and its applications*. CRC press, New York, NY.
- Fairchild, Keith Wm and Misra, Lalatendu and Shi, Yilun (2016). Using triangular distribution for business and finance simulations in Excel. *Journal of Financial Education*, 42(3-4):313–336.
- van Brummelen, S. P. J., de Kort, W. L., e van Dijk, N. M. (2015). Waiting time computation for blood collection sites. *Operations Research for Health Care*, 7:70–80.