UEPB
UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS I
CENTRO DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA
CURSO DE BACHARELADO EM ESTATÍSTICA

DÉBORA DE SOUSA CORDEIRO

ESTUDO DE SIMULAÇÃO DE UM SISTEMA COM FILA DO TIPO M/M/C COM DESISTÊNCIAS

CAMPINA GRANDE - PB
2020

DÉBORA DE SOUSA CORDEIRO

ESTUDO DE SIMULAÇÃO DE UM SISTEMA COM FILA DO TIPO M/M/C COM DESISTÊNCIAS

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de bacharel em Estatística.

Orientador: Profa. Dra. Divanilda Maia Esteves
Coorientador: Prof. Dr. Gustavo Henrique Esteves

CAMPINA GRANDE - PB
2020

C794e Cordeiro, Débora de Sousa.
Estudo de simulação de um sistema com fila do tipo M/M/C com desistências [manuscrito] / Debora de Sousa Cordeiro. - 2020.
27 p.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia , 2020.

"Orientação : Profa. Dra. Divanilda Maia Esteves , Coordenação do Curso de Estatística - CCT."

1. Filas markovianas. 2. Filas com desistência. 3. Simulação. I. Título

21. ed. CDD 519.2

Elaborada por Giulianne M. Pereira - CRB - 15/714 BC/UEPB

DÉBORA DE SOUSA CORDEIRO

ESTUDO DE SIMULAÇÃO DE UM SISTEMA COM FILA DO TIPO $M/M/C$ COM DESISTÊNCIAS

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de bacharel em Estatística.

Trabalho aprovado em 11 de Dezembro de 2020.

BANCA EXAMINADORA

Prof. Dr. Gustavo Henrique Esteves
(Coorientador)
Universidade Estadual da Paraíba (UEPB)

Tiago Almeida de Oliveira

Prof. Dr. Tiago Almeida de Oliveira
Universidade Estadual da Paraíba (UEPB)

Prof. Ms. Ednário Barbosa de Mendonça
Universidade Estadual da Paraíba (UEPB)

Dedico este trabalho primeiramente a Deus que me permitiu a realização de tudo e a meu querido amigo Bruno Farias Gomes (in memoriam) que, assim como muitos entes queridos, passou a me proteger de um outro plano.

AGRADECIMENTOS

Minha vida sempre foi regida por sonhos. Muitos deles pareciam utopia, mas cada um se tornou um degrau na escada de metas pessoais a serem vencidas.

Não sonhava em ser estatística, no entanto essa profissão floresceu dentro de mim como um carinho na alma. Relembrando tudo que passei, tantos desafios e muita luta, gostaria de fazer um agradecimento especial a Deus, guia de toda a minha trajetória nessa emocionante graduação.

A meu irmão Welson que é minha inspiração desde o dia que tornou-se meu pai aos seis anos de idade, quando ficamos órfãos de mãe e caímos em um novo mundo e, por isso, é o motivo das minhas buscas constantes em ser alguém melhor.

A meus pais/avós Dona Vera e Inácio Primo (*in memoriam*) que após cuidarem de onze filhos receberam de braços abertos mais dois netos para passar todo conhecimento e caráter que tinham, ensinando-nos desde muito cedo a enfrentar os desafios de frente, nunca perder a essência, nem esquecer de onde viemos.

A todos os tios e familiares que me incentivaram a estudar e buscar uma profissão, visto que tive a oportunidade de ir a uma escola, Corrinha, Aninha, Luciana, Isa, Jailson, Jacinto e meus tios do coração Alcimar, Tony e Jack. E também agradeço aos que me desmotivaram, pois foram combustível na minha luta diária.

Acredito que na vida de todos, na minha não seria diferente, além da família ser um grande pilar para que eu conseguisse chegar até aqui, um combustível fundamental na minha jornada foram meus amigos, que chegaram por acaso, mas se tornaram indispensáveis em meu coração. Dessa forma, faço meus agradecimentos a cada um de vocês.

Aos que me acompanham desde a infância Gizele e Iuri que, de certa forma, tornaram-se irmãos e são responsáveis por boa parte da pessoa que me tornei.

A Moisés e Demmily que chegaram por acaso, mas hoje me guiam e criticam quando preciso e por isso me mantém firme diante dos obstáculos.

Aos parceiros da graduação Hiago, Pedro, Leo, Iago, Lucas, Janaína e Adenilson que foram minhas companhias diárias e muitas vezes o abraço que acalentou minhas lágrimas, a força que eu precisava.

Os encontros do destino Mayara, Mateus, Gabriel e Elizandra que foram cruciais em determinados momentos.

Aos que não estão comigo diariamente mas que posso contar Francinildo e Ramon.

Aos demais colegas que o fio do destino foi trazendo e levando e de alguma forma me ajudaram.

A Vitor Brasil que desde os projetos iniciais na academia contribuiu seja descrevendo fórmulas ou dando apoio moral.

Não menos importante, faço minhas honras aos meus queridos professores de todo o período escolar que me lapidaram por meio dos conhecimentos compartilhados e motivação

à leitura, em especial Nildinho e tia Jailma.

A todos os professores da UEPB que tive o prazer de agregar conhecimento: Tanise, Ednário, Tiago, Ricardo, Ana Patrícia, Érika, Kléber, Juarez.

Aqueles que de alguma forma marcam de um jeito especial pela forma que passam conhecimento ou pelo carinho e respeito extra classe: meu coorientador Gustavo e minha amiga de outras vidas Vitória.

A diretora do Centro, Isabelle, por todo o carinho e pelo exemplo de mulher batalhadora que luta pelo direito de nós estudantes sem medir esforços e ao ex diretor e meu pai acadêmico Mará (*in memoriam*) pelo acolhimento no dia a dia e por sempre ter me apoiado nas lutas pelo movimento estudantil.

As pessoas que hoje me protegem de um outro plano espiritual mas, continuam habitando meu coração e deixam uma saudade enorme: minha mãe Luciene, Artulho, Mariana e Bruno.

A mulher que me inspirou desde a primeira vista por seu caráter ímpar e destaque numa área que ainda é muito masculinizada, minha orientadora Diana. A pessoa que teve paciência com minhas particularidades e meu tempo para realizar este trabalho e sem dúvidas a que mais agregou durante toda trajetória pela humildade e por nunca ter poupado esforços para me ensinar algo.

"Seu trabalho vai preencher uma grande parte da sua vida, e a única maneira de ficar realmente satisfeito é fazer o que você acredita ser um ótimo trabalho. E a única maneira de fazer um excelente trabalho é amar o que você faz."

(Steve Jobs)

RESUMO

Esse trabalho teve por objetivo principal escrever uma função para simular dados de uma fila $M/M/c$ com ênfase nas situações em que alguns clientes desistem de esperar e saem da fila. Além disso foi executado um estudo teórico do modelo implementado. Faz parte do cotidiano da maioria das pessoas, situações em que, ao se procurar por um serviço, precisa-se esperar para ser atendido. Essa possibilidade de espera resulta em uma fila, seja ela física ou não. Muitas vezes, as filas para realizar os atendimentos estão congestionadas, fazendo com que os clientes fiquem impacientes, levando-os muitas vezes a desistir do serviço. Em geral, embora seja cada vez mais comuns serviços nos quais o cliente é impaciente diante da fila de espera, não é fácil conseguir banco de dados que contenham informações acerca da desistência. Dessa forma, a simulação é uma técnica ferramental importante pois otimiza as análises de modelos dessa natureza por fornecer a geração desses dados, que normalmente não são fáceis de encontrar, a partir de funções que contenham as características do modelo desejado.

Palavras-chave: Filas markovianas. Filas com desistência. Simulação.

ABSTRACT

The main objective of this work was to write a function to simulate data from an M/M/c queue with an emphasis on situations in which some customers arrive to receive service and, due to their impatience with the waiting time, end up giving up the service. Also, a theoretical study of the implemented model was carried out. It is part of the daily lives of most people, situations in which, when looking for a service, one has to wait to be served. This possibility of waiting results in a queue, whether physical or not. Often, the lines to perform the calls are congested, making customers impatient, leading them to give up the service. In general, although it is increasingly common for services in which the client is impatient in front of the waiting line, it is not easy to obtain a database containing information about the withdrawal. Thus, simulation is an important tooling technique as it optimizes modeling analyzes of this nature by providing the generation of this data, which is usually not easy to find, from functions that contain the characteristics of the desired model.

Key-words: Markovian queues. Queues with withdrawal. Simulation.

SUMÁRIO

1 INTRODUÇÃO

Faz parte do cotidiano da maioria das pessoas, situações em que, ao se procurar por um serviço, precisa-se esperar para ser atendido, pois o que se deseja não está à disposição imediata. Essa possibilidade de espera resulta em uma fila, seja ela física ou não, formada, por exemplo, em supermercados, bancos, impressoras, serviços de atendimento ao consumidor. Muitas vezes, as filas para realizar os atendimentos estão congestionadas, fazendo com que os clientes fiquem impacientes, levando-os muitas vezes a desistir do serviço.

Segundo Magalhães (1996), as filas de espera são formadas quando clientes que chegam para receber um certo tipo de serviço são impossibilitados de um atendimento imediato. Diante disso, Fogliatti e Mattos (2007) definem a Teoria de Filas como uma modelagem analítica de processos ou até mesmo sistemas que resultam em espera. Essa teoria tem por objetivo avaliar e determinar quantidades, chamadas **medidas de desempenho**, que expressam a produtividade e operacionalidade desses processos. Tais quantidades são de suma importância, visto que são base para a tomada de decisão no que se refere a modificar ou até mesmo na manutenção da operacionalidade do sistema no estado atual. Elas também facilitam o dimensionamento de processos que geram as filas, entre outras coisas que visam um bom desempenho corroborando com a ideia de que a Teoria de Filas é fundamental nos quesitos de gerência e a administração de sistemas produtivos. Entre tais medidas pode-se citar o número de usuários no sistema e o tempo de espera pelo atendimento.

Em diversas circunstâncias, o usuário acaba desistindo de esperar por um serviço devido ao tempo de espera ser “longo”. Essa desistência, muitas vezes, é causada devido à falta de informação no sistema quanto ao início do atendimento por parte do cliente. A exemplo disso, Baccelli, Boyer e Hebuterne (1984) citam a busca de informações ou dúvidas, por ligação, referente a cartões de crédito, em que as filas são consideradas invisíveis e o cliente não poderá avaliar a velocidade do atendimento. Outra forma de interrupção da busca por algum serviço é considerar que o cliente já chega no sistema impaciente e por isso estará propenso a esperar apenas uma pequena parcela de tempo. Para avaliar essa impaciência do usuário, que tem sido comum nos serviços, faz-se necessário o cálculo de medidas específicas como o tempo de paciência do n-ésimo cliente, a taxa individual de abandono, o tempo virtual e real de espera, entre outras que irão dimensionar adequadamente a capacidade do sistema. Segundo Oliveira (2009), tais medidas permitem avaliar a eficácia do sistema no que se refere a um melhor atendimento e ao menor custo de operação gerando, assim, um equilíbrio entre a gerência e o usuário. Contudo, não se encontra a demonstração dessas medidas na literatura, portanto, elas não serão abordadas com tanta ênfase.

Em geral, embora seja cada vez mais comum serviços nos quais o cliente é impaciente diante da fila de espera, não é fácil conseguir bancos de dados que contenham informações acerca da desistência, especialmente porque o processo de coleta desse tipo de dados não é simples. Dessa forma, a simulação é uma técnica ferramental importante pois otimiza as análises de modelagem dessa natureza por fornecer a geração desses dados, que normalmente não são fáceis de encontrar, a partir de funções que contenham as características do modelo desejado. Ou seja, o uso de simulação possibilita a reprodução de modelos probabilísticos, de forma que permita a análise e inferência no mesmo.

Mesmo com a escassez de ferramental para analisar modelos de filas, encontra-se na literatura estudos relevantes utilizando como base para a tomada de decisão a Teoria de Filas, consolidando sua importância. A exemplo desses estudos, pode-se citar o uso dessa teoria para otimizar as filas de atendimento numa loja de artigos do Aeroporto Internacional de Guarulhos - SP por Correia et al. (2019), bem como a aplicação de um modelo de teoria de filas em um restaurante universitário a fim de estudar o tempo de atendimento do local, proposto por Soares et al. (2016) e a avaliação de desempenho, com base nessa teoria, de uma rede de comunicação congestionada de uma Universidade por Vazquez, Morabito e Marcondes (2018).

Deste modo, esse trabalho teve por objetivo principal escrever uma função para simular dados de uma fila $M/M/c$ com ênfase nas situações em que há clientes que chegam para receber um serviço e, devido a sua impaciência quanto ao tempo de espera, acabam desistindo do atendimento.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo serão apresentados alguns tópicos importantes que foram estudados acerca da teoria das Filas. Inicialmente será apresentado o conceito de Processo de Nascimento e Morte, seguido das ideias gerais a respeito de um modelo de fila. Posteriormente, serão abordadas definições e resultados importantes referente a essa teoria, como as medidas de desempenho e, por fim, será apresentada uma síntese do modelo escolhido para esse estudo e particularidades deste ao se deparar com situações envolvendo desistência e/ou impaciência de clientes nas filas de espera.

2.1 O Processo de Nascimento e Morte

Um processo estocástico é uma sequência de variáveis aleatórias $\{X(t) \mid t \in T\}$ que descreve a evolução de um processo ao longo do tempo. Em particular, diz-se que um processo estocástico é markoviano (de ordem 1) quando a distribuição de probabilidade condicional de $X(t)$ para os valores $X(t_0), X(t_1) \ldots X(t_n)$ depende apenas de $X(t_n)$ dada uma sequência de tempos $t_0 < t_1 < \ldots < t_n < t$, implicando afirmar que para prever a posição do processo no futuro apenas a informação do presente é relevante. Ou seja,

$$P\left[X(t) \leq x \mid X(t_n) = x_n, X(t_{n-1}) = x_{n-1}, \ldots, X(t_0) = x_0\right] = P\left[X(t) \leq x \mid X(t_n) = x_n\right].$$

Quando as variáveis $X(t)$ são discretas, então tem-se uma cadeia de Markov. Neste caso, para $s > 0$, a probabilidade

$$P_{x,y}(t, t+s) = P\left[X(t+s) = y | X(t) = x\right]$$

é chamada probabilidade de transição. Em particular, se $P_{x,y}(t, t+s)$ não depende de t, então a cadeia é homogênea no tempo e então a probabilidade de ir de um estado x para um estado y em uma unidade de tempo será denotada simplesmente por $P(x, y)$. Diz-se ainda que uma cadeia é irredutível quando todos os estados do processo podem ser alcançados pelos demais.

No contexto deste trabalho, será explorado uma cadeia de Markov em particular que é o **processo de nascimento e morte**. Para esse processo considera-se que os eventos referente aos nascimentos e as mortes são estatisticamente independentes e que, se o sistema está no estado n, em um intervalo de tempo $(t, t+\Delta t)$, em que Δt é tão pequeno quanto se queira, a probabilidade de ocorrência de:

- um nascimento será dada por $\lambda_n \Delta t + o(\Delta t)$;
- uma morte será dada por $\mu_n \Delta t + o(\Delta t)$;
- nascimento(s) e/ou morte(s), que resultam em mais de um evento simultaneamente, são desprezíveis, sendo igual a $o(\Delta t)$, em que,

$$\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0.$$

O processo de nascimento e morte é homogêneo, irredutível e permite transições de um estado n, em uma unidade de tempo t, apenas para os vizinhos imediatos, sendo classificado como nascimento quando a transição é realizada para o estado $n+1$, bem como uma morte quando essa transição é, por sua vez, para o estado $n-1$. Esse processo será considerado, para todo $n > 0$, com taxas:

$$
\begin{aligned}
P\,(n, n+1) &= \lambda_n. \\
P\,(n, n-1) &= 1 - \lambda_n = \mu_n. \\
P\,(n, k) &= 0 \quad \forall k \neq n-1, n+1.
\end{aligned}
$$

Aqui, λ_n e μ_n são chamadas, respectivamente, de taxa de nascimento e taxa de morte.

Por meio dessas taxas, obtém-se a distribuição limite dos estados do sistema $(P_0, P_1, P_2 ...)$, também definida como a distribuição do estado do regime estacionário, completamente determinada pelas taxas de nascimento e morte. Para obter tal distribuição, inicialmente denominamos as equações de balanço ou de equilíbrio que utiliza o fato de que para estado, "o fluxo que entra é igual ao fluxo que sai". Em outras palavras (ROSS, 1995) tem-se que para qualquer estado $n \geq 1$:

$$\lambda_n P_n + \mu_n P_n = \lambda_{n-1} P_{n-1} + \mu_{n+1} P_{n+1},$$

ou seja,

$$\lambda_n P_n - \mu_{n+1} P_{n+1} = \lambda_{n-1} P_{n-1} - \mu_n P_n.$$

Para o estado 0,

$$\lambda_0 P_0 = \mu_1 P_1. \tag{2.1}$$

Recursivamente,

$$
\begin{aligned}
\lambda_n P_n - \mu_{n+1} P_{n+1} &= \lambda_{n-1} P_{n-1} - \mu_n P_n \\
&= \lambda_{n-2} P_{n-2} - \mu_{n-1} P_{n-1} \\
&= \dots \\
&= \lambda_0 P_0 - \mu_1 P_1 \\
&= 0,
\end{aligned}
$$

sendo que a última igualdade é consequência da Equação (2.1). Assim,

$$P_{n+1} = \frac{\lambda_n}{\mu_{n+1}} P_n,$$

que é uma fórmula de recorrência cujo termo geral é

$$P_n = P_0 \prod_{i=1}^{n} \frac{\lambda_{i-1}}{\mu_i} \quad n \geq 1. \tag{2.2}$$

Como $\sum\limits_{n \geq 0} P_n = 1$, conclui-se que

$$P_0 = \frac{1}{1 + \sum\limits_{n \geq 1} \prod\limits_{i=1}^{n} \frac{\lambda_{i-1}}{\mu_i}}. \tag{2.3}$$

2.2 O Modelo de Fila

Segundo Fogliatti e Mattos (2007), um sistema com fila tem sua composição física formada por um espaço estabelecido para a espera, canais ou postos que prestarão serviços/atendimentos e por usuários. Estes usuários chegarão caracterizando um processo de chegadas e devem ser atendidos nos postos de serviço segundo um padrão definido por meio das disciplinas pela gerência. Nos momentos em que os postos estão sendo utilizados, os demais clientes aguardam atendimento numa fila única no espaço atribuído para a espera. Assim, quando o(s) canal(is) fica(m) livre(s), o usuário que está na fila é convidado para o atendimento e liberado do sistema logo que o serviço seja concluído.

De acordo com Magalhães (1996), a forma como os usuários procuram um serviço descreve um processo de chegadas, o qual é classificado como determinístico quando se tem conhecimento prévio acerca dos instantes de tempo que as chegadas ocorrem e o número destas. Por sua vez, quando não se sabe essas quantidades tem-se um processo aleatório, caracterizado por uma distribuição de probabilidade, em que o parâmetro taxa de chegadas corresponde ao número médio de usuários que ingressam no sistema por unidade de tempo. O processo de atendimento, é análogo ao processo de chegadas, com a distribuição dependendo do estado do sistema ou do tipo do usuário, em que tal processo é estabelecido pelo modo do fluxo destes usuários que são atendidos.

Quanto à maneira como os clientes serão selecionados para receber um certo tipo de serviço quando um posto fica disponível usa-se uma disciplina de atendimento pré-estabelecida pela gerência do sistema. Na literatura, encontram-se comumente as seguintes rotinas de atendimento:

- FCFS (do inglês, *first come first served*) ou FIFO (do inglês, *first in - first out*), onde os atendimentos são de acordo com a ordem de chegada. Essa é a rotina mais comum e aparece em situações como filas de supermercado e atendimento bancário.
- LCFS (*last come first served*) ou LIFO (*last in - first out*), em que o usuário que será atendido primeiro é o que chegou por último, como modelos de arquivo, busca por disco rígidos ou contêineres em navios.
- PRI (*priority service*), no qual o atendimento terá algum tipo de prioridade estabelecida pela gerência do sistema, como cirurgias de alto risco, idosos, etc. Nessa disciplina, é necessário que seja especificado como se dará a ordem de atendimento dentro dessa mesma classe de prioridade e, nesses casos, utiliza-se a FCFS ou FIFO;
- SIRO (*service in random order*), referente ao atendimento aos usuários de forma aleatória sendo independente do tempo de chegada dos mesmos, como a contemplação de consórcios.

Os postos de serviço também podem ser chamados de canais de atendimento e definem-se como os locais designados para atender o usuário que busca por serviço. Estes podem ser físicos ou não, como guichês e centrais de atendimento *online*, e podem ser finitos ou infinitos.

O número máximo de clientes que um sistema pode comportar é definido pela capacidade do sistema que também pode ser finita ou infinita. A capacidade designada em um sistema inclui a quantidade de clientes na fila e no atendimento. Quando esta capacidade é do tipo finita, os usuários que tentam ingresso após ela ser atingida são rejeitados.

Para descrever um sistema com fila será utilizada apenas a notação proposta por Kendall (1953), dada por A/B/C/D/E, em que as letras A e B referem-se, respectivamente, às distribuições dos tempos entre chegadas sucessivas e de atendimento, enquanto C denota o número de postos de atendimento em paralelo, D diz respeito à capacidade física do sistema adotado e E representa a disciplina de atendimento dos clientes no serviço. Quando D e E são omitidas, será considerado que o sistema possui capacidade infinita e disciplina FCFS/FIFO. Para as distribuições de A e B, as notações mais utilizadas são:

- M: Distribuição de Probabilidade Exponencial (Markoviana);
- D: Distribuição determinística ou degenerada;
- G: Distribuição geral, em que não há especificação;
- E_k: Distribuição Erlang do tipo k.

As medidas de desempenho ou medidas de operacionalidade são a forma como a Teoria de Filas avalia a eficiência de um sistema analisando suas respectivas características. Como estas normalmente se alteram ao longo do tempo, serão representadas por variáveis aleatórias, no qual, seus valores esperados podem ser utilizados como medidas do sistema quando o regime for estacionário. Esse regime é dado no intervalo de tempo de funcionamento do sistema $[t^*, t)$, em que t^* é o instante no qual tais medidas se mantêm estáveis. Dentre essas medidas tem-se:

- Número médio de usuários na fila (L_q) e no sistema (L).
- Tempo médio de espera de um usuário qualquer na fila (W_q).
- Tempo médio de permanência de um usuário qualquer no sistema (W).
- Probabilidade de se ter no máximo um número n_0 pré fixado de usuários no sistema, $P(N \leq n_0)$.
- Probabilidade de um usuário qualquer ter que aguardar mais do que um determinado tempo t na fila, $P(T_q > t)$.

- Probabilidade de se ter algum servidor ocioso em um sistema com c postos de atendimento, $P(N < c)$.

Vale salientar que considera-se N como o número de usuários no sistema e T_q o tempo de espera de um usuário qualquer na fila.

2.3 O Modelo $M/M/1$

O modelo M/M/1, que também pode ser escrito por M/M/1/∞/FCFS em sua forma sem abreviação, é caracterizado por ter o tempo entre chegadas sucessivas e o tempo de atendimento ambos com distribuições de probabilidade exponenciais, pela existência de um único servidor, por não haver limitação no que se refere ao espaço destinado para a fila de espera e, por fim, a ordem em que os usuários acessam o serviço acontece de acordo com a ordem de chegada destes ao sistema.

Esse modelo possui as taxas constantes de ingresso ao sistema e de atendimento dadas, respectivamente, por

$$\lambda_n = \lambda, \quad \forall n \geq 0 \quad \text{e} \quad \mu_n = \mu, \quad \forall n \geq 1.$$

Consequentemente, substituindo essas taxas em (2.2) e (2.3) tem-se:

$$P_n = \frac{\lambda^n}{\mu^n} P_0, \quad \forall n \geq 1 \quad \text{e} \quad P_0 = \left[\sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu} \right)^n \right]^{-1}.$$

Se $\frac{\lambda}{\mu} < 1$, a série em P_0 converge e então se conclui que

$$P_0 = 1 - \frac{\lambda}{\mu},$$

sendo que $\rho = \frac{\lambda}{\mu}$ é denominado a taxa de ocupação do sistema.

2.4 O Modelo $M/M/c$

Assim como o modelo anterior, este possui o tempo entre chegadas sucessivas e o tempo de atendimento seguindo distribuições de probabilidade exponenciais, entretanto, há c servidores disponíveis. Vale salientar que este modelo não contém restrição quanto o número de usuários na fila de espera e que estes ingressam no serviço pela ordem de chegada no sistema.

Por conter as chegadas e atendimentos caracterizando um processo de nascimento e morte, as respectivas taxas são:

$$\lambda_n = \lambda, \quad \forall n \geq 0$$

e

$$\mu_n = \begin{cases} n\mu, & \text{se } 1 \leq n < c, \\ c\mu, & \text{se } n \geq c. \end{cases}$$

A taxa de utilização do sistema é dada por

$$\rho = \frac{r}{c} = \frac{\lambda}{c\mu},$$

Substituindo essas taxas acima em (2.2) e (2.3), tem se:

$$P_n = \begin{cases} P_0 \dfrac{r^n}{n!}, & 1 \leq n < c, \\ \\ P_0 \dfrac{r^n}{c^{n-c}c!}, & n \geq c, \end{cases}$$

com

$$P_0 = \left(\sum_{n=0}^{c-1} \frac{r^n}{n!} + \frac{cr^c}{c!(c-r)} \right)^{-1}.$$

2.5 O Modelo com Desistência/Impaciência

O modelo $M/M/c$ considerando a desistência é um diferencial ao comum, pois considera-se que estas desistências podem ser causadas por uma impaciência por parte do cliente que pode estar ligada a uma tolerância para o tempo de espera em um determinado serviço, bem como no que diz respeito em como a gerência do sistema está sendo eficaz e eficiente. Neste estudo, considera-se um sistema de filas markovianas (tempos entre chegadas e de atendimentos exponenciais) operando com taxa de chegada λ e taxa de atendimento μ e c servidores, no qual, o usuário que busca por um dado serviço desiste sempre que o tempo de espera na fila for superior ao tempo que ele está disposto a aguardar o atendimento. Esse tempo que o usuário fica esperando é chamado **tempo de tolerância ou paciência** (OLIVEIRA, 2009).

O estudo desse modelo deve-se ao fato de eles serem frequentes em situações práticas. Frequentemente, o cliente aguarda para receber atendimento por um tempo limitado e deixa tal serviço quando não é atendido durante esse período, levando-o a decidir (de forma subjetiva) se o serviço buscado "vale o tempo" de espera. Há um entendimento que, além da paciência e necessidade do usuário, o tempo de tolerância também evidencia a eficiência da gerência e operacionalidade do sistema. Observa-se desistência gerado pela impaciência do usuário em filas de bancos, postos de gasolina, supermercados, entre outros. Kok e Tijms (1985) também citam sistemas de inventário com mercadoria perecível. Eles afirmam que o primeiro modelo de fila considerando a impaciência do cliente, em que os serviços seguiam uma função densidade de probabilidade exponencial foi estudado por Gnedenko e Kovalenko (1968).

Segundo Oliveira (2009), o modelo de fila Erlang - A é o mais simples que considera o abandono de clientes pois enfatiza que estes chegam impacientes no sistema com os respectivos tempos de paciência seguindo uma distribuição Exponencial (de média ω^{-1}). Este também pode ser representado por $M/M/n + M$. Garnett, Mandelbaum e Reiman

(2002 apud OLIVEIRA, 2009) estudaram esse modelo e constataram boas aproximações do mesmo para as medidas de desempenho estudadas. Brown et al. (2005) falam ainda que o tempo de paciência pode ser modelado usando outras distribuições que não a exponencial.

Na seção anterior, foram citadas algumas das medidas de desempenho mais utilizadas para examinar quão bem um sistema está operando. Quando se leva em conta a tolerância, novas medidas passam a ser relevantes para dimensionar adequadamente o impacto causado quando um cliente desiste de um serviço. A ocupação e as medidas de desempenho envolvendo abandono e espera quantificam o nível da operacionalidade do serviço que é de suma importância para a gerência (FEINBERG et al., 2000). No entanto, há uma dificuldade visível em encontrar, na literatura, a demonstração dessas medidas. Dentre estas, tem-se a fração de cliente perdido, o atraso médio da fila de um cliente, o tempo virtual de espera e, talvez a mais importante, probabilidade de abandono. Vale salientar que nesse trabalho não serão utilizadas as medidas de desempenho relacionadas a desistência do usuário inerentes a esse modelo, portanto, será omitido o cálculo destas. Como estes serviços possuem um problema para serem analisados devido a alta complexidade de coleta dos dados, a simulação passa a ser uma técnica viável para a modelagem desses sistemas.

3 SIMULAÇÃO

Problemas relacionados a serviços com desistências devido ao tempo de espera pelo atendimento ser superior a tolerância do cliente são rotineiros. Os avanços tecnológicos na área computacional proporcionam a simulação de sistemas desta natureza, possibilitando conhecer melhor seu funcionamento e favorecendo a tomada de decisões. A procura por modelagem e em especial pela simulação tem crescido, principalmente para a análise e desenvolvimento de sistemas (FURLANETTO, 2016). Essa ferramenta (simulação) contribui em termos de eficiência e permite a modificação nas características de um dado sistema em estudo para a análise e avaliação sem que este seja necessariamente físico. Ela pode ter um papel importante de investigar como um sistema se comporta (VAZQUEZ; MORABITO; MARCONDES, 2018).

Segundo Sakurada e Miyake (2003), a simulação trata-se de uma ferramenta baseada em modelos para reproduzir um sistema estudado e, então, atuar na resolução de problemas em que é inviável solucionar analiticamente.

Contudo, Furlanetto (2016) atenta para o fato de que não há ferramentas consolidadas disponíveis que gerem simuladores de filas. Dessa forma neste trabalho, serão simuladas filas M/M/c com desistências em que os tempos entre chegadas sucessivas, tempos de atendimento e tempos de tolerância seguem uma distribuição Exponencial com taxas λ, μ e τ, respectivamente. Para tal foi criada a função `r.MMC_pacience` utilizando o *software* R (R Core Team) (2016). A função mencionada já está disponível no pacote `r.queuepb` e pode ser obtido em <https://github.com/mayara124/queuepb>. Para a instalação no *software R* inicialmente instala-se o pacote `devtools` pelos comandos

```
>install.packages("devtools")
>devtools::install_github("mayara124/queuepb").
```

Essa função, utilizando operadores condicionais e de *loop*, possui recursos que viabilizam a modelagem de sistemas considerando clientes impacientes representando, adequadamente, um sistema que na prática é complexo e, portanto, apresenta ao usuário como este funciona no cotidiano. Ela simula bancos de dados contendo informações acerca dos tempos de chegada, de começo do serviço, de saída e guichê que normalmente são muito difíceis de coletar. Assim, a partir dos parâmetros de entrada informados pelo usuário - número de clientes, guichês, taxas médias de chegada, atendimento e tolerância - o compilador oferece como saída uma matriz com as informações, em que as colunas representam os tempos descritos acima e cada linha o n-ésimo cliente.

A seguir será representado o *script* da função e posteriormente a sequência de procedimentos utilizados para especificar, parametrizar e simular o modelo de estudo:

```
r.MMC_pacience <- function(n,c,lambda,mu,tau){
```

```
##Vetor para os tempos de saída de cada guichê
saida.guiches = vector("numeric", c)

##n tempos entre chegadas sucessivas com taxa lambda
time.arrival <- rexp(n,1/lambda)
tdc <- comsum(time.arrival)   #tempos de chegadas

##n tempos de atendimento com taxa mu
time.service <- rexp(n,1/mu)

##n tempos de tolerÂncia com taxa tau
time.tolerance <- rexp(n,1/tau)

##Vetores de armazenamento:
saida <- NULL            ##Tempos de saida
tcs <- NULL              ##Tempos de começo de serviço
guiche <- NULL           ##Guiche de atendimento
tempo.waitting <- NULL   ##Tempo de espera

##Função auxiliar para evitar o problema de usar o sample diretamente
resample <- function(x, ...) x[sample.int(length(x), ...)]

for(i in 1:n) {
idx = which(saida.guiches < tdc[i])  #Verifica se existem guiches livres

if(length(idx)> 0) { #Caso exista
atendente <- resample(idx, 1)   #Sorteia um entre os livres
tcs[i] <- tdc[i]  #Inicio do serviço coincide com a chegada
guiche[i] = atendente
saida[i] <- saida.guiches[atendente] <- tcs[i] + time.service[i]
}
else { #Caso todos estejam ocupados
atendente <- which.min(saida.guiches)  #Pega o primeiro a desocupar

tcs[i] <- saida.guiches[atendente]  #Inicio do serviço quando desocupar o 1

tempo.waitting[i] = tcs[i] - tdc[i]
if(tempo.waitting[i] < time.tolerance[i]) {
```

```
#Tempo de espera menor que a tolerância
guiche[i] = atendente
saida[i] <- saida.guiches[atendente] <- tcs[i] + time.service[i]
}
else { #Caso contrário, o cliente desiste
tcs[i] <- guiche[i] <- NA
saida[i] <- tdc[i] + time.tolerance[i]
}
}

}

##Matriz com os valores simulados
return(cbind(tdc, tcs, saida, guiche))

}
```

No corpo da função o processo de simulação é dividido de acordo com as estruturas condicionais e de *loop*. O início do código possui um vetor para guardar os parâmetros de entrada, em que os tempos simulados seguem uma distribuição exponencial, bem como uma função auxiliar para gerar os números de simulações desejadas.

Em seguida, há um operador de *loop* com o comando **for** para que as condições do modelo desejado sejam testadas para o número de simulações estabelecidas. Na primeira estrutura condicional testa-se o caso em que há guichês livres de imediato, sorteando um entre os habilitados e fazendo o cálculo dos tempos de início de serviço e de saída, bem como guardando a informação do guichê ao qual o usuário foi atendido. Para o caso dos guichês estarem ocupados, na segunda estrutura condicional o comando `which.min` destinará o cliente simulado para o primeiro guichê a desocupar, em seguida calculará o tempo no qual o serviço começou e o tempo de espera do n-ésimo usuário.

Dentro dessa estrutura há um outro operador para solucionar a questão da desistência. No primeiro bloco verifica-se se o tempo de espera é menor que o tempo de tolerância, caso em que o cliente permanece no sistema. No segundo, considera-se os casos em que o tempo de espera foi maior ou igual à tolerância, o que levará o cliente a abandonar o serviço. Se o n-ésimo cliente desiste, o tempo de começo de serviço e guichê de atendimento receberão na matriz o código NA e, portanto, o tempo de saída será apenas o tempo de chegada adicionado ao tempo de tolerância.

Por último há o comando `return`, responsável por fazer a função retornar a matriz com o modelo de fila simulado. Desse modo, após a compilação é obtido um banco de dados que representa um sistema de filas com desistências. Vale salientar que a função também permite calcular o tempo na fila utilizando as duas primeiras colunas, bem como o tempo de atendimento com os tempos de saída e do começo dos serviços.

4 RESULTADOS E DISCUSSÃO

A fim de visualizar a matriz de saída do banco de dados simulado, um exemplo será mostrado a seguir. Suponha que se deseja simular um sistema de fila com 2 guichês de atendimento, no qual passaram 30 usuários. Considere que os tempos entre chegadas sucessivas e de atendimento tem taxas médias $\lambda = 5$ minutos e $\mu = 4$ minutos, respectivamente, e que a taxa de tolerância $\tau = 2$ minutos.

Neste caso, o comando usado seria

```
>r.MMC_pacience(30,2,5,4,2)
```

Como o *software* R não tem semente fixa, cada vez que esse comando for compilado, o resultado variará. Abaixo tem-se uma saída que foi obtida.

```
          tdc        tcs       saida     guiche
[1,]    2.107462   2.107462   3.910068      2
[2,]    4.270652   4.270652   5.144090      1
[3,]    4.991835   4.991835  12.558764      2
[4,]   20.995457  20.995457  24.206657      1
[5,]   23.962813  23.962813  28.317463      2
[6,]   25.543120  25.543120  25.667078      1
[7,]   25.961526  25.961526  28.344772      1
[8,]   26.826635         NA  27.620739     NA
[9,]   43.916696  43.916696  48.570289      2
[10,]   45.820256  45.820256  53.836613      1
[11,]   48.508842  48.570289  57.820423      2
[12,]   50.385516         NA  51.010247     NA
[13,]   53.522652  53.836613  55.475387      1
[14,]   55.506252  55.506252  59.752504      1
[15,]   66.086124  66.086124  69.448188      2
[16,]   67.237384  67.237384  69.750859      1
[17,]   76.840437  76.840437  76.871140      2
[18,]   92.210930  92.210930  97.437063      2
[19,]   95.174793  95.174793  95.494829      1
[20,] 102.904233 102.904233 103.137845      1
[21,] 107.055745 107.055745 109.241824      2
[22,] 112.530833 112.530833 113.441310      2
[23,] 113.395707 113.395707 136.858294      1
[24,] 116.362382 116.362382 120.272776      2
[25,] 118.705526 120.272776 120.597219      2
[26,] 119.173001         NA 119.420111     NA
```

```
[27,] 136.411855 136.411855 136.571525     2
[28,] 137.100559 137.100559 140.354564     2
[29,] 137.835009 137.835009 148.206784     1
[30,] 137.917234         NA 139.067396    NA
```

Em cada linha há quatro valores referentes a um usuário:

- **tdc:** tempo de chegada, ou seja, quantos minutos se passaram entre o momento que o sistema começou a funcionar e a chegada desse cliente;
- **tcs:** quantos minutos se passaram entre o momento que o sistema começou a funcionar e o instante em que o cliente começa a ser atendido;
- **saida:** quantos minutos se passaram entre o momento que o sistema começou a funcionar e o instante em que o cliente termina o atendimento;
- **guiche:** número do guichê no qual o cliente foi atendido.

Nas linhas 8, 12, 26 e 30 os clientes saíram da fila antes de chegar sua vez de ser atendidos, ou seja, houve desistência e por isso os NA's em tcs e guiche.

O uso de ferramentais como a simulação é de suma importância para a otimização de análises estatísticas como os modelos de filas, pois tais técnicas trazem a rapidez e precisão que a forma analítica por vezes é inviável. Um trabalho de simulação, que utiliza um modelo de filas de maneira complementar, pode dar a exata dimensão das necessidades de aumento de estrutura (SOARES et al., 2016).

Apesar da modelagem de filas ser cada vez mais comum, não há uma diversidade de pacotes que possam analisar esse tipo de sistema. Por isso, foi criada a função `r.MMC_pacience` no pacote `r.queuepb` para a simulação de modelos de filas com desistências permitindo, agora, a interação com sistemas virtuais que dão uma visão de como funcionam serviços em que os usuários são impacientes e, portanto, a possibilidade de implementar melhorias.

Correia et al. (2019) analisaram filas em pontos de vendas do aeroporto de Guarulhos-SP com o auxílio da simulação permitindo a organização e redução de filas nesse local. Vazquez, Morabito e Marcondes (2018) utilizaram a simulação para avaliar o desempenho e propor soluções numa rede de comunicação congestionada de uma universidade. Furlanetto (2016) desenvolveu um gerador de simuladores baseado em filas. Estes autores utilizaram a simulação com o auxílio de *softwares* e pacotes diferentes do proposto nesse trabalho o que consolida a importância do mesmo.

No desenvolvimento deste trabalho a maior dificuldade foi conseguir atender, na função, a ideia de reproduzir o funcionamento de um dado sistema que possui várias hipóteses, pois para cada uma delas ser representada, precisou-se de operadores executando cada uma das condições que precisavam estar bem correlacionadas para que a simulação retornasse informações confiáveis.

5 CONCLUSÃO

Este trabalho proporcionou o estudo de modelos de filas markovianas, bem como o processo de simulação de dados que representam serviços do cotidiano, no qual os usuários possuem uma certa tolerância quanto ao tempo de espera e por isso acabam desistindo.

O aumento da modelagem de filas, a dificuldade de coleta dos dados e a escassez de ferramental para simular modelos desta natureza, culminou com a implementação no *solftware* estatístico R da função `r.MMC_pacience` para simular sistemas que representam filas com desistências. Deve-se compreender que a criação desta função foi de suma importância, visto que o processo de elaboração expande os conhecimentos computacionais e consolida toda base teórica utilizada.

O objetivo ao propor a simulação das filas $M/M/c$ é reproduzir, de forma autêntica, o modelo e assim facilitar o processo de análise e inferência. Todavia, embora a função esteja finalizada, tem-se a necessidade, posteriormente, de implementar no pacote `r.queuepb` funções para o cálculo das medidas de desempenho do modelo trabalhado pois estas são importantes na tomada de decisão embora possuam um difícil acesso no que se refere à demonstração por serem dificilmente encontradas na literatura. Sendo assim, o uso da simulação neste estudo, que é utilizada comumente em outras modelagens mas não em sistemas considerando os clientes impacientes dentro do serviço, confirma a importância desse ferramental. Em análises bibliográficas, foi verificado uma gama de simuladores utilizados para reproduzir sistemas mas a maioria dos autores utilizavam outros *solftwares* como auxílio que não fosse o R, tampouco faziam o uso de funções específicas para filas markovianas ou era considerada a desistência de quem buscava pelo serviço.

Diante disso, foi interessante a escolha de um modelo divergente aos comumente apresentados na literatura voltado para a simulação com o uso de uma função inédita que está disponível para todos. Dessa forma, espera-se que a função desenvolvida neste trabalho tenha uma contribuição computacional para o avanço da modelagem de sistemas de filas. Entretanto há muito o que aprender sobre modelos de filas que considerem outras especificações, sobretudo modelos mais robustos. Portanto, o conhecimento aqui adquirido será designado para o progresso em modelagem com simulações dado o horizonte de expansão para pesquisa e estudo de filas e, assim este trabalho é uma opção a quem deseja adentrar nesse ramo da Estatística.

REFERÊNCIAS

BACCELLI, F.; BOYER, P.; HEBUTERNE, G. Single-server queues with impatient customers. *Advances in Applied Probability*, Cambridge University Press, v. 16, n. 4, p. 887–905, 1984. Citado na página 10.

BROWN, L. et al. Análise estatística de uma central de atendimento telefônico: uma perspectiva da ciência das filas. *Jornal da Associação Estatística Americana*, Taylor and Francis, p. 36–50, 2005. Citado na página 18.

CORREIA, P. F. da C. et al. Simulação em ponto de venda de varejo de viagens internacionais no aeroporto de são paulo. *Revista Produção Industrial e Serviços*, v. 6, n. 2, p. 30–41, 2019. Citado 2 vezes nas páginas 11 e 24.

FEINBERG, R. A. et al. Operational determinants of caller satisfaction in the call center. *International Journal of Service Industry Management*, MCB UP Ltd, 2000. Citado na página 18.

FOGLIATTI, M. C.; MATTOS, N. M. C. *Teoria de filas*. Rio de Janeiro: Interciência, 2007. Citado 2 vezes nas páginas 10 e 14.

FURLANETTO, G. C. Geração de simuladores de filas para diferentes contextos com estudo de casos para redes de computadores. Universidade Estadual Paulista (UNESP), 2016. Citado 2 vezes nas páginas 19 e 24.

GARNETT, O.; MANDELBAUM, A.; REIMAN, M. Designing a call center with impatient customers. *Manufacturing & Service Operations Management*, INFORMS, v. 4, n. 3, p. 208–227, 2002. Citado 2 vezes nas páginas 17 e 18.

GNEDENKO, B. V.; KOVALENKO, I. N. *Introduction to Queueing Theory*. [S.l.]: Israel Program for Scientific Translation, Jerusalem, 1968. Citado na página 17.

KENDALL, D. G. Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded markov chain. *The Annals of Mathematical Statistics*, JSTOR, p. 338–354, 1953. Citado na página 15.

KOK, A. G. D.; TIJMS, H. C. A queueing system with impatient customers. *Journal of Applied Probability*, Cambridge University Press, v. 22, n. 3, p. 688–696, 1985. Citado na página 17.

MAGALHÃES, M. N. *Introdução à rede de filas*. [S.l.]: 12 SINAPE Associação Brasileira de Estatística, 1996. Citado 2 vezes nas páginas 10 e 14.

OLIVEIRA, C. C. de. *Espera e abandono na fila M/M/n+ G e variantes*. Tese (Doutorado) — Universidade de São Paulo, 2009. Citado 3 vezes nas páginas 10, 17 e 18.

R (R Core Team). *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2016. Disponível em: <https://www.R-project.org/>. Citado na página 19.

ROSS, S. M. *Processos estocásticos*. [S.l.]: Wiley New York, 1995. v. 2. Citado na página 13.

SAKURADA, N.; MIYAKE, D. I. Estudo comparativo de softwares de simulação de eventos discretos aplicados na modelagem de um exemplo de loja de serviços. *XXIII ENEGEP-Encontro Nacional de Engenharia de Produção*, 2003. Citado na página 19.

SOARES, M. d. S. et al. Aplicação de um modelo de teoria das filas em um restaurante universitário: estudo do tempo de atendimento. Universidade Federal da Grande Dourados, 2016. Citado 2 vezes nas páginas 11 e 24.

VAZQUEZ, M.; MORABITO, R.; MARCONDES, C. Caracterização, modelagem e simulação de enlace congestionado de uma universidade. *Gestão & Produção*, SciELO Brasil, v. 25, n. 3, p. 583–594, 2018. Citado 3 vezes nas páginas 11, 19 e 24.