

Estimação da intensidade de tráfego de filas markovianas multi-servidoras via SIR

Sandy P. Alves

Universidade Federal de Minas Gerais; Departamento de Estatística - ICEX, Belo Horizonte - MG
sandy2013@ufmg.br

Frederico R. B. Cruz

Universidade Federal de Minas Gerais; Departamento de Estatística - ICEX, Belo Horizonte - MG
fcruz@est.ufmg.br

Roberto C. Quinino

Universidade Federal de Minas Gerais; Departamento de Estatística - ICEX, Belo Horizonte - MG
roberto@est.ufmg.br

Resumo: Filas multi-servidoras com chegadas Poisson e tempos de serviço exponenciais são o foco deste trabalho, mais especificamente, o problema de estimação bayesiana da sua intensidade de tráfego, definida como a razão entre a taxa de chegada e a taxa de serviço. Tais filas são importantes como modelos aproximados em diversas situações práticas, tais como em redes de computadores e de telecomunicações, sistemas de manufatura, de serviços e de saúde, entre outros problemas similares. É investigado aqui o uso do método de amostragem/reamostragem de importância (ou SIR, do inglês *sampling/importance resampling*) e sua implementação na plataforma R. A determinação da intensidade de tráfego é ponto de partida para obtenção de outras medidas de desempenho importantes, tais como, a probabilidade de o sistema estar vazio, o tamanho médio da fila, o número esperado de usuários no sistema, dentre outras. Resultados numéricos são apresentados, para demonstrar a eficácia e a eficiência do SIR, diante de amostras de tamanho finito.

Palavras-chave: Métodos Estocásticos e Estatísticos. Filas Markovianas. Inferência em filas. SIR.

Introdução

As filas são descritas, basicamente, por um processo de chegada de clientes, ou produtos, em um determinado sistema de atendimento, ou de produção, que provê um ou mais serviços, que será executado por certa quantidade de servidores. O foco deste estudo foram as filas markovianas infinitas com c servidores em paralelo, ou, na notação de Kendall (1953), filas $M/M/c$ (ou $M/M/c/\infty$). A intensidade de tráfego, ρ , definida como a relação entre a taxa de chegada (λ) e a taxa de serviço (μ), representa a proporção média de tempo que cada servidor permanece ocupado e é a medida de desempenho que este artigo tem por foco estimar. A partir desta, outras medidas de desempenho podem ser determinadas, como o tamanho médio da fila ou o número esperado de usuários no sistema.

O objetivo é o estudo empírico de estimadores estatísticos para o parâmetro ρ , em filas $M/M/c$, e seu desempenho frente a amostras de tamanho finito (≤ 250). A ideia é utilizar um método Monte Carlo não iterativo, denominado amostragem/reamostragem de importância (ou SIR, do inglês *sampling/importance resampling*). Ferramenta simples e conhecida desde o final dos anos 1980's (RUBIN, 1988) e início dos anos 1990's (SMITH; GELFAND, 1992), o SIR é uma técnica de estatística bayesiana que tem se mostrado viável em novas aplicações, inclusive em estimação em filas $M/M/c$ (CRUZ; QUININO; HO, 2017), principalmente devido aos avanços computacionais que possibilitam a precisão e a velocidade necessárias aos cálculos envolvidos. Para uma revisão ampla e atualizada sobre inferência em filas em geral, ver Asanjarani, Nazarathy e Pollett (2017).

Neste artigo, uma implementação do SIR em R (R CORE TEAM, 2017) é apresentada e tornada disponível para a comunidade, uma vez que o R é um ambiente de software livre para computação estatística e gráfica que é bastante utilizado pela comunidade de matemática aplicada e computacional e estatística. O restante do artigo está

organizado da seguinte forma. A próxima seção detalha as equações e algoritmos envolvidos. Em seguida, os resultados computacionais são apresentados e discutidos. Finalmente, o artigo é concluído com algumas observações finais e tópicos para futuras pesquisas na área.

Materiais e métodos

Embora seja do ponto de vista matemático um dos modelos de filas mais simples, as filas markovianas multi-servidoras infinitas são aproximações bastante satisfatórias de diversos sistemas reais. Elas são o foco deste estudo. O modelo probabilístico de tais filas é apresentado na Figura 1, em que o número k de clientes que chegam na unidade de tempo segue uma distribuição Poisson, com função de probabilidade

$$p(k) = \frac{e^{-\lambda} \lambda^k}{k!}, k = 0, 1, 2, \dots, \quad (1)$$

e os tempos de serviço t seguem uma distribuição Exponencial, com função densidade de probabilidade

$$g(t) = \mu e^{-\mu t}, t \geq 0. \quad (2)$$

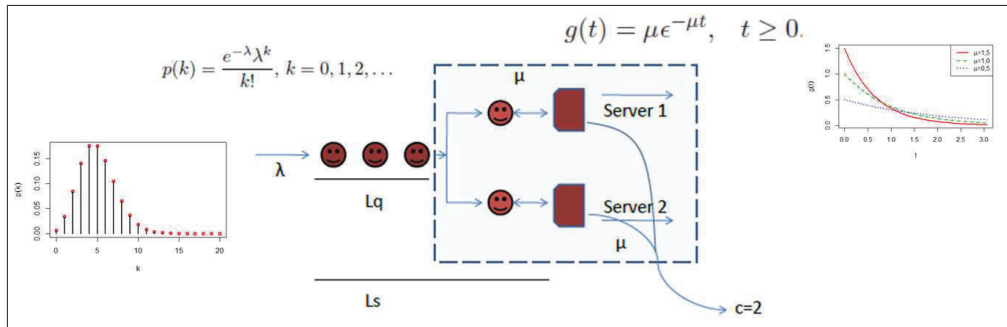


Figura 1: Um exemplo de fila $M/M/c$, com $c = 2$ servidores.

À esquerda da Figura 1, tem-se o gráfico da função de probabilidade de uma distribuição Poisson com uma determinada taxa λ . Este gráfico é apresentado com o intuito de ilustrar uma modelagem do número de chegadas de clientes ou produtos em um sistema de atendimento. A abscissa k refere-se ao número de chegadas em um determinado período de tempo e a ordenada $p(k)$ representa a probabilidade de ocorrer tais chegadas. Ao lado direito da Figura 1, encontra-se o gráfico da função densidade de probabilidade para a distribuição Exponencial, com várias taxas μ . Este gráfico é mostrado com o propósito de ilustrar uma modelagem do tempo de serviço, em que a abscissa t representa o tempo de serviço, em unidades de tempo, e a ordenada a respectiva densidade de probabilidade $g(t)$.

Assumindo que a fila é observada por um longo período de operação, até alcançar um estado estacionário, e que a intensidade de tráfego é tal que $\rho < 1$, para assegurar que o número de usuários em espera não cresça indefinidamente, é possível deduzir a seguinte distribuição estacionária do número de clientes (N) no sistema (GROSS et al., 2009).

$$P(N = n) \equiv p_n = \begin{cases} \frac{(c\rho)^n}{n!} p_0, & 0 \leq n \leq c, \\ \frac{c^c \rho^n}{c!} p_0, & n > c, \end{cases} \quad (3)$$

em que p_n é a probabilidade de se encontrar n clientes no sistema em determinado momento e a probabilidade de o sistema estar vazio, $P(N = 0) \equiv p_0$, é dada a partir da condição de que as probabilidades devem somar 1,

$$p_0 = \left[\sum_{j=0}^{c-1} \frac{(c\rho)^j}{j!} + \frac{c^c \rho^c}{c!} \frac{1}{1-\rho} \right]^{-1}. \quad (4)$$

Supondo agora que $\mathbf{x} = \{x_1, x_2, \dots, x_\ell\}$ constitui uma amostra, de tamanho ℓ , em que os x_i representam o número de clientes observados no sistema (i.e., os clientes em serviço e em espera, na fila), a seguinte função de verossimilhança pode ser escrita (CRUZ; QUININO; HO, 2017),

$$L(\mathbf{x}|\rho) = \prod_{i=1}^{\ell} \left[\frac{(c\rho)^{x_i}}{x_i!} p_0 I_{\{0 \leq x_i \leq c\}} + \frac{c^c \rho^{x_i}}{c!} p_0 I_{\{x_i > c\}} \right], \quad (5)$$

em que $I_{\{\bullet\}}$ é a função indicadora. Note-se que x_i deve ser observado em instantes aleatórios e suficientemente espaçados, para evitar correlação entre as observações. A determinação desse espaçamento, do ponto de vista teórico, não é trivial e está fora do escopo deste trabalho. No entanto, do ponto de vista prático, se em cada observação, x_i , os clientes forem todos diferentes daqueles da observação anterior, x_{i-1} , tem-se então um espaçamento adequado.

Uma vez que $0 < \rho < 1$, qualquer distribuição com suporte $(0, 1)$ é uma candidata natural a distribuição a priori para este parâmetro. Uma distribuição que já vem sendo usada com sucesso em inferência em filas (GUPTA; NADARAJAH, 2004) é

$$\Pi(\rho) \propto \rho^{(a-1)}(1-\rho)^{(b-1)}, a > 0, b > 0, \quad (6)$$

que, após a normalização, é a conhecida distribuição beta, com parâmetros a e b , $\text{Beta}(a, b)$. A beta será utilizada aqui por comodidade, uma vez que oferece a flexibilidade necessária para a modelagem do conhecimento a priori sobre o parâmetro ρ , conforme pode ser visto na Figura 2, além de facilitar a comparação com estudos anteriores, nos quais foi utilizada. Vários métodos para obtenção dos parâmetros a e b em situações reais podem ser encontrados em Gupta e Nadarajah (2004).

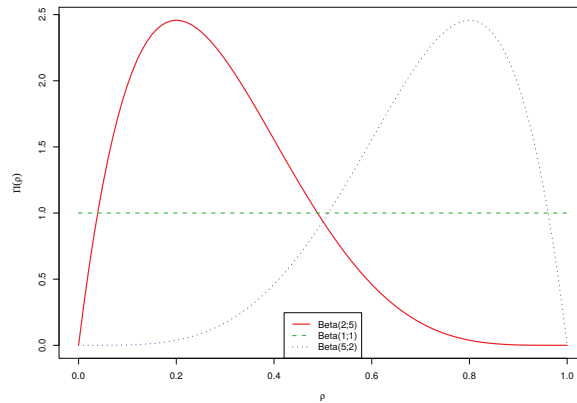


Figura 2: Várias distribuições a priori $\text{Beta}(a, b)$.

A distribuição a posteriori, correspondente à distribuição a priori dada pela Eq. (6), é

$$\Pi(\rho|\mathbf{x}) \propto \rho^{(a-1)}(1-\rho)^{(b-1)} \prod_{i=1}^{\ell} \left[\frac{(c\rho)^{x_i}}{x_i!} p_0 I_{\{0 \leq x_i \leq c\}} + \frac{c^c \rho^{x_i}}{c!} p_0 I_{\{x_i > c\}} \right]. \quad (7)$$

Entretanto, como seria difícil obter exatamente a distribuição de probabilidade a posteriori $\Pi(\rho|\mathbf{x})$, dada pela Eq. (7), será utilizado o SIR. Discutido em detalhe por Rubin (1988) e Smith e Gelfand (1992), este algoritmo é descrito brevemente como segue. Uma amostra aleatória de tamanho k (igual ou superior a 5.000), $\rho_i, i = 1, 2, \dots, k$, é retirada da densidade a priori de ρ , que neste trabalho é considerada uma distribuição beta, com parâmetros a e b , $\text{Beta}(a, b)$. Um peso ω_i , é associado a cada $\rho_i, i = 1, 2, \dots, k$, proporcional ao lado direito da Eq. (7). Finalmente, uma reamostra de tamanho k com reposição, ρ_i^* , é feita dos ρ_i , de acordo com os pesos ω_i . Dessa forma, podem-se ser calculadas as medidas usuais da inferência bayesiana. Por exemplo, um estimador pontual é a média de ρ_i^* . Uma implementação simplificada do SIR em R pode ser vista na Listagem 1.

```

1 sirmmsRo<-function(ce, amostra, a, b){
  # amostragem da priori
3  sirSampSiz<-5000
  rhoPrior<-rbeta(sirSampSiz, shape1=a, shape2=b)
5 # função de probabilidade, Eq.(3)
  dmms<-function(m, rho, ce){
7  p0<-sum(sapply(c(0:(ce-1)), function(j) (ce*rho)^j/factorial(j))+
          ce^ce*rho^ce/factorial(ce)/(1-rho)
9  p0<-1/p0
  p<-sapply(floor(m), function(m) if(m<=ce) (ce*rho)^m/factorial(m)
11  else ce^ce*rho^m/factorial(ce))*p0
  return(p)}
13 # função de verossimilhança
  like.f<-function(rho, ce, amostra){
15  ss<-length(amostra)
  likef<-sapply(1:length(rho), function(n) prod(dmms(amostra[1:ss], rho[n], ce)))
17  return(likef)}
  # SIR
19  maxss<-50
  init<-1
21  fin<-maxss
  ssize<-length(amostra)
23  rhoPost<-rhoPrior
  while (init<ssize) {
25  subsample<-amostra[init:min(fin, ssize)]
  wght<-like.f(rhoPost, ce, subsample)
27  wght<-wght/sum(wght)
  rhoPost<-sample(rhoPost, size=sirSampSiz, replace=TRUE, prob=wght)
29  init<-min(fin, ssize)+1
  fin<-fin+maxss}
31  return(mean(rhoPost))}

```

Listagem 1: Estimador bayesiano para ρ via SIR e distribuição a priori beta

Sobre a Listagem 1, é necessário chamar à atenção suas linhas finais (24 a 30), que dividem a amostra original em partes de tamanho máximo igual a 50 e, iterativamente, calculam posterioris intermediárias, que serão utilizadas como prioris no passo seguinte, até considerar todos os elementos da amostra original. Este processo bayesiano sequencial é necessário para evitar erros numéricos que tornam a função de verossimilhança praticamente plana, e, em consequência, anulam os pesos ω_i , utilizados no SIR. Isso ocorre com frequência quando as amostras têm tamanhos superiores a $\ell > 50$.

Resultados e discussão

A Tabela 1 apresenta os valores médios e os desvios-padrões para a média a posteriori da intensidade de tráfego (ρ), obtida a partir de um procedimento de simulação Monte Carlo repetido 100 vezes com o tamanho da amostra do SIR igual a $k = 5.000$, para vários tamanhos de amostra, ℓ , e $c = 3$.

Dá para perceber pela Figura 3-a que os erros médios de estimação mantêm-se próximos de zero para todos os valores de ρ quando tem-se a priori não-informativa, Beta(1, 1), e que as prioris informativas têm melhor desempenho quando as modas são próximas ao valor ρ a ser estimado. Na Figura 3-b, pode-se observar que o desvio-padrão é maior para ρ igual a 0,5 e menor quando ρ é igual a 0,9. Vê-se também, na Figura 3-c, que os erros médios de estimação diminuem à medida que o tamanho da amostra aumenta, para todas as prioris testadas, para amostras de tamanho $\ell < 100$. A partir desse tamanho de amostra os erros começam a aumentar para a priori Beta(2, 5) e, para amostras de tamanho $\ell > 150$, os erros começam a aumentar para a priori Beta(1, 1), o que levanta a suspeita de que a dificuldade de cálculo da função de verossimilhança para amostras grandes começa a crescer. De fato, à medida que a amostra cresce, a função de verossimilhança apresenta multiplicações de números muito próximos de zero, fazendo com que a posteriori não incorpore a informação proveniente dos dados. Por isso é

Tabela 1: Média e o desvio-padrão (entre parênteses) para a intensidade de tráfego (ρ) a posteriori.

Priori	ρ	Tamanho da amostra (ℓ)					
		25	50	100	150	200	250
Beta(2;5)	0,20	0,2096 (0,0476)	0,2052 (0,0296)	0,2029 (0,0227)	0,2020 (0,0173)	0,2002 (0,0164)	0,1985 (0,0146)
	0,50	0,4744 (0,0603)	0,4891 (0,0404)	0,4937 (0,0307)	0,4954 (0,0235)	0,4943 (0,0218)	0,4933 (0,0186)
	0,90	0,8633 (0,0327)	0,9003 (0,0135)	0,8990 (0,0139)	0,8876 (0,0204)	0,8827 (0,0194)	0,8802 (0,0172)
Beta(1;1)	0,20	0,2101 (0,0508)	0,2046 (0,0304)	0,2024 (0,0227)	0,2016 (0,0177)	0,2001 (0,0164)	0,1983 (0,0148)
	0,50	0,4940 (0,0631)	0,4992 (0,0408)	0,4988 (0,0302)	0,4990 (0,0229)	0,4971 (0,0215)	0,4957 (0,0184)
	0,90	0,8924 (0,0209)	0,8963 (0,0135)	0,8977 (0,0101)	0,8986 (0,0073)	0,8983 (0,0067)	0,8981 (0,0056)
Beta(5;2)	0,20	0,2601 (0,0512)	0,2287 (0,0328)	0,2136 (0,0212)	0,2085 (0,0163)	0,2051 (0,0152)	0,2021 (0,0151)
	0,50	0,5169 (0,0575)	0,5098 (0,0393)	0,5033 (0,0300)	0,5016 (0,0232)	0,4995 (0,0216)	0,4976 (0,0187)
	0,90	0,8907 (0,0210)	0,8955 (0,0131)	0,8972 (0,0096)	0,8981 (0,0071)	0,8979 (0,0067)	0,8977 (0,0056)

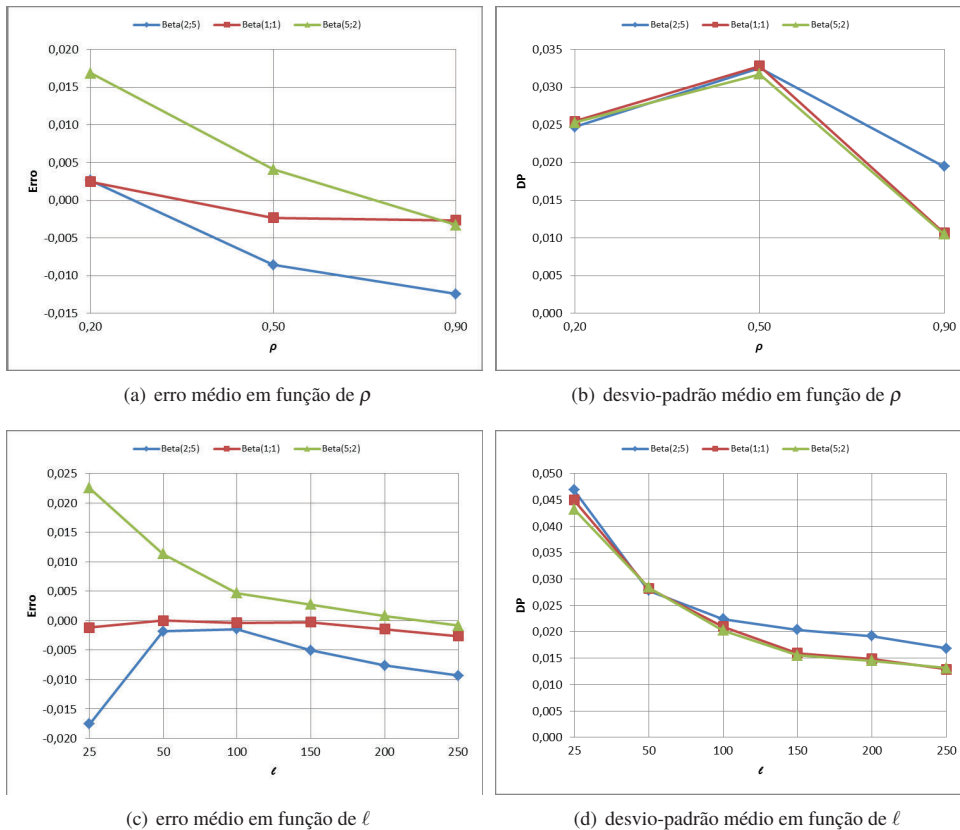


Figura 3: Desempenho médio das estimativas para ρ .

necessário lançar mão do processo bayesiano sequencial, mencionado na seção anterior, que permite a redução dos erros numéricos que podem tornar a função de verossimilhança praticamente plana em amostras grandes. Pode-se observar ainda, na Figura 3-d, que o desvio-padrão decresce em função do tamanho da amostra (ℓ), e que para as priors Beta(1, 1) e Beta(5, 2) o valor desse decréscimo foi bastante próximo. Em resumo, o desempenho do SIR é satisfatório para uma priori vaga, Beta(1, 1), para quaisquer valores de ρ , ou quando a priori tem moda próxima ao valor correto de ρ , conforme visto na Figura 3-a, onde as priors Beta(2, 5) e Beta(5, 2) têm o melhor desempenho quando $\rho = 0, 20$ e $0, 90$, respectivamente.

Conclusões

No contexto testado, o algoritmo implementado para estimação da intensidade de tráfego (ρ) de filas markovianas multi-servidoras ($M/M/c$), pelo método de amostragem/reamostragem de importância (SIR), no software R, produz resultados satisfatórios para amostra do número de usuários no sistema, em tempos aleatórios. Os erros médios já são inferiores a 0,002 quando o tamanho chega a 100. A partir desse tamanho de amostra, os erros não decaem mais, principalmente para uma priori com moda muito baixa, como a Beta(2, 5). Tem-se, portanto, uma segurança na utilização do SIR em situações reais, se a informação a priori é vaga ou se esta informação representa bem o valor a ser estimado para ρ . O ambiente R é bastante conveniente, por ser uma plataforma aberta, gratuita e que está tendo um uso cada vez mais difundido entre pesquisadores e usuários das áreas de matemática aplicada e de estatística.

Tópicos para trabalhos futuros incluem o estudo da influência do número de servidores (s) na qualidade da estimação, da aplicabilidade e eficácia do método SIR na estimação de outras medidas de desempenho (p.e., tamanho médio da fila, L_q , número médio de usuários no sistema, L , probabilidade de sistema vazio, p_0). Aplicações do método em estimação em filas truncadas $M/M/c/N$, ou filas mais gerais, $GI/G/c/M$ também podem render bons resultados.

Agradecimentos

Esta pesquisa foi parcialmente financiada pelo *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq), processo 304671/2014-2, e pela *Fundação de Amparo à Pesquisa do Estado de Minas Gerais* (FAPEMIG), processo CEX-PPM-00564-17, aos quais os autores expressam seus agradecimentos.

Referências

- ASANJARANI, A.; NAZARATHY, Y.; POLLETT, P. K. **Parameter and State Estimation in Queues and Related Stochastic Models**: A bibliography. Disponível em: <http://www.maths.uq.edu.au/~7Epk/papers/Qest/Qest.html>. Acesso em: 10 fev. 2017.
- CRUZ, F. R. B.; QUININO, R. C.; HO, L. L. Bayesian estimation of traffic intensity based on queue length in a multi-server $M/M/s$ queue. **Communications in Statistics - Simulation and Computation**, v.46, n.9, p.7319–7331, 2017.
- GROSS, D. et al.. **Fundamentals of Queueing Theory**. 4^a ed. New York, NY, EUA: Wiley-Interscience, 2009.
- GUPTA, A. K.; NADARAJAH, S. **Handbook of Beta Distribution and its Applications**. CRC press, 2004.
- KENDALL, D. G. Stochastic processes occurring in the theory of queues and their analysis by the method of embedded Markov chains. **Annals Mathematical Statistics**, v.24, p.338–354, 1953.
- R CORE TEAM. **R**: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2017. <http://www.R-project.org/>
- RUBIN, D. B. Using the SIR algorithm to simulate posterior distributions. **Bayesian Statistics**, v.3, p.395–402, 1988.
- SMITH, A. F. M.; GELFAND, A. E. Bayesian statistics without tears: A sampling-resampling perspective. **The American Statistician**, v.46, n.2, p.84–88, 1992.