



**Universidade de Brasília
Departamento de Estatística**

Estudo da Teoria de Filas com Aplicações

Jéssica Fernandes de Araujo

Projeto apresentado para obtenção do título
de Bacharel em Estatística.

**Brasília
2015**

Jéssica Fernandes de Araujo

Estudo da Teoria de Filas com Aplicações

Orientador:
Prof. Dr. **LUCAS MOREIRA**

Projeto apresentado para obtenção do título de Bacharel em
Estatística.

Brasília
2015

SUMÁRIO

1 INTRODUÇÃO	5
2 RESUMO	7
3 CONCEITOS FUNDAMENTAIS	9
3.1 Teoria das Filas	9
3.2 Notações	10
3.3 Processos Estocásticos	11
3.3.1 Processos de Markov	12
3.3.2 Processos de Nascimento e Morte	12
3.4 Processos de Poisson	16
3.4.1 Processos de Poisson Homogêneo	16
3.4.2 Processos de Poisson Não Homogêneo	17
3.5 Distribuição Exponencial	18
3.6 Teste Kolmogorov-Smirnov de Aderência	19
3.7 Sistemas de Filas	20
4 FILAS MARKOVIANAS	23
4.1 Sistemas $M/M/1$	23
4.2 Sistemas $M/M/\infty$	25
4.3 Sistemas $M/M/c$	27
4.4 Sistemas $M/M/1/k$	28
4.5 Sistemas $M/M/c/k$	32
4.5.1 Sistemas $M/M/c/c$	35
5 APLICAÇÕES	37
5.1 Fluxos de paletes em indústria química	37
5.2 Processos de atendimento de emergência	42
6 SIMULAÇÃO PARA O MODELO $M/M/1$	49
6.1 Estimativas médias	49
6.2 Intervalos de confiança para as estimativas	50
7 CONSIDERAÇÕES FINAIS	51
REFERÊNCIAS	53

1 INTRODUÇÃO

O objetivo principal deste trabalho é apresentar aplicações da Teoria das Filas aos sistemas de Pronto Atendimento de um hospital e ao armazenamento de paletes em uma Indústria Química. Serão propostas parametrizações diferentes para diminuir o tempo de espera nessas filas e um serviço de boa qualidade. Faremos um estudo simulado utilizando o software R-project 3.30 a fim de comparar a eficácia das medidas de desempenho implementadas à fórmulas teóricas do modelo em que temos um único servidor, intervalos entre as chegadas e o tempo de atendimento exponencialmente distribuídos.

Com o acúmulo populacional e o aumento da demanda por serviços na sociedade moderna surge a necessidade de recursos para o bom fluxo de serviços, as filas. Filas são definidas como qualquer agregado de sujeitos que esperam por um atendimento e a formação destas ocorre quando a demanda de clientes excede a oferta por serviços. A importância de estudar esse processo se dá pela presença constante no nosso dia-a-dia, sejam em supermercados, lotéricas, agências bancárias, lojas, centrais telefônicas, no tráfego de veículos, transporte de cargas, sistemas de estocagem, processamento de informações, etc.

A Teoria das Filas é um campo específico da pesquisa operacional que permite medir a qualidade do atendimento para um determinado serviço. Apresenta um comportamento estocástico probabilístico, ou seja, considera-se que a chegada de unidades ao sistema é aleatória com certa distribuição, o atendimento é imediato quando o servidor está ocioso, e caso contrário, esperam na fila, saindo após serem atendidos.

O matemático dinamarquês Agner Krarup Erlang (1878), no início do século XX desenvolveu o primeiro estudo sobre filas, relacionado ao problema das redes de telefones, quando as ligações para fora da vila em estudo tinham que esperar porque todas as linhas estavam ocupadas. Com a análise dos problemas de tráfego de ligações e utilizando a teoria de probabilidades nesse estudo, Agner prova o Processo de Poisson para a distribuição aleatória dos tempos entre as chegadas de unidades ao sistema.

Em meados dos anos cinquenta os objetivos eram determinar o tamanho máximo admissível para a fila, a disciplina e o tempo de espera adequados. Para isso, o sistema seguia pressupostos Markovianos, cujo modelo admitia entradas independentes e aleatórias, tempo de serviço exponencial e chegadas de novos clientes baseadas no Processo de Poisson,

ou seja, os intervalos entre estas são exponencialmente distribuídos. Os trabalhos de Kendall, Smith e Cox tiveram grande destaque sobre a teoria das filas, e outros textos publicados de autores como Saaty, Benes, Cohen, Gnedenko e Kovalenko apresentaram conhecimentos teóricos e aplicações importantes para o estudo das filas naquela época.

Nos anos sessenta, alguns estudiosos tinham alcançado resultados impressionantes e soluções para vários modelos de filas, tais como aqueles em que as entradas são independentes e com distribuição de Poisson, único servidor e tempo de serviço exponencial. Até mesmo filas com prioridade foram bastante estudadas. Nesse momento, surgiu a ideia de que o estudo sobre a teoria de filas estava completo, mas foi precipitada, pois ainda hoje existem muitos matemáticos e estatísticos explorando e trabalhando com novos modelos de filas e aplicações.

Conforme Andrade (1998), um sistema de filas é caracterizado por seis componentes: modelo de chegada e serviço, número de servidores, capacidade do sistema, tamanho da população e a disciplina da fila, sendo os três primeiros obrigatórios. Estes compõem a Notação de Kendall, que classifica o sistema das filas por meio de símbolos. Para cada modelo serão calculadas algumas medidas de desempenho, tais como: o número médio de clientes na fila e no sistema, o tempo médio na fila e no sistema, a proporção de tempo que n clientes permanecem no sistema e quando o sistema está ocioso. Essas medidas permitem que encontremos um ponto de equilíbrio que satisfaça o cliente e o servidor.

Este trabalho está dividido da seguinte maneira: Na Seção 2, apresentamos a Teoria das Filas e suas características; Na seção 3, discutimos os Sistemas de Filas; A Seção 4 trata das aplicações; Na Seção 5, fazemos as simulações para o modelo $M/M/1$.

2 RESUMO

3 CONCEITOS FUNDAMENTAIS

A seguir, apresentamos as definições básicas da Teoria das filas, os processos estocásticos que a englobam, os tipos de modelo de filas, componentes de uma fila, distribuições das taxas e o teste de aderência para a Distribuição Exponencial.

3.1 Teoria das Filas

As filas de espera são encontradas em diversas situações do nosso cotidiano, são desagradáveis e responsáveis pela perda de tempo. Elas podem causar transtornos pessoais, pagamento de multas, prejuízos nos negócios futuros e na produtividade.

A Teoria das filas é uma ferramenta matemática que trata de eventos aleatórios e permite retratar previamente o comportamento de um sistema de filas, matematicamente. Podemos encontrar soluções que equilibrem tanto o congestionamento de clientes tanto a taxa de atendentes ociosos, e através desse estudo propor sistemas que sejam eficazes economicamente para provedor de serviço e que atendam a demanda.

Em aplicações, o estudo da Teoria das Filas tem como objetivo propor melhorias para o desempenho do sistema, ou seja, melhor utilização dos recursos de serviços considerando o ponto de vista do cliente e do servidor. As medidas de desempenho mais importantes de um sistema, como: número médio de clientes no sistema e na fila, comprimento médio da fila, tempo médio de espera no sistema e na fila e a taxa média de ocupação do serviço, auxiliam na tomada de decisão para minimizar o tamanho das filas e a permanência de clientes no sistema.

O processo de geração de filas é governado por leis probabilísticas, leis que englobam a Teoria de Processos Estocásticos. As filas em que a taxa de chegadas de clientes e o número de clientes atendidos por unidade de tempo obedecem um processo de Poisson são exemplos especiais de Cadeias de Markov, onde as transições de estado são processos contínuos em relação ao tempo de permanência. A cadeia de Markov de tempo contínuo aplicado aos modelos de filas nesse trabalho é chamada Processo de Nascimento e Morte, onde uma chegada representada um nascimento e uma saída representa uma morte.

Nas Seções seguintes, descrevemos as notações utilizadas nesse trabalho e as vertentes de Processos Estocásticos associados ao Sistema de Filas.

3.2 Notações

Em todas as filas de espera existem características muito úteis, indicando a sua notação habitual. A letra L , que significa length, será usada em questões relacionadas com o comprimento da fila e a letra W (waiting) em relação ao tempo de espera. As letras Q e S indicam respectivamente a fila (queue) e o sistema (system). A seguir são descritas as nomenclaturas e medidas de desempenho utilizadas nesse trabalho

1. **Taxa de chegada** (λ): número médio de clientes que chegam por unidade de tempo. Quando a taxa de chegadas depende do número de clientes n denotamos por λ_n ;
2. Intervalo médio entre duas chegadas consecutivas $\left(\frac{1}{\lambda}\right)$;
3. **Taxa de atendimento** (μ): número médio de clientes que cada servidor atende por unidade de tempo. Quando a taxa de serviço depende do número de clientes, denotamos por μ_n ;
4. Tempo médio de atendimento $\left(\frac{1}{\mu}\right)$;
5. Número médio de clientes no sistema (L) ou (L_S)= Fila + Atendimento;
6. Comprimento médio da fila (L_Q)= Em espera por atendimento;
7. Tempo médio de espera no sistema (W) ou (W_S);
8. Tempo médio de espera na fila (W_Q) (este tempo exclui o tempo que o cliente leva a ser atendido);
9. Taxa média de ocupação do serviço (ρ);
10. Proporção do tempo em que há n clientes no sistema (P_n);
11. S = Tempo de atendimento (serviço) de um cliente;
12. $\lambda E(S)$ = Número médio de clientes em atendimento;
13. Taxa média de entrada no sistema quando a taxa de chegada depende do estado, e o sistema está em equilíbrio $\left(\lambda' = \sum_{n=0}^k \lambda_n P_n\right)$;

14. Quando um processo está no estado estacionário, são aplicáveis as fórmulas de Little, apresentadas por este nos anos 60, que afirmam que

- (a) $L = \lambda W$, onde $L = \sum_{n=0}^{\infty} nP_n$ (média dos estados possíveis do sistema ponderados pelas suas probabilidades de ocorrência);
- (b) $L_Q = \lambda W_Q$, onde $L = \sum_{n=c+1}^{\infty} (n - c)P_n$;

Quando λ depender do estado do sistema usamos a expressão de λ' ao invés de λ , podendo aplicar normalmente as fórmulas de Little.

3.3 Processos Estocásticos

Um processo estocástico é um conjunto de variáveis aleatórias X_t ou $X(t)$, onde t é um parâmetro que percorre um conjunto de índices T . Geralmente, toma-se para conjunto T um dos conjuntos: \mathbb{N}_0 ou $[0, \infty)$, e neste último caso, a situação mais comum será que t represente o tempo. Os processos estocásticos são distinguidos pelo seu espaço de estados, ou seja, pelo conjunto dos valores que as v.a's envolvidas X_t podem tomar, pelo seu conjunto de índices T e pelas relações de dependência entre as v.a's $X_t, t \in T$. Dependendo da natureza da faixa de tempo, o processo é classificado como um processo de parâmetro contínuo ou de parâmetro discreto como discutimos a seguir.

- (i) Se T é um conjunto enumerável, por exemplo, $T = \{\dots, -2, -1, 0, 1, 2, \dots\}$ ou $T = \{0, 1, 2, \dots\}$, então o processo estocástico $\{X_t, t \in T\}$ é dito ser um processo de parâmetro discreto definido no conjunto de índices de T .
- (ii) Se T é um intervalo ou uma combinação de intervalos reais, por exemplo, $T = \{t : -\infty < t < +\infty\}$ ou $T = \{t : 0 < t < +\infty\}$, então o processo estocástico $\{X_t, t \in T\}$ é chamado de processo de parâmetro contínuo definido no conjunto de índices de T .

Um processo estocástico é a abstração matemática de um processo empírico cujo desenvolvimento é governado por leis probabilísticas, um exemplo são os Processos de Markov.

3.3.1 Processos de Markov

Um processo de Markov $\{X_t\}, t \in T$ é um processo estocástico onde a probabilidade de qualquer comportamento futuro do processo, quando o estado atual é conhecido, não é alterado pelo conhecimento adicional relativo ao comportamento passado. Isto é,

$$P[X_{t_{k+1}} = x_{k+1} | X_{t_k} = x_k, \dots, X_{t_1} = x_1] = P[X_{t_{k+1}} = x_{k+1} | X_{t_k} = x_k],$$
 para os instantes arbitrários $t_1 < \dots < t_k < t_{k+1}$ no caso discreto;

$$P[a \leq X_{t_{k+1}} \leq b | X_{t_k} = x_k, \dots, X_{t_1} = x_1] = P[a \leq X_{t_{k+1}} \leq b | X_{t_k} = x_k]$$
 para os instantes arbitrários $t_1 < \dots < t_k < t_{k+1}$ no caso contínuo.

Pode-se dizer que, dado a condição “presente” do processo, o “futuro” é independente do “passado”, ou seja, o processo está sem memória.

Um número real x é dito ser estado do processo estocástico $\{X_t, t \in T\}$ se existe um instante $t, t \in T$, tal que a $P(x - h < X(t) < x + h)$ é positivo para todo $h > 0$. O conjunto de estados possíveis constitui o espaço de estado do processo. Se o espaço de estado é discreto, o processo de Markov é chamado de cadeia de Markov, e se os processos de Markov têm parâmetro contínuo com espaço de estado discreto são conhecidos como cadeias de Markov a tempo contínuo. Quando um processo de Markov tem um espaço de estados contínuo e um espaço de parâmetro discreto, ele é conhecido como processo de Markov de parâmetro discreto. Por outro lado, se ambos os espaço de estados e espaço de parâmetros forem contínuos então o processo de Markov será chamado de processo de Markov de parâmetro contínuo.

Utilizamos os Processos de Markov de parâmetro contínuo para descrever os Processos de Nascimento e Morte, na próxima seção.

3.3.2 Processos de Nascimento e Morte

O Processo de Nascimento e Morte é uma cadeia de Markov de tempo contínuo, onde, uma cadeia com espaço de estados $S = \{0, 1, \dots\}$ ou $S = \{0, 1, \dots, s\}$, estando no estado n , no próximo passo somente poderá estar em $n, n + 1$ ou $n - 1$.

Segundo Marins (2009), os Sistemas das Filas podem ser identificados como Processos de Nascimento e Morte e apresenta as seguintes características

(a) Um Nascimento representa a entrada de um cliente no sistema;

- (b) Uma Morte representa a saída de um cliente no sistema;
- (c) $\lambda_n =$ Taxa Média de Chegada quando há n clientes no sistema;
- (d) $\mu_n =$ Taxa Média de Serviço quando há n clientes no sistema.

Através da representação de transições de estados, obtemos equações de balançamento que comprovam que o fluxo de entrada em um estado é igual ao seu fluxo de saída. As taxas de nascimento e morte, no estado $i \in S$, são dadas por λ_i e μ_i , respectivamente. A matriz geradora ou matriz de intensidade Q é constituída pela taxa na qual o processo move do estado i para o estado j , (q_{ij}) , ou seja,

$$Q = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & \dots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & 0 & \dots \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}.$$

O tempo entre chegadas e o tempo de atendimento são supostos exponencialmente distribuídos com taxas λ e μ . Então, o tempo até que um cliente entre ou saia do sistema têm distribuição Exponencial com taxa $\lambda + \mu$. Quando $j = i$ temos que $q_{ii} = -\sum_{j \neq i} q_{ij}$. Resumidamente,

$$q_{ij} = \begin{cases} \lambda_i, & \text{se } j = i + 1; \\ \mu_i, & \text{se } j = i - 1. \quad (\mu_i \neq 0); \\ -(\lambda_i + \mu_i), & \text{se } j = i. \quad (q_0 = \lambda_0); \\ 0, & \text{se } j = i + 2 \quad \text{ou} \quad j = i - 2. \end{cases}$$

Além disso, pode-se observar que a razão $\rho = \frac{\lambda}{\mu}$ (taxa de ocupação do serviço) é tal que

- (i) Se $\frac{\lambda}{\mu} > 1$, há mais chegadas do que saídas de clientes. O número de clientes no sistema é ilimitado, o sistema é instável;
- (ii) Se $\frac{\lambda}{\mu} < 1$, há mais saídas do que chegadas de clientes. Então, existe uma solução estacionária para o sistema;

- (iii) Se $\frac{\lambda}{\mu} = 1$, significa que chega, em média, o mesmo número de clientes que saem do sistema. Qualquer número de clientes no sistema é equiprovável e o sistema é instável.

A Teoria das Filas tem interesse sobre quanto tempo o sistema fica em um dado espaço, e pra onde ele se moverá, dado o estado atual. Por isto, as transições probabilísticas de estados podem ser descritas como uma Cadeia de Markov de tempo contínuo, na qual o processo das filas tem probabilidade de movimento, somente, para o estado $n + 1$ ou $n - 1$, estando no estado n . A matriz de transição P é definida como

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & \dots \\ \frac{\mu_1}{(\lambda_1 + \mu_1)} & 0 & \frac{\lambda_1}{(\lambda_1 + \mu_1)} & 0 & 0 & \dots \\ 0 & \frac{\mu_2}{(\lambda_2 + \mu_2)} & 0 & \frac{\lambda_2}{(\lambda_2 + \mu_2)} & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}.$$

Assim,

$$p_{ij} = \begin{cases} \frac{\lambda_i}{\lambda_i + \mu_i}, & \text{se } (j = i + 1, i \geq 1); \\ \frac{\mu_i}{\lambda_i + \mu_i}, & \text{se } (j = i - 1, i \geq 1); \\ 1, & \text{se } (i = 0, j = 1); \\ 0, & \text{caso contrário.} \end{cases}$$

Um padrão de chegada ou atendimento que não muda com o tempo (a distribuição de probabilidade é independente do tempo) apresenta a propriedade de estacionaridade. Os modelos são ditos não estacionários quando

- (i) O número médio de entradas por unidade de tempo é diferente do número médio de saídas, em cada estado;
- (ii) A fila tende para infinito porque a taxa de chegada excede a taxa de atendimento;
- (iii) A taxa de chegada varia ao longo do tempo.

As equações de balanceamento para um Processo de Nascimento e Morte com as condições apropriadas em μ_i e λ_i , podem ser obtidas através da equação estacionária do

sistema, ou seja, $PQ = 0$. Com esse resultado, podemos encontrar as equações de balanço do processo para deduzimos os valores de P_n e P_0 . Dessa forma, temos

$$\lambda_{j-1}P_{j-1} + \mu_{j+1}P_{j+1} - (\lambda_j + \mu_j)P_j = 0, \quad \text{para } j \geq 1. \quad (1)$$

Para $j = 0$ em (1), temos

$$\begin{aligned} \mu_1 P_1 - \lambda_0 P_0 &= 0; \\ P_1 &= \frac{\lambda_0}{\mu_1} P_0. \end{aligned}$$

Isolando P_{j+1} na equação (1), obtemos

$$P_{j+1} = \frac{(\lambda_1 + \mu_1)}{\mu_{j+1}} P_j - \frac{\lambda_{j-1}}{\mu_{j+1}} P_{j-1}. \quad (2)$$

Agora, vamos fazer $j = \{0, 1, 2, \dots, n-1\}$ na equação (2) e substituir os valores anteriores de P_{j+1} já definidos com o objetivo de escrever P_{j+1} em função de P_0 . Daí,

$$\begin{aligned} j = 0 &\Rightarrow P_1 = \frac{\lambda_0}{\mu_1} P_0; \\ j = 1 &\Rightarrow P_2 = \frac{(\lambda_1 + \mu_1)}{\mu_2} P_1 - \frac{\lambda_0}{\mu_2} P_0 = \frac{(\lambda_1 + \mu_1)}{\mu_2} \frac{\lambda_0}{\mu_1} P_0 - \frac{\lambda_0}{\mu_2} P_0 = \frac{\lambda_1 \lambda_0 P_0}{\mu_1 \mu_2}; \\ j = 2 &\Rightarrow P_3 = \frac{(\lambda_2 + \mu_2)}{\mu_3} P_2 - \frac{\lambda_1}{\mu_3} P_1 = \frac{(\lambda_2 + \mu_2)}{\mu_3} \frac{\lambda_1 \lambda_0 P_0}{\mu_1 \mu_2} - \frac{\lambda_1}{\mu_3} \frac{\lambda_0}{\mu_1} P_0 = \frac{\lambda_2 \lambda_1 \lambda_0 P_0}{\mu_3 \mu_2 \mu_1}; \\ &\vdots \\ j = (n-1) &\Rightarrow P_n = \frac{\lambda_{n-1} \lambda_{n-2} \dots \lambda_0}{\mu_n \mu_{n-1} \dots \mu_1} P_0. \end{aligned}$$

Logo, para $j = (n-1)$ encontramos a expressão de P_n em função de P_0 , que pode ser reescrita como

$$P_n = P_0 \prod_{i=1}^n \left(\frac{\lambda_{i-1}}{\mu_i} \right), \quad n \geq 1. \quad (3)$$

Sabemos que $\sum_{n=0}^{\infty} P_n = 1$. Então

$$P_0 + \sum_{n=1}^{\infty} P_n = 1. \quad (4)$$

Substituindo o resultado de P_n encontrado em (3) na equação (4), podemos encontrar o valor de P_0 , ou seja

$$\begin{aligned} 1 &= P_0 + \sum_{n=1}^{\infty} \left[P_0 \prod_{i=1}^n \left(\frac{\lambda_{i-1}}{\mu_i} \right) \right]; \\ 1 &= P_0 + P_0 \sum_{n=1}^{\infty} \left[\prod_{i=1}^n \left(\frac{\lambda_{i-1}}{\mu_i} \right) \right]; \end{aligned}$$

Isolando P_0 , obtemos

$$P_0 = \left(1 + \sum_{n=1}^{\infty} \left[\prod_{i=1}^n \left(\frac{\lambda_{i-1}}{\mu_i} \right) \right] \right)^{-1}. \quad (5)$$

Temos que P_0 por (5), é a proporção do tempo em que o sistema fica vazio, ou seja, os atendentes ficam ociosos. Pela expressão em (3) é possível calcular os valores de probabilidade de ter n clientes no sistema. Essas equações de P_n e P_0 , obtidas através desse Processo de Nascimento e Morte são muito úteis para gerar uma variedade de Sistema de Filas. Por exemplo, adotando-se $\lambda_n = \lambda$ e $\mu_n = \mu$ obtém-se uma fila $M/M/1$, descrita na Seção (4.1).

3.4 Processos de Poisson

3.4.1 Processos de Poisson Homogêneo

Definição 3.1 *Um processo estocástico $\{X_t\}, t \in T$ é dito ser um processo de contagem se X_t representa o número total de chegadas ocorridos até o tempo t . Um processo de Poisson Homogêneo com taxa (número de clientes que chegam ou são atendidos por unidade de tempo) $\lambda > 0$, deve satisfazer as seguintes condições*

1. $X_0 = 0$ e $X_t > 0$;
2. Se $s < t$, então $X_s < X_t$;

3. $X_t - X_s$ é o número de chegadas que ocorreram no intervalo $(s, t]$.

Para qualquer coleção de instantes $t_0 < t_1 < \dots < t_n$, os incrementos dos processos são variáveis aleatórias independentes do tempo. Em outras palavras, para um tempo arbitrário $t \geq 0$, o número de chegadas no período de tempo $(t, t + s)$ é independente do número de chegadas até o tempo t .

Pela propriedade, a distribuição do número de chegadas não muda com tempo. É dita ser Poisson com parâmetro λt em t unidades de tempo, ou seja,

$$P(X_{s+t} - X_s = n) = \frac{e^{-\lambda t} (\lambda t)^n}{n!}.$$

Pode-se mostrar que a média e a variância são iguais a

$$E(X_t) = Var(X_t) = \lambda t.$$

3.4.2 Processos de Poisson Não Homogêneo

Definição 3.2 No Processo de Poisson Não Homogêneo a taxa não é constante, mas uma função dependente do tempo λ_t , diferentemente do Processo de Poisson Homogêneo quando $\lambda(t) = \lambda$ para $\forall t \geq 0$. Por definição, temos que

$$X_t \sim \text{Poisson} \left(\int_0^t \lambda(u) du \right), \quad \text{onde } \lambda(t) = \lim_{\varepsilon \downarrow 0} \frac{P(x_{t+\varepsilon} - X_t = 1)}{\varepsilon}.$$

De forma geral, $X_t - X_s \sim \text{Poisson} \left(\int_s^t \lambda(u) du \right)$ para $0 < s < t$.

Como $X_t = n$, os tempos T_1, \dots, T_n das ocorrências de interesse se distribuem como as estatísticas de ordem de n variáveis aleatórias: Y_1, \dots, Y_n independentes e identicamente distribuídas com Função de densidade de probabilidade dada por

$$f(u) = \frac{\lambda(u)}{\int_0^t \lambda(v) dv}, \quad 0 \leq u \leq t.$$

Observação 3.1 *Dado que $T_n = s$, temos*

$$P(T_{n+1} - T_n > \mu) = \exp\left(-\int_s^\mu \lambda(v)dv\right).$$

Geralmente, um Processo de Poisson Não Homogêneo é o que melhor caracteriza as chegadas num sistema de filas. De fato, os clientes vão mais ao banco no começo do mês, o supermercado fica mais cheio no dia de frutas frescas ou o tráfego de carros é maior no horário da manhã e da tarde. As análises desses tipos de modelos se dá somente por simulações e utilizando softwares. Por outro lado, se analisarmos em um período pequeno de tempo, temos que as taxas de chegadas e de atendimento são homogêneas e podemos utilizar as fórmulas fechadas definidas para o Processo de Poisson Homogêneo. E assim, faremos ao longo do nosso trabalho.

3.5 Distribuição Exponencial

Os tempos entre as chegadas e os tempos de atendimento são independentes e identicamente distribuídos. Podem pertencer às distribuições: Geral (G), Erlang (E_r) e Exponencial (M de Memoryless), sendo essa última, a distribuição utilizada para os Sistemas de Filas descritos e aplicados nesse trabalho.

Definição 3.3 *Uma variável aleatória X tem distribuição Exponencial com parâmetro $\lambda > 0$, se tem função densidade de probabilidade definida por*

$$f(t) = \begin{cases} \lambda e^{-\lambda t}, & \text{se } t > 0; \\ 0, & \text{se } t \leq 0. \end{cases}$$

A distribuição acumulada da variável $X \sim \exp(\lambda)$ é dada por

$$F(t) = P(T \leq t) = \begin{cases} 1 - e^{-\lambda t}, & \text{se } t > 0; \\ 0, & \text{se } t \leq 0. \end{cases}$$

Pode-se mostrar que a esperança e a variância são, respectivamente

$$E(X) = \frac{1}{\lambda} \quad \text{e} \quad \text{Var}(X) = \frac{1}{\lambda^2}.$$

A distribuição exponencial satisfaz uma propriedade importante, chamada perda de memória, pois

$$P(X > t + s | X > t) = P(X > s), \quad t, s > 0.$$

3.6 Teste Kolmogorov-Smirnov de Aderência

Durante este trabalho, para verificar se os dados possuem distribuição exponencial para os tempos entre as chegadas e o tempos de serviço realizamos o teste Kolmogorov-Smirnov de aderência que assume as seguintes hipóteses

$$\begin{cases} H_0 : \text{A amostra provém de uma Distribuição Exponencial;} \\ H_A : \text{A amostra não segue a Distribuição Exponencial.} \end{cases}$$

Dada uma amostra aleatória X_1, X_2, \dots, X_t referente a uma variável aleatória X , a estatística do teste é a diferença máxima absoluta entre as funções de distribuição acumulada teórica e empírica, ou seja,

$$D = \max_x |S_t(x) - F_0(x)|.$$

Se $X \sim \exp(\lambda)$, temos que

1. $F_0(x) = \int_0^z e^{-\lambda u} du = 1 - e^{-\lambda z}, (z \leq t);$
2. $S_n(x) = \frac{N(x)}{n}$, onde $N(x)$ é o número de observações $\leq x_n$.

Se a estatística D for maior que o D_{critico} , obtido por meio da tabela “Quantis do teste Kolmogorov-Smirnov” rejeitamos a hipótese nula. Se for menor, assumimos que a amostra provém de uma distribuição exponencial.

3.7 Sistemas de Filas

Os sistemas de filas se descrevem, de forma geral, por um processo de chegada de clientes a um sistema de atendimento para receber um ou mais serviços, executados por certa quantidade de servidores. As formações de filas ocorrem porque a procura pelo serviço é maior do que a capacidade de atendimento do sistema.

Os tipos de modelos de filas é definido a partir da Notação de Kendall, que representa cada cadeia de filas pelos símbolos A, S, m, k, N, Q da seguinte forma

$$A/S/m/k/N/Q.$$

Onde,

$A \rightarrow$ Distribuição dos tempos entre as chegadas;

$S \rightarrow$ Distribuição dos tempos de serviços;

$m \rightarrow$ Número de Servidores ($m \in \mathbb{N}$);

$k \rightarrow$ Capacidade do Sistema;

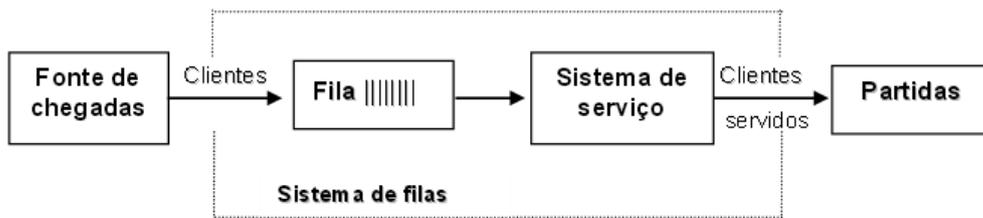
$N \rightarrow$ Tamanho da população;

$Q \rightarrow$ Disciplina da Fila.

Quando os três últimos símbolos são omitidos, como por exemplo no modelo $M/M/1$, assume-se capacidade ilimitada, população infinita e disciplina de atendimento FCFS ou FIFO. As letras que definem as distribuições do tempo entre chegadas de clientes ou do serviço são: M (de Memoryless) para a distribuição exponencial, E_r (de Erlang) para a distribuição de Erlang e G (de Genérica) para uma distribuição genérica.

Um sistema de filas é constituído por seis componentes. A figura a seguir mostra a estrutura de um sistema baseada nos seus elementos.

Figura 1 – Estrutura de um sistema de filas



Agora definiremos cada um dos seis componentes,

1. Modelo de chegada

A distribuição das chegadas apresenta um comportamento estocástico e pode ser descrita pelo tempo entre duas chegadas ou pelo número médio de chegadas por unidade de tempo (taxa). Podem ser constantes quando existem intervalos de tempo fixos entre chegadas sucessivas ou aleatórias quando os intervalos de tempo entre chegadas não podem ser previstos, usando-se neste caso distribuições de probabilidade. Também é necessário conhecer se os clientes chegam sozinhos ou em grupo, se o cliente muda de fila, desiste ou sempre decide ficar no sistema e se o padrão das chegadas é estacionário ou não-estacionário.

2. Padrão de Serviço

A distribuição dos tempos de serviço também são variáveis aleatórias independentes e identicamente distribuídas. A taxa de serviço corresponde ao número médio de clientes que são atendidos por cada servidor e por unidade de tempo. O processo de serviço pode ser simples ou batch e depender do número de clientes esperando pelo serviço. A situação na qual o serviço depende do número de clientes na fila é conhecida como serviço dependente do estado. Serviços, como chegadas, podem ser estacionários ou não estacionários com respeito ao tempo.

3. Número de Servidores

É o número de posições de atendimento à disposição no sistema. Podem ser idênticos ou distintos e a fila pode ser única por servidor ou por grupo de servidores (vários clientes podem ser atendidos ao mesmo tempo no mesmo servidor). Ou seja, sistema com canal simples ou com multicanais.

4. Capacidade do Sistema

É o número máximo de clientes que podem permanecer no sistema, inclui clientes em serviço ou à espera de atendimento com capacidade infinita ou finita, ou seja, o cliente só entra quando o outro sai, quando o sistema está cheio. A capacidade é definida por restrições econômicas, espaço ou tempo de espera.

5. Tamanho da População

É o número de clientes finito ou infinito que podem chegar a um sistema de filas. Quando for infinita a probabilidade de ocorrer uma nova chegada não é influenciada pelo número de clientes que já se encontram no sistema.

6. Disciplina da fila

Descreve a ordem do qual os clientes são atendidos. Os tipos mais usuais são

- (a) FCFS (First Come, First Served) ou FIFO (First In, First Out): As filas com características FCFS (ou FIFO) são as filas onde o primeiro cliente que chega é o primeiro a ser atendido e a sair. Estas são as filas mais comuns no cotidiano.
- (b) LCFS (Last Come, First Served) ou LIFO (Last In, First Out): São as filas onde o último cliente a chegar é o primeiro a ser atendido e a sair.
- (c) SIRO (Service In Random Order): São filas em que o serviço é feito de forma aleatória.
- (d) PRI (Prioritárias): As filas PRI são as filas com prioridade, onde é atribuída uma prioridade a cada cliente, podendo um cliente que entra com maior prioridade ser atendido imediatamente, interrompendo o cliente que estava sendo atendido ou o cliente com maior prioridade ser colocado no início da fila, e sendo o próximo a ser atendido após a saída do cliente que estava recebendo o atendimento.
- (e) RR (Round-Robin): Nas filas que seguem a regra de Round-Robin cada cliente recebe uma fatia de tempo do serviço, dentro da qual é atendido. Após terminar esse tempo, mesmo que a atividade não tenha sido completada, o cliente é retirado e outro passa a ser atendido. O cliente cujo serviço foi interrompido, retorna ao serviço, posteriormente.
- (f) GD (General Discipline): As filas GD seguem uma disciplina genérica, ou seja, nestas filas não é especificada a disciplina de atendimento.

4 FILAS MARKOVIANAS

As filas markovianas abrangem os principais sistemas de filas onde as chegadas em um intervalo de tempo $X_t \in (0, T]$, seguem um processo de Poisson com taxa λ , e os tempos de serviço seguem uma distribuição exponencial de parâmetro μ , ou seja,

$$P(X_t = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!};$$

Onde,

$$f(t) = \lambda e^{-\lambda t} \text{ (distribuição do tempo entre as chegadas);}$$

$$f(t) = \mu e^{-\mu t} \text{ (distribuição do tempo de atendimento).}$$

4.1 Sistemas $M/M/1$

Os sistemas $M/M/1$ corresponde ao modelo básico de filas markovianas, onde o sistema tem chegadas seguindo um processo de poisson, tempos de atendimento com distribuição exponencial, capacidade do sistema e população são infinitas. Além disso, esse modelo possui as seguintes características

1. Taxa média de chegada = λ ;
2. Taxa média de atendimento = μ ;
3. Número de atendentes = 1;
4. Disciplina de atendimento = FIFO ou FCFS.

As funções densidade para o tempo entre as chegadas e tempos de atendimento são dadas, respectivamente

$$f(t) = \lambda e^{-\lambda t},$$

$$f(t) = \mu e^{-\mu t}.$$

Onde $\frac{1}{\lambda}$ é o tempo médio entre chegadas e $\frac{1}{\mu}$ é o tempo médio de serviço, sendo os tempos independentes.

O modelo $M/M/1$ é um processo de nascimento e morte onde as taxas de chegada e serviço independem do estado do sistema, isto é, $\lambda_n = \lambda$ e $\mu_n = \mu$, para todo $n = \{1, 2, 3, \dots\}$. Assim, as chegadas podem ser consideradas como nascimentos para o sistema, desde que, se o sistema está no estado n e uma chegada ocorre, o estado do sistema é atualizado para $n + 1$. Por outro lado, a ocorrência de uma partida pode ser considerada como uma morte, de maneira que, quando este evento ocorre, o sistema será levado para o estado $n - 1$, estando ele no estado n . Note que, isto vale para todos os valores onde $n \geq 1$.

Utilizando-se as equações de balanceamento, temos que Taxa de Entrada = Taxa de Saída, ou seja, para que haja um estado de equilíbrio é necessário que o número de entradas e de saídas por unidade de tempo sejam iguais. Através destas, podemos calcular P_{i+1} de cada estado i (número de clientes no sistema) em função de P_0 . Os passos são descritos na seguinte tabela

Tabela 1 – Dedução de P_n para o modelo $M/M/1$

Estado	Taxa de entrada	=	Taxa de Saída	P_{i+1}
0	μP_1	=	λP_0	$P_1 = \frac{\lambda}{\mu} P_0$
1	$\lambda P_0 + \mu P_2$	=	$\lambda P_1 + \mu P_1$	$P_2 = \frac{1}{\mu}(\lambda + \mu)P_1 - \lambda P_0 \Rightarrow P_2 = \frac{\lambda^2}{\mu^2} P_0$
2	$\lambda P_1 + \mu P_3$	=	$\lambda P_2 + \mu P_2$	$P_3 = \frac{\lambda^3}{\mu^3} P_0$
⋮				
$n - 1$	$\lambda P_{n-2} + \mu P_n$	=	$\lambda P_{n-1} + \mu P_{n-1}$	$P_n = \left(\frac{\lambda}{\mu}\right)^n P_0$
n	$\lambda P_{n-1} + \mu P_{n+1}$	=	$\lambda P_n + \mu P_n$	$P_{n+1} = \frac{1}{\mu}(\lambda + \mu)P_n - \frac{\lambda}{\mu}P_{n-1} \Rightarrow P_{n+1} = \frac{\lambda}{\mu}P_n$

De acordo com o padrão para cada estado i , deduzimos que

$$P_{n+1} = \left(\frac{\lambda}{\mu}\right)^{n+1} P_0, \quad \text{para } \forall n \geq 0. \quad (6)$$

Por definição temos que $\sum_{n=0}^{\infty} P_n = 1$. Com esse resultado podemos encontrar o valor de P_0 , ou seja,

$$\begin{aligned} 1 &= \sum_{n=0}^{\infty} P_n = \sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^n P_0 = P_0 \frac{1}{1 - \lambda/\mu} \\ \Rightarrow P_0 &= \frac{\mu - \lambda}{\mu} = 1 - \frac{\lambda}{\mu}. \end{aligned} \quad (7)$$

Substituindo o resultado (7) na equação (6), obtemos

$$P_n = \left(\frac{\lambda}{\mu}\right)^n \left(1 - \frac{\lambda}{\mu}\right), \quad n = 0, 1, 2, \dots$$

Como dito anteriormente, esse modelo é dito ser um processo de nascimento e morte. Logo, podemos encontrar esses resultados a partir das fórmulas (5) e (3). Onde

$$P_0 = \left(1 + \sum_{n=1}^{\infty} \left[\prod_{i=1}^n \left(\frac{\lambda_{i-1}}{\mu_i}\right) \right]\right)^{-1} = \frac{1}{\sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^n} = \frac{1}{\frac{1}{1-\lambda/\mu}} = 1 - \frac{\lambda}{\mu}, \quad (8)$$

$$P_n = P_0 \prod_{i=1}^n \left(\frac{\lambda_{i-1}}{\mu_i}\right) = P_0 \left(\frac{\lambda}{\mu}\right)^n = \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^n. \quad (9)$$

Se fizermos $\rho = \lambda/\mu$ em (9) temos que $P_n = \rho^n(1 - \rho)$, ou seja, se N é o número de clientes no sistema, então $N \sim \text{Geométrica}(\rho)$.

Seja $E(S)$ o tempo médio de atendimento, as medidas de desempenho são baseadas nas fórmulas de Little e são dadas por

1. $L = E(N) = \frac{\rho}{1 - \rho} = \frac{\lambda/\mu}{1 - \lambda/\mu} = 1 - \frac{\lambda}{\mu}$ (número médio de clientes no sistema);
2. $W = \frac{L}{\lambda} = \frac{1}{\mu - \lambda}$ (tempo médio de cada unidade no sistema);
3. $W_Q = W - E(S) = \frac{1}{\mu - \lambda} - \frac{1}{\mu} = \frac{\lambda}{\mu(\mu - \lambda)}$ (tempo médio de cada unidade na fila);
4. $L_Q = \lambda W_Q \Rightarrow L_Q = \frac{\lambda^2}{\mu(\mu - \lambda)}$ (número médio de clientes na fila).

4.2 Sistemas $M/M/\infty$

A diferença entre os modelos $M/M/\infty$ e $M/M/1$ é que neste, o número de servidores disponíveis é infinito, ou seja, não existe fila. Para encontrarmos as expressões de P_n e P_0 usamos a mesma ideia do modelo anterior, logo

Tabela 2 – Dedução de P_n para o modelo $M/M/\infty$

Estado	Taxa de entrada	=	Taxa de Saída	P_{i+1}
0	μP_1	=	λP_0	$P_1 = \frac{\lambda}{\mu} P_0$
1	$\lambda P_0 + 2\mu P_2$	=	$\lambda P_1 + \mu P_1$	$(\lambda + \mu)P_1 = (\lambda + \mu)\frac{\lambda}{\mu}P_0 \Rightarrow P_2 = \frac{1}{2}\left(\frac{\lambda}{\mu}\right)^2 P_0$
2	$\lambda P_1 + 3\mu P_3$	=	$\lambda P_2 + 2\mu P_2$	$P_3 = \frac{1}{3\mu} \frac{\lambda^3}{2\mu^2} P_0 = \frac{1}{3!}\left(\frac{\lambda}{\mu}\right)^3 P_0$
⋮				
$n - 2$	$\lambda P_{n-3} + (n - 1)\mu P_{n-1}$	=	$\lambda P_{n-2} + (n - 2)\mu P_{n-2}$	$P_{n-1} = \frac{1}{(n-1)!} \left(\frac{\lambda}{\mu}\right)^{n-1} P_0$
$n - 1$	$\lambda P_{n-2} + n\mu P_n$	=	$\lambda P_{n-1} + (n - 1)\mu P_{n-1}$	$P_n = \frac{1}{n\mu} [(\lambda + (n - 1)\mu)] P_{n-1} - \frac{\lambda P_{n-2}}{n\mu}$

Expandindo a equação para $(n - 1)$ e substituindo os valores já encontrados, temos que

$$P_n = \frac{1}{n} \left(\frac{\lambda}{\mu}\right) P_{n-1} = \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n P_0 \quad (10)$$

Usando a restrição $\sum_{n=0}^{\infty} P_n = 1$ podemos encontrar P_0 , dada por

$$1 = \sum_{n=0}^{\infty} P_n = P_0 \sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^n = P_0 e^{\lambda/\mu} \Rightarrow P_0 = e^{-\lambda/\mu} \quad (11)$$

Substituindo o resultado (11) na equação (10), obtemos

$$P_n = \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n P_0 = \frac{e^{\lambda/\mu} (\lambda/\mu)^n}{n!}, \quad n = 0, 1, \dots \quad (12)$$

Com o uso das equações (5) e (3) de nascimento e morte geral com $\lambda_n = \lambda$ e $\mu_n = n\mu$ (taxa de saída quando se tem n clientes) encontramos os mesmos resultados obtidos em (12) e (11).

Observamos que P_n tem distribuição Poisson. Logo, se N é a variável aleatória que indica o número de clientes no sistema num instante qualquer, $N \sim \text{Poisson}(\lambda/\mu)$. Então

- $L = E(N) = \frac{\lambda}{\mu}$ (número médio de clientes no sistema);
- $L_Q = W_Q = 0$, pois nenhum cliente espera pelo atendimento;
- $L = \lambda W \Rightarrow W = \frac{L}{\lambda} = \frac{1}{\mu}$ (tempo de espera médio de cada unidade no sistema).

4.3 Sistemas $M/M/c$

Nos modelos $M/M/c$, diferente dos anteriores, o número de atendentes é restrito igual a c . Cada servidor possui distribuição de tempo de serviço exponencial independente e identicamente distribuída exponencialmente, com o processo de chegada de novo assumido ser Poisson.

Sob essas condições, tem-se novamente um processo de nascimento e morte, onde as razões de chegadas continuam não dependendo do estado do sistema, e as razões de saídas variam de acordo com o número de clientes. São definidas como

$$\begin{aligned} \lambda_n &= \lambda, \quad n \geq 0; \\ \mu_n &= \begin{cases} n\mu, & 0 \leq n < c; \\ c\mu, & n \geq c \end{cases} \quad (\text{Se todos os servidores estão ocupados}). \end{aligned} \quad (13)$$

Aplicando as restrições definidas em (13) na equação (3), obtemos

$$P_n = \begin{cases} \frac{\lambda^n}{n!\mu^n} P_0, & 1 \leq n < c, \\ \frac{\lambda^n}{c^{n-c}c!\mu^n}, & n \geq c. \end{cases} \quad (14)$$

Para encontrar P_0 , utilizamos a restrição de que as probabilidades devem somar um, e isto implica que

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

Utilizando expressões para soma de uma progressão geométrica, e fazendo $r = \lambda/\mu$ e $\rho = \lambda/c\mu$, podemos reescrever P_0 como

$$P_0 = \left(\sum_{n=0}^{c-1} \frac{r^n}{n!} + \frac{r^c}{c!(1-\rho)} \right)^{-1}, \quad \rho < 1. \quad (15)$$

Para um estado de equilíbrio, a taxa média de chegada deve ser menor que a taxa média máxima de serviço do sistema, que é intuitivamente esperado. Note que, quando $c = 1$ aplicados nas fórmulas de P_n e P_0 para esse modelo, reduzimos para as equações do sistema $M/M/1$. Baseando-se nas probabilidades de estado de equilíbrio dada nas equações

(14) e (15) podemos encontrar as medidas desempenho desse modelo de maneira similar as utilizadas nos sistemas anteriores. Logo,

- $L_Q = \sum_{n=c+1}^{\infty} (n-c)P_n = \sum_{n=c+1}^{\infty} (n-c) \frac{r^n}{c^{n-c}c!} P_0 = \frac{r^c}{c!} P_0 \sum_{n=c+1}^{\infty} (n-c) \frac{r^{n-c}}{c^{n-c}} = \frac{(r^c \rho)}{c!(1-\rho)^2} P_0;$
- $W_Q = \frac{L_Q}{\lambda} = \frac{r^c}{c!(c\mu)(1-\rho)^2} P_0;$
- $L = \lambda W = L_Q + r = \frac{(r^c \rho)}{c!(1-\rho)^2} P_0 + r;$
- $W = W_Q + \frac{1}{\mu} = \frac{1}{\mu} + \frac{r^c}{c!(c\mu)(1-\rho)^2} P_0.$

4.4 Sistemas $M/M/1/k$

Neste modelo, consideramos só um atendente e capacidade do sistema igual a k , ou seja, a probabilidade de estar num estado $k+1$ é nula pois o sistema só aceita até k clientes. A população é infinita e a disciplina da fila é a FIFO como as das anteriores.

Nos sistemas $M/M/1/k$, as taxas de chegada e saída dependem do estado n , ou seja,

$$\lambda_n = \begin{cases} \lambda, & n = 0, 1, 2, \dots, k-1; \\ 0, & n \geq k. \end{cases}$$

$$\mu_n = \begin{cases} \mu, & 0 \leq n = 0, 1, 2, \dots, k; \\ 0, & n > k. \end{cases}$$

O limite do sistema P_n é dado por

$$P_n = \begin{cases} \rho^n P_0, & \text{se } n = 0, 1, \dots, k; \\ 0, & \text{se } n > k. \end{cases} \quad (16)$$

onde $\rho = \frac{\lambda}{\mu}$ e $P_0 + P_1 + \dots + P_k = 1$. Utilizando essas relações podemos determinar P_0 . De fato,

Quando $\lambda = \mu$ temos $\rho = 1$. Nesse caso, $P_n = P_0$ então

$$\begin{aligned} 1 &= P_0 + P_1 + \dots + P_k = P_0 + \dots + 1^k P_0 \\ &= P_0(k+1) \\ P_0 &= \frac{1}{k+1}. \end{aligned}$$

Para $\rho \neq 1$ utilizamos a mesma relação, ou seja,

$$\begin{aligned} 1 &= P_0 + P_1 + \dots + P_k = P_0 + \dots + \rho^k P_0 \\ &= P_0 \sum_{n=0}^k \rho^n \\ P_0 &= \frac{1-\rho}{1-\rho^{k+1}}. \end{aligned}$$

Substituindo os resultados de P_0 encontrados em (16) concluímos que

$$P_n = \begin{cases} \rho^n \frac{1-\rho}{1-\rho^{k+1}}, & \text{se } n = 0, 1, \dots, k. \quad (\rho \neq 1); \\ \frac{1}{k+1}, & \text{se } n = 0, 1, \dots, k. \quad (\rho = 1); \\ 0, & \text{se } n > k. \end{cases}$$

Para calcular o número médio de clientes no sistema fazemos

Para $\rho \neq 1$,

$$\begin{aligned} L &= \sum_{n=0}^k n P_n = \sum_{n=1}^k n \rho^n \left(\frac{1-\rho}{1-\rho^{k+1}} \right) = \frac{1-\rho}{1-\rho^{k+1}} \rho \sum_{n=1}^k n \rho^{n-1} \\ &= \frac{1-\rho}{1-\rho^{k+1}} \rho \sum_{n=1}^k \frac{d\rho^n}{d\rho} = \frac{1-\rho}{1-\rho^{k+1}} \rho \frac{d}{d\rho} \sum_{n=1}^k \rho^n \\ &= \frac{1-\rho}{1-\rho^{k+1}} \rho \frac{d}{d\rho} \left(\frac{1-\rho^{k+1}}{1-\rho} \right) \\ &= \frac{1-\rho}{1-\rho^{k+1}} \rho \left(\frac{-(k+1)\rho^k(1-\rho) + (1-\rho^{k+1})}{(1-\rho)^2} \right) \\ &= \frac{\rho}{1-\rho} - \frac{(k+1)\rho^{k+1}}{1-\rho^{k+1}}. \end{aligned}$$

Para $\rho = 1$,

$$\begin{aligned} L &= \sum_{n=0}^k nP_n = \frac{1}{k+1}(1+2+3+\dots+k) \\ &= \frac{k(k+1)}{2(k+1)} = \frac{k}{2}. \end{aligned}$$

Agora vamos encontrar a expressão para o número médio de clientes na fila. Como λ depende do estado do sistema, devemos calcular λ' .

Para $\rho \neq 1$, temos

$$\begin{aligned} \lambda' &= \sum_{n=0}^{\infty} \lambda_n P_n = \lambda \sum_{n=0}^{k-1} P_n = \lambda(1 - P_k) \\ &= \lambda \left(1 - \rho^k \left(\frac{1 - \rho}{1 - \rho^{k+1}} \right) \right) \end{aligned} \quad (17)$$

$$\begin{aligned} &= \lambda \left(\frac{1 - \rho^{k+1}}{1 - \rho^{k+1}} - \frac{\rho^k - \rho^{k+1}}{1 - \rho^{k+1}} \right) \\ &= \lambda \left(\frac{1 - \rho^k}{1 - \rho^{k+1}} \right). \end{aligned} \quad (18)$$

Então

$$\begin{aligned} L_Q &= L - \frac{\lambda'}{\mu} = \frac{\rho}{1 - \rho} - \frac{(k+1)\rho^{k+1}}{1 - \rho^{k+1}} - \frac{\lambda \left(\frac{1 - \rho^k}{1 - \rho^{k+1}} \right)}{\mu} \\ &= \frac{\rho}{1 - \rho} - \frac{(k+1)\rho^{k+1}}{1 - \rho^{k+1}} - \rho \left(\frac{1 - \rho^k}{1 - \rho^{k+1}} \right) \\ &= \frac{\rho}{1 - \rho} + \frac{-k\rho^{k+1} - \rho^{k+1} - \rho + \rho^{k+1}}{1 - \rho^{k+1}} \\ &= \frac{\rho}{1 - \rho} - \frac{k\rho^{k+1} + \rho}{1 - \rho^{k+1}}. \end{aligned}$$

Para $\rho = 1$, temos

$$\lambda' = \lambda(1 - P_k) \lambda \left(1 - \frac{1}{k+1} \right) = \quad (19)$$

$$= \lambda \left(\frac{k}{k+1} \right). \quad (20)$$

Então

$$\begin{aligned}
L_Q &= L - \frac{\lambda'}{\mu} = \frac{k}{2} - \frac{\lambda \left(\frac{k}{k+1} \right)}{\mu} = \frac{k}{2} - \rho \left(\frac{k}{k+1} \right) \\
&= \frac{k(k+1) - 2k}{2(k+1)} = \frac{k^2 + k - 2k}{2(k+1)} \\
&= \frac{k(k-1)}{2(k+1)}.
\end{aligned}$$

O cálculo do tempo de espera se dá a partir da fórmula de Little. Com o resultado de λ' encontrado em (17), temos para $\rho \neq 1$,

$$\begin{aligned}
W_Q &= \frac{L_Q}{\lambda'} = \frac{\frac{\rho}{1-\rho} - \frac{k\rho^{k+1} + \rho}{1-\rho^{k+1}}}{\lambda \left(\frac{1-\rho^k}{1-\rho^{k+1}} \right)} \\
&= \frac{\frac{\rho - \rho^{k+2}}{(1-\rho)(1-\rho^{k+1})} - \frac{k\rho^{k+1} + \rho - k\rho^{k+2} - \rho^2}{1-\rho^{k+1}}}{\lambda \left(\frac{1-\rho^k}{1-\rho^{k+1}} \right)} \\
&= \frac{\rho - \rho^{k+2} - (k\rho^{k+1} + \rho - k\rho^{k+2} - \rho^2)}{\lambda(1-\rho)(1-\rho^k)} \\
&= \frac{k\rho^{k+1}(\rho - 1) + \rho^2(1 - \rho^k)}{\lambda(1-\rho)(1-\rho^k)} \\
&= \frac{\rho^2}{\lambda(1-\rho)} - \frac{k\rho^{k+1}}{\lambda(1-\rho^k)} \\
&= \frac{\rho}{\mu(1-\rho)} - \frac{k\rho^k}{\mu(1-\rho^k)}.
\end{aligned}$$

Com o resultado de λ' encontrado em (18) temos, para $\rho = 1$,

$$\begin{aligned}
W_Q &= \frac{L_Q}{\lambda'} = \frac{\frac{k(k-1)}{2(k+1)}}{\lambda \left(\frac{k}{k+1} \right)} = \frac{k(k-1)}{2\lambda k} \\
&= \frac{k-1}{2\rho\mu}.
\end{aligned}$$

Agora, vamos calcular no tempo médio de espera no sistema.

Para $\rho \neq 1$ e utilizando a expressão de λ' encontrada em (17) então

$$W = \frac{L}{\lambda'} = \frac{\frac{\rho}{1-\rho} - \frac{(k+1)\rho^{k+1}}{1-\rho^{k+1}}}{\lambda \left(\frac{1-\rho^k}{1-\rho^{k+1}} \right)}$$

$$\begin{aligned}
&= \frac{\frac{\rho - \rho^{k+2}}{(1-\rho)(1-\rho^{k+1})} - \frac{(k+1)\rho^{k+1} - (k+1)\rho^{k+2}}{(1-\rho)(1-\rho^{k+1})}}{\lambda \left(\frac{1-\rho^k}{1-\rho^{k+1}} \right)} \\
&= \frac{\rho - \rho^{k+2} - ((k+1)\rho^{k+1} - (k+1)\rho^{k+2})}{\lambda(1-\rho)(1-\rho^k)} \\
&= \frac{k\rho^{k+1}(\rho - 1) + \rho(1 - \rho^k)}{\lambda(1-\rho)(1-\rho^k)} \\
&= \frac{1}{\mu(1-\rho)} - \frac{k\rho^k}{\mu(1-\rho^k)}.
\end{aligned}$$

Para $\rho = 1$ e considerando a expressão de λ' encontrada em (18), segue que

$$\begin{aligned}
W &= \frac{L}{\lambda'} = \frac{k/2}{\lambda \left(\frac{k}{k+1} \right)} \\
&= \frac{k(k+1)}{2\lambda k} \\
&= \frac{k+1}{2\rho\mu}
\end{aligned}$$

4.5 Sistemas $M/M/c/k$

Os modelos $M/M/c/k$ são caracterizados por filas com canais paralelos e limitação na capacidade do sistema igual a c . Com abordagem análoga empregada no modelo $M/M/c$, exceto que a taxa de chegada que agora deve ser 0, $\forall n \geq k$. Considerando $c < k$ (pois quando $c \geq k$ não tem fila) e que ambas as taxas dependem do estado do sistema, temos

$$\lambda_n = \begin{cases} \lambda, & n = 0, 1, 2, \dots, k-1; \\ 0, & n \geq k. \end{cases}$$

$$\mu_n = \begin{cases} n\mu, & n = 0, 1, 2, \dots, c-1; \\ c\mu, & n = c, \dots, k. \end{cases}$$

Segue da equação (14) que P_n é dada por

$$P_n = \begin{cases} \frac{r^n}{n!} P_0, & 1 \leq n < c; \\ \frac{c^n \rho^n}{c!} P_0, & c \leq n \leq k. \end{cases} \quad (21)$$

No cálculo de P_0 temos que as séries são finitas e $\rho = \frac{\lambda}{c\mu}$ não necessariamente é menor do que 1. Logo,

$$P_0 = \left[\sum_{n=0}^{c-1} \frac{\lambda^n}{n!\mu^n} + \sum_{n=c}^k \frac{\lambda^n}{c^{n-c}c!\mu^n} \right]^{-1}. \quad (22)$$

Fazendo $r = \frac{\lambda}{\mu}$ e $\rho = \frac{r}{c}$ no somatório da segunda equação em (20) temos que

$$\begin{aligned} \sum_{n=c}^k \frac{r^n}{c^{n-c}c!} &= \left(\frac{r^c}{c!} \right) \sum_{n=c}^k \rho^{n-c} \\ &= \begin{cases} \frac{r^c}{c!} \frac{1 - \rho^{k-c+1}}{1 - \rho}, & \rho \neq 1, \\ \frac{r^c}{c!} (k - c + 1), & \rho = 1. \end{cases} \end{aligned} \quad (23)$$

Substituindo o resultado encontrado em (21) na equação (20) obtemos

$$P_0 = \begin{cases} \left(\sum_{n=0}^{c-1} \frac{r^n}{n!} + \frac{r^c}{c!} \left(\frac{1 - \rho^{k-c+1}}{1 - \rho} \right) \right)^{-1}, & \rho \neq 1, \\ \left(\sum_{n=0}^{c-1} \frac{r^n}{n!} + \frac{r^c}{c!} (k - c + 1) \right)^{-1}, & \rho = 1. \end{cases} \quad (24)$$

A taxa média de entradas no sistema, denotado por λ' é dada por

$$\lambda' = \sum_{n=0}^{k-1} \lambda P_n = \lambda(1 - P_k) = \lambda \left(1 - \frac{c^c \rho^k}{c!} P_0 \right)$$

Agora, adaptamos a derivação do número médio de clientes do modelo $M/M/c$ para um somatório finito, pois a capacidade do sistema é restrita. Ou seja,

$$L_Q = \sum_{n=c+1}^k (n - c) P_n$$

Para o caso $\rho \neq 1$, temos

$$L_Q = \sum_{n=c+1}^k (n - c) \frac{c^c \rho^n}{c!} P_0 = P_0 \frac{c^c}{c!} \rho^{c+1} \sum_{n=c+1}^k (n - c) \rho^{n-c-1}$$

$$\begin{aligned}
&= P_0 \frac{c^c}{c!} \sum_{j=1}^{k-c} j \rho^{j-1} = P_0 \frac{c^c}{c!} \rho^{j-1} = P_0 \frac{c^c}{c!} \rho^{s+1} \frac{\partial}{\partial \rho} \sum_{j=0}^{k-c} \rho^j = P_0 \frac{c^c}{c!} \rho^{c+1} \frac{\partial}{\partial \rho} \left(\frac{1 - \rho^{k-c+1}}{1 - \rho} \right) \\
&= P_0 \frac{c^c}{c!} \rho^{c+1} \left(\frac{1 - \rho^{k-c} - k\rho^{k-c} + k\rho^{k-c+1} + c\rho^{k-c} - c\rho^{k-c+1}}{(1 - \rho)^2} \right) \\
&= P_0 \frac{c^c}{c!} \rho^{c+1} \left(\frac{1 - \rho^{k-c}(1 + (1 - \rho))(k - c)}{(1 - \rho)^2} \right) \\
&= \frac{P_0 c^c \rho^{c+1}}{c!(1 - \rho)^2} [1 - \rho^{k-c}(1 + (1 - \rho)(k - c))].
\end{aligned}$$

Quando $\rho = 1$, segue que

$$\begin{aligned}
L_Q &= \sum_{n=c+1}^k (n - c) \frac{c^c}{c!} P_0 = \frac{c^c}{c!} P_0 \sum_{n=c+1}^k (n - c) = \frac{c^c}{c!} P_0 \sum_{j=1}^{k-c} j \\
&= \frac{c^c}{c!} P_0 \frac{(k - c)(k - c + 1)}{2}.
\end{aligned}$$

O cálculo do número médio de clientes no sistema se dá a partir das equações de L_Q .

Para $\rho \neq 1$ temos

$$\begin{aligned}
L &= L_Q + \frac{\lambda'}{\mu} = \frac{P_0 c^c \rho^{c+1}}{c!(1 - \rho)^2} [1 - \rho^{k-c}(1 + (1 - \rho)(k - c))] + \frac{\lambda(1 - \frac{c^c \rho^k}{c!} P_0)}{\mu} \\
&= \frac{P_0 c^c \rho^{c+1}}{c!(1 - \rho)^2} [1 - \rho^{k-c}(1 + (1 - \rho)(k - c))] + c\rho \left(1 - \frac{c^c \rho^k}{c!} P_0 \right) \\
&= \frac{P_0 c^c \rho^{c+1}}{c!(1 - \rho)^2} [1 - \rho^{k-c}(1 + (1 - \rho)(k - c))] + c\rho - \frac{c^{c+1} \rho^{k+1}}{c!} P_0 \\
&= \frac{P_0 c^c \rho^{c+1}}{c!(1 - \rho)^2} [1 - \rho^{k-c}(1 + (1 - \rho)(k - c)) - c\rho^{k-c}(1 - \rho)^2] + c\rho \\
&= \frac{P_0 c^c \rho^{c+1}}{c!(1 - \rho)^2} [1 - \rho^{k-c}(1 + (1 - \rho)(k - c\rho))] + c\rho.
\end{aligned}$$

Para $\rho = 1$ temos

$$\begin{aligned}
L &= L_Q + \frac{\lambda'}{\mu} = \frac{c^c}{c!} P_0 \frac{(k - c)(k - c + 1)}{2} + c \left(1 - \frac{c^c}{c!} P_0 \right) \\
&= \frac{c^c}{c!} P_0 \frac{(k - c)(k - c + 1)}{2} + c - \frac{c^{c+1}}{c!} P_0 \\
&= \frac{c^c}{c!} P_0 \frac{(k - c)(k - c + 1)}{2} + c.
\end{aligned}$$

O limite das funções L_Q e L quando $\rho = \frac{\lambda}{\mu}$ tende a infinito é respectivamente, $k - c$ e k . O tempo médio de espera na fila para $\rho \neq 1$ é dado por

$$\begin{aligned}
 W_Q &= \frac{L_Q}{\lambda'} = \frac{\frac{P_0 c^c \rho^{c+1}}{c!(1-\rho)^2} [1 - \rho^{k-c}(1 + (1 - \rho)(k - c))]}{\lambda \left(1 - \frac{c^c \rho^k}{c!} P_0\right)} \\
 &= \frac{\frac{P_0 c^{c-1} \rho^c}{c!(1-\rho)^2} [1 - \rho^{k-c}(1 + (1 - \rho)(k - c))]}{\mu \left(1 - \frac{c^c \rho^k}{c!} P_0\right)} \\
 &= \frac{P_0 c^{c-1} \rho^c}{\mu(1-\rho)^2 (c! - c^c \rho^k P_0)} [1 - \rho^{k-c}(1 + (1 - \rho)(k - c))] .
 \end{aligned}$$

Agora para $\rho = 1$ temos

$$\begin{aligned}
 W_Q &= \frac{L_Q}{\lambda'} = \frac{\frac{c^c P_0}{c!} \frac{(k-c)(k-c+1)}{2}}{\lambda \left(1 - \frac{c^c}{c!} P_0\right)} = \frac{\frac{c^c P_0}{c!} \frac{(k-c)(k-c+1)}{2}}{c\mu \left(1 - \frac{c^c}{c!} P_0\right)} \\
 &= \frac{c^{c-1} P_0 (k-c)(k-c+1)}{2\mu(c! - cP_0)} .
 \end{aligned}$$

Por definição, o tempo de espera no sistema é calculo como

$$W = W_Q + \frac{1}{\lambda}$$

4.5.1 Sistemas $M/M/c/c$

Os modelos $M/M/c/c$, caso particular do modelo $M/M/c/k$ para o qual $k = c$, ou seja, a fila não pode ser formada pois temos o mesmo número de servidores e de clientes. As taxas de chegada e de atendimento são definidas, respectivamente, por

$$\lambda_n = \begin{cases} \lambda, & \text{se } n = 0, 1, 2, \dots, c - 1; \\ 0 & \text{se } n \geq c. \end{cases}$$

$$\mu_n = \begin{cases} n\mu, & \text{se } n = 0, 2, \dots, c; \\ 0, & \text{se } n \geq c + 1. \end{cases}$$

Fazendo $k = c$ em (19) obtemos a expressão de P_n . Isto é,

$$P_n = \frac{\frac{\lambda/\mu^n}{n!}}{\sum_{j=0}^c \frac{\lambda/\mu^j}{j!}} = \frac{\frac{(c\rho)^2}{n!}}{\sum_{j=0}^c \frac{c\rho^j}{j!}}. \quad (25)$$

O tamanho médio da fila e o tempo médio de espera de um cliente na fila é zero, ou seja

$$L_Q = W_Q = 0.$$

Obtemos W e L à partir das fórmulas de Little definida anteriormente, logo

$$W = W_Q + \frac{1}{\mu} = \frac{1}{\mu},$$

$$L = L_Q + \frac{\lambda'}{\mu} = \frac{\lambda(1 - P_c)}{\mu} = c\rho(1 - P_c).$$

onde $\lambda' = \lambda(1 - P_c)$ e P_c é a proporção do tempo em que há c clientes no sistema.

5 APLICAÇÕES

5.1 Fluxos de paletes em indústria química

A Indústria Química da cidade de Campos dos Goytacazes no estado do Rio de Janeiro, fabrica uma ampla lista de produtos e de matérias-primas obtidas por mineração e agricultura com utilidade para variados setores industriais. Produtos como: tintas (para proteger e dar cor às casas), medicamentos e vacinas (para saúde de homens e animais), fertilizantes e defensivos agrícolas (para quantidade e qualidade de alimentos).

Nesse estudo de caso, será analisado o processo de geração de filas por paletes já empacotados pela empilhadeira, sendo esse último momento o atendimento. Paletes são estrados de madeira, metal ou plástico que são utilizados para movimentação de cargas e sua função é viabilizar a otimização do transporte através do uso de empilhadeiras.

Para otimizar o serviço e encontrar pontos críticos na capacidade do sistema, é necessário coletar amostras para calcular os tempos de armazenagem dos paletes com o uso da empilhadeira. O objetivo é observar os números atingidos pela empresa e propor melhorias no desempenho do sistema que retorne maiores lucros, por armazenar mais paletes e diminuir a probabilidade de perdas de processo por excesso de paletes na esteira.

Para isto, foram realizadas medições no tempo de chegada dos paletes na esteira rolante e os tempos médios de armazenamento. O levantamento de dados ocorreu nos meses de Setembro a Dezembro de 2010 por estudantes da Universidade Estadual do Norte Fluminense.

Para cada amostra tem-se o tempo médio de chegada e o tempo médio de serviço. Foram doze observações, de meia hora cada durante uma semana e horários padrões.

A disciplina de atendimento desse processo de filas é a FIFO (Primeiro a chegar, primeiro a sair) e como os paletes não seguem um padrão determinado de armazenagem, os tempos são aleatórios.

A Tabela 3 mostra os instantes de chegadas dos paletes na esteira e os tempos médios de armazenagem, em segundos, coletados.

Tabela 3 – Taxas de chegadas na esteira e armazenamento (em segundos)

Amostra	Chegada(s)	Armazenamento(s)
1	500	530
2	570	524
3	420	566
4	460	563
5	576	500
6	570	502
7	548	600
8	720	659
9	490	530
10	500	549
11	498	580
12	596	508

A média de intervalo entre chegadas de paletes na esteira em minutos e o número de paletes que chegam por hora, são respectivamente,

$$\bar{x}_1 = \frac{6448}{12} \frac{1}{60} = 8.9556 \quad \text{e} \quad \lambda = \frac{1}{a_1} \times 60 = 6,6997$$

O tempo médio de armazenamento em minutos e o número médio de paletes que são armazenados por hora, são respectivamente,

$$\bar{x}_2 = \frac{6611}{12} \frac{1}{60} = 9.1819 \quad \text{e} \quad \mu = \frac{1}{a_2} \times 60 = 6,5345$$

De acordo com esses resultados, os intervalos entre as chegadas e o tempos de atendimento possuem distribuição Exponencial com média λ e μ , respectivamente.

O modelo aplicado pelos pesquisadores a esse processo de filas foi o $M/M/1/k$ com $\lambda = 6,6997$ número de paletes que chegam por hora e $\mu = 6,5345$ paletes atendimentos por hora. O valor de k é igual 4, ou seja, três vagas na esteira para paletes e um palete sendo armazenado. Lembrando que para $n > k$ temos que os valores de λ e μ são iguais a zero.

A taxa de ocupação do serviço é dada por

$$\rho = \frac{\lambda}{\mu} = 1.0252$$

Esse resultado nos diz que chegam mais paletes à esteira do que a empilhadeira consegue armazenar, ou seja, demanda maior que a oferta. Agora, vamos calcular a probabilidade de atendimento imediato, dada por

$$P_0 = \begin{cases} \frac{1}{k+1} = 0.2, & \text{para } \rho = 1 \\ \frac{1-\rho}{1-\rho^{k+1}} = 0.1901 & \text{para } \rho \neq 1 \end{cases}$$

A proporção do tempo em que há n paletes na esteira esperando pela empilhadeira ou sendo armazenado, é igual a

$$P_n = \begin{cases} \rho^n \frac{1-\rho}{1-\rho^{k+1}} = 1.02^n \times 0.19, & \text{se } n = 0, 1, \dots, k. \quad (\rho \neq 1); \\ \frac{1}{k+1} = 0.2, & \text{se } n = 0, 1, \dots, k. \quad (\rho = 1); \\ 0, & \text{se } n > k. \end{cases}$$

Com isto podemos encontrar os valores para as medidas de desempenho desse modelo $M/M/1/4$, que são os seguintes

1. $L = \begin{cases} 2, & \text{para } \rho = 1 \\ 2.0497, & \text{para } \rho \neq 1 \end{cases}$ (número médio de clientes no sistema);
2. $W_Q = \begin{cases} 0.2239, & \text{para } \rho = 1 \\ 0.2343, & \text{para } \rho \neq 1 \end{cases}$ (tempo médio de cada unidade na fila);
3. $W = \begin{cases} 0.3731, & \text{para } \rho = 1 \\ 0.3873, & \text{para } \rho \neq 1 \end{cases}$ (tempo médio de cada unidade no sistema);

$$4. L_Q = \begin{cases} 1.2, & \text{para } \rho = 1 \\ 1.2399, & \text{para } \rho \neq 1 \end{cases} \quad (\text{número médio de clientes na fila}).$$

Calculamos todas as medidas de desempenho nesse modelo específico, tanto para $\rho = 1$ quanto para $\rho \neq 1$, as diferenças entre os resultados encontrados são sutis, pois $\rho \approx 1$.

Agora, propomos uma parametrização diferente. A fim de maximizar a armazenagem e diminuir o tempo de espera por paletes na fila (esteira), imaginamos o seguinte caso

1. Taxa média de chegada (Poisson) = $\lambda_n = \begin{cases} 6.6997, & n = 0, 1, \dots, k-1; \\ 0, & n \geq k. \end{cases}$
2. Taxa de serviço médio (Exponencial) = $\mu_n = \begin{cases} n \times 6.5345, & n = 0, 1, \dots, c-1; \\ c \times \mu = 13.069, & n = 2, \dots, 4. \end{cases}$
3. Número de atendentes (empilhadeira) = 2;
4. Capacidade do sistema (esteira + armazenagem) = 4 (1 em atendimento e 3 na fila);
5. Disciplina de atendimento = FIFO.

Essas características definem um modelo $M/M/c/k$, com $c = 2$ empilhadeiras e $k = 4$ número máximo de paletes no sistema. A taxa de ocupação de serviço desse modelo é dado por $\rho = \lambda/c\mu = 0.5126$, ou seja, o servidor está ocupado em 51% do tempo realizando atendimento.

Adotando $r = \lambda/\mu = 1.0252$, podemos calcular a probabilidade de atendimento imediato e a proporção do tempo em que há n paletes no sistema, dada pela equação (22).

$$P_0 = \begin{cases} 0.3380, & \rho \neq 1, \\ 0.2776, & \rho = 1. \end{cases}$$

No nosso modelo proposto, $M/M/2/4$, temos $\rho \neq 1$, por isso usamos a primeira expressão de P_0 para calcular P_n dada pela expressão em (19). Então,

$$P_n = \begin{cases} \frac{1.0252^n}{n!} \times 0.3380, & 1 \leq n < c; \\ \frac{2^n \times 0.5126^n}{2} \times 0.3380, & c \leq n \leq k \\ 0, & \text{caso contrário.} \end{cases}$$

Agora vamos determinar os valores para as medidas de desempenho desse modelo para $\rho \neq 1$, que são dados por

1. $L_Q = \frac{P_0 c^c \rho^{c+1}}{c!(1-\rho)^2} [1 - \rho^{k-c}(1 + (1-\rho)(k-c))] = 0.1844;$
2. $L = \frac{P_0 c^c \rho^{c+1}}{c!(1-\rho)^2} [1 - \rho^{k-c}(1 + (1-\rho)(k-c\rho))] + c\rho = 1.2598;$
3. $W_Q = \frac{P_0 c^{c-1} \rho^c}{\mu(1-\rho)^2(c! - c^c \rho^k P_0)} [1 - \rho^{k-c}(1 + (1-\rho)(k-c))] = 0.039; \cdot$
4. $W = W_Q + \frac{1}{\lambda} = 0.1818.$

Nota-se que o modelo $M/M/2/4$ apresenta melhores resultados que o modelo proposto pelos pesquisadores, pois o número médio de paletes na fila para armazenagem diminuiu de 1.4 para quase zero. E o tempo de cada palete no sistema passa a ser 10.9 minutos, sendo que destes, apenas 2.4 minutos de espera na fila e o restante de tempo para o atendimento (armazenagem pela empilhadeira). A probabilidade do sistema estar cheio P_4 e a produção parar cai de 21% para 5%.

Esse novo sistema de filas proposto atende às expectativas da motivação inicial, que através da introdução de uma nova empilhadeira há redução da probabilidade de parada e da porcentagem de perdas no processo, com custo absorvido plenamente pelo custo de oportunidade de armazenar mais paletes.

5.2 Processos de atendimento de emergência

O processo de geração de filas nesse caso está associado à recepção do Pronto Atendimento na unidade de São Mateus, situado no estado do Espírito Santo. Sendo que a população são os hospitais do setor público de saúde desse município.

A superlotação nos hospitais provocam grandes esperas nesses locais de atendimento e o tempo perdido em filas podem causar impactos e prejudicar a saúde dos pacientes. Os indicadores de desempenho aqui propostos auxiliarão na tomada de decisão para minimizar o tamanho das filas e tempo de espera para a recepção dessa unidade.

A coleta de dados foi realizada por Karina Pedrini Fraga e três voluntariados da Universidade Federal do Espírito Santo em 2011, por meio de planilhas para o tempo observado. Não houve distinção entre os usuários do sistema em relação à priorização de atendimento a deficientes físicos, gestantes ou idosos.

O levantamento de dados ocorreu numa segunda-feira no dia 05 de Setembro de 07:30 às 10:30, com início uma hora antes do horário de pico e fim uma hora depois do horário de pico. Ele é delimitado da porta pela qual os usuários entram no Pronto Atendimento até um guichê na sala de recepção, onde o atendimento é realizado. Ou seja, os pacientes entram no sistema pela porta e aguardam na fila, são atendidos de acordo com a disciplina FIFO (primeiro a chegar, primeiro a ser atendido). O processo de atendimento ocorre no guichê quando preenchem uma ficha com os seus dados pessoais e a especialidade médica de que necessitam.

As chegadas dos pacientes são variáveis independentes, os eventos são espontâneos e não possui interferência dos pesquisadores. O critério de probabilidade se estabelece através da aleatorização da primeira unidade amostral, sendo um processo probabilístico não aleatório que caracteriza uma amostragem sistemática. Tem como propósito cobrir a população de pacientes com um modelo simples e uniforme, visto que foi observado o que realmente tem importância para o estudo, ou seja, planejado.

Os horários de chegadas e os intervalos entre chegadas dos pacientes ao Pronto

Atendimento, em segundos, são descritos na Tabela 4.

Tabela 4 – Tempo e Intervalo de chegadas (em segundos)

Paciente	Tempo	Intervalo	Paciente	Tempo	Intervalo
1	27210	-	34	30710	320
2	27290	80	35	30850	140
3	27345	55	36	30980	130
4	27380	35	37	31330	350
5	27580	200	38	31460	130
6	27625	45	39	31505	45
7	27660	35	40	31665	160
8	27740	80	41	31770	105
9	28245	505	42	32045	275
10	28360	115	43	32145	100
11	28530	170	44	32245	100
12	28740	210	45	32290	45
13	28780	40	46	32730	440
14	28805	25	47	32790	60
15	28910	105	48	32825	35
16	28940	30	49	33020	195
17	29075	135	50	33400	380
18	29110	35	51	33465	65
19	29160	50	52	33540	75
20	29180	20	53	33685	145
21	29300	120	54	34120	435
22	29310	10	55	34415	295
23	29320	10	56	34460	45
24	29600	280	57	34750	290
25	29745	145	58	34770	20
26	29840	95	59	35020	250
27	29960	120	60	35490	470
28	29990	30	61	35650	160
29	30185	195	62	35690	40
30	30205	20	63	36020	330
31	30230	25	64	36435	415
32	30380	150	65	36630	195
33	30390	10			

O número de pacientes que chegam por minuto e o intervalo médio entre chegadas, são respectivamente,

$$\lambda = \frac{65}{36630 - 27210} \times 60 = 0.4140 \quad \text{e} \quad \bar{x}_1 = \frac{9420}{64 \times 60} = 2.4531$$

Esses resultados indicam que: chegam em média 2 pacientes a cada 5 minutos, e o intervalo médio entre cada chegada é de apenas 2.45 minutos.

Antes do momento de início da coleta de dados, já haviam 5 clientes no Pronto Atendimento, por isso 70 pacientes foram atendidos e 65 chegaram no sistema durante o tempo de observação. Somente os 64 intervalos entre chegadas são contabilizados, já que o primeiro paciente chega no momento 0 e não existe intervalo anterior.

Para verificar se o intervalo entre as chegadas possui distribuição Exponencial, realizamos o teste Kolmogorov-Smirnov de aderência. A estatística $D = \max_x |S_n(x) - F_0(x)| = 0.0645$ é menor que o quantil do teste que é igual a 0.2037 com $\alpha = 0.01$ (99% de confiança), ou seja, não rejeitamos a hipótese nula e podemos assumir que os dados se ajustam a uma distribuição exponencial.

Agora faremos a análise descritiva para os dados de atendimento. A Tabela 5 contém o tempo de serviço, para cada um dos 70 pacientes observados.

Tabela 5 – Tempo, em segundos, de atendimento:

Paciente	Tempo(s)	Paciente	Tempo(s)
1	26	36	48
2	22	37	47
3	118	38	21
4	204	39	24
5	43	40	124
6	102	41	83
7	110	42	47
8	168	43	90
9	87	44	20
10	177	45	37
11	95	46	99
12	184	47	22
13	107	48	11
14	125	49	27
15	135	50	228
16	129	51	38
17	142	52	57
18	141	53	301
19	37	54	239
20	150	55	173
21	201	56	45
22	305	57	165
23	128	58	19
24	178	59	26
25	78	60	170
26	228	61	124
27	145	62	148
28	141	63	203
29	115	64	54
30	358	65	218
31	135	66	217
32	148	67	157
33	166	68	169
34	33	69	117
35	46	70	133

O número médio de pacientes atendidos por minuto e o tempo médio de atendimento são, respectivamente,

$$\mu = \frac{70}{8408} \frac{1}{60} = 0.4995 \quad \text{e} \quad \bar{x}_2 = \frac{8408}{70 \times 60} = 2.3352$$

Esses resultados indicam que: são atendidos em média 1 paciente a cada 2 minutos, e o tempo médio de atendimento é 2.3 minutos por paciente.

Para verificar se o tempo de atendimento possui distribuição Exponencial, realiza-se o teste Kolmogorov-Smirnov de aderência. A estatística $D = \max_x |S_n(x) - F_0(x)| = 0.1614$ é menor que o quantil do teste que é igual a 0.1948 com $\alpha = 0.01$ (99% de confiança), ou seja, não rejeitamos a hipótese nula e podemos assumir que os dados se ajustam a uma distribuição exponencial.

Como ambas as taxas são exponencialmente distribuídas, o modelo $M/M/1$ se adequou nesse estudo com $\lambda = 0.4140$ número de chegada por minuto e $\mu = 0.4995$ número de atendimentos por minuto. O único servidor está ocupado realizando atendimento em 82% do tempo (taxa de ocupação). A probabilidade do sistema estar vazio e a proporção do tempo em que há n pacientes no Pronto Atendimento, são dados pelas equações (8) e (9).

$$P_0 = 1 - \frac{\lambda}{\mu} = 1 - 0.8210 = 0,1790$$

$$P_n = \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^n = (1 - 0.8210)(0.8210)^n = (0.1790)(0.8210)^n$$

O resultado encontrado para P_0 indica que em 18% do tempo o sistema está ocioso e indica grande demanda ao serviço desse setor.

Os valores encontrados para as medidas de desempenho foram os seguintes

1. $L = 4.56$ (número médio de clientes no sistema);
2. $W_Q = 9.11$ (tempo médio de cada unidade na fila);
3. $W = 11.11$ (tempo médio de cada unidade no sistema);

4. $L_Q = 3.74$ (número médio de clientes na fila).

Visto que, essa fila é somente para o atendimento no guichê, o tempo de espera médio é relativamente alto e esse tempo pode ser crucial para o bem-estar do paciente.

As filas nos setores públicos devem ser rápidas e eficientes. Para isto, a fim de maximizar o atendimento e diminuir o tempo de espera na fila, proporemos uma parametrização diferente, da seguinte forma

1. Taxa média de chegada (Poisson) = $\lambda_n = 0.4140$;
2. Taxa de serviço médio (Exponencial) = $\mu_n = 0.4995$;
3. Número de atendentes = 2;
4. Disciplina de atendimento = FIFO.

Essas características definem um modelo $M/M/c$, com $c = 2$ guichês, ou seja, propomos um guichê adicional ao sistema de Pronto Atendimento. Temos que $r = \lambda/\mu = 0.8210$ e $\rho = \lambda/c\mu = 0.4105$, então a probabilidade de atendimento imediato, dada por (15) e a proporção do tempo em que há n pacientes no sistema, dada por (14), são iguais a

$$P_0 = \left(1 + 0.82 + \frac{0.82^2}{2!(1 - 0.41)} \right)^{-1} = \frac{1}{2.38983} = 0.4184.$$

$$P_n = \begin{cases} \frac{(0.414)^n}{n!(0.4995)^n} \times 0.4184, & 1 \leq n < 2, \\ \frac{(0.414)^n}{2^{(n-2)}2!(0.4995)^n} \times 0.4184, & n \geq 2. \end{cases}$$

O resultado de P_0 nos diz que em 41% do tempo o sistema fica vazio, e os atendentes dos guichês ficam ociosos. Agora, calculamos as medidas de desempenho para esse modelo de filas, encontradas na seção 4.3.

1. $L_Q = \frac{0.8210^2 \times 0.4105}{2!(1 - 0.4105)^2} \times 0.4184 = 0.1665$;

$$2. W_Q = \frac{0.8210^2}{2!(2 \times 0.4995)(1 - 0.4105)^2} \times 0.4184 = 0.4061;$$

$$3. L = L_Q + r = 0.1665 + 0.8210 = 0.9875$$

$$4. W = W_Q + \frac{1}{\mu} = 0.4061 + \frac{1}{0.4995} = 2.4081.$$

Esses resultados comprovam que o modelo $M/M/2$ otimizaria o atendimento e reduziria o tempo de espera na fila pelos pacientes para a recepção nesse Pronto Atendimento. Visto que, o número médio de usuários na fila passaria de 3,7 para quase zero e o número médio de usuários no sistema não chega a um paciente, ou seja, a fila inexistiria. Além disso, o tempo médio de espera na fila passaria a ser 24 segundos em relação ao modelo $M/M/1$ que era aproximadamente 9 minutos.

A demanda de pacientes por serviço médico nesse Pronto Atendimento seria atendida com sucesso e sem filas. O sistema de filas com 2 servidores tornaria a recepção dessa unidade eficiente e de qualidade. Do ponto de vista do servidor, a não existência de filas indica um superdimensionamento do serviço com alta ociosidade dos funcionários, 41% do tempo. Considerando ainda os valores das medidas de desempenho para o sistema com um guichê, a inserção de mais um guichê de atendimento pode não se justificar devido ao investimento necessário para manter mais um atendente.

6 SIMULAÇÃO PARA O MODELO $M/M/1$

O objetivo dessas simulações é obter estimativas para o tempo médio de espera na fila, o tamanho médio da fila e a taxa de ocupação do servidor para o modelo $M/M/1$. A ideia principal é realizar n simulações com números aleatórios usando implementações em contexto R, obter intervalos de confiança para a estimativa média e comparar esses resultados com as fórmulas teóricas encontrados na Seção (4.1) para esse modelo específico.

6.1 Estimativas médias

Para encontrar as estimativas médias do tempo de espera na fila, taxa de ocupação e tamanho da fila, fazemos as simulações de acordo com a tolerância do erro estipulado. O valor de n é calculado a partir da fórmula que maximiza a variância pelo método conservativo, dada por

$$n = \left(\frac{z_{0.95} \times \sigma}{\epsilon} \right)^2.$$

onde,

1. $z_{0.95}$ é o quantil da normal com 95% de confiança;
2. $\epsilon = 0.01$ (erro de precisão);
3. $\sigma = 1/4$ (variância).

Com isto temos que

$$n = \left(\frac{1.644854}{4 \times 0.01} \right)^2 = 1690.965.$$

Ou seja, o sistema será simulado 1691 vezes. Os números aleatórios e independentes que foram gerados para os tempos entre as chegadas e para os tempos de atendimento provém da distribuição exponencial, com parâmetros $\lambda = 2$ e $\mu = 3$, respectivamente. O tempo total de simulação foi definido por 500, logo o número médio de clientes é $2 \times 500 = 1000$.

A Tabela (6) contém a estimativa ponderada simulada pelo tamanho de n , (\bar{x}) , para algumas medidas de desempenho, e seu respectivo valor teórico do modelo $M/M/1$:

Tabela 6 – Estimativas médias e valor teórico

Estimativa	Simulada	Teórica
Tempo de espera médio na fila	0.6645945	0.666667
Taxa de ocupação média	0.6654305	0.666667
Tamanho médio da fila	1.332143	1.33333

6.2 Intervalos de confiança para as estimativas

Os intervalos das estimativas foram calculados com 95% de confiança. As variâncias para o tempo de espera médio na fila, taxa de ocupação média e tamanho médio da fila são respectivamente, 0.0215, 0.0008 e 0.0993. Os intervalos de confiança são calculados através da seguinte fórmula

$$IC = \bar{x} \pm z_{\alpha/2} \times \sqrt{\sigma^2}$$

Os intervalos das estimativas médias são descritos na Tabela (7).

Tabela 7 – Intervalos inferiores e superiores para as estimativas médias

Estimativa	Intervalo inferior	Intervalo Superior
Tempo de espera médio na fila	0.3770789	0.9521101
Taxa de ocupação média	0.6070223	0.7238387
Tamanho médio da fila	0.7143333	1.949953

Os resultados nos mostram que: a maior amplitude do confiança é a do tamanho médio da fila, com 1.2356. Esperada, por ter a maior variância; as estimativas encontradas através de simulações são bem próximas do valor real; os intervalos de confiança contém o verdadeiro valor da fórmula teórica. E, por fim, a eficiência desse algoritmo torna capaz a inserção de dados que não sejam exponencialmente distribuídos.

7 CONSIDERAÇÕES FINAIS

Ao longo deste trabalho foram estudados e apresentados alguns dos modelos de filas mais utilizados, que são as filas markovianas. O objetivo foi alcançado, compreendemos seus comportamentos, assim como toda a estatística envolvida, a partir de análises do tamanho médio da fila, da quantidade média de usuários no sistema e para os tempos médios de espera na fila e de permanência no sistema.

O estudo da Teoria das Filas tem grande importância, uma vez que estão presentes em diversas situações do cotidiano. Por isto, ainda há muito o que explorar sobre a Teoria das Filas e modelos mais complexos.

Na aplicação sobre fluxo de paletes em uma indústria química, o modelo de filas aplicado pelos pesquisadores foi o sistema $M/M/1/4$, e propomos uma empilhadeira adicional, substituindo o modelo anterior por um sistema $M/M/2/4$. Provamos que o nosso modelo proposto apresenta ótimos resultados, onde a fila quase inexistiria.

Já no estudo de caso sobre o serviço em um Pronto Atendimento, o modelo adotado foi o $M/M/1$. Propomos um guichê adicional, o sistema $M/M/2$, e este apresentou valores médios significativos para as medidas de desempenho e qualidade de atendimento.

Com o estudo simulado para o modelo de fila $M/M/1$, vimos que as estimativas médias encontradas através de simulações são bem próximas do valor teórico. Tornando nosso algoritmo, um sucesso. A partir deste, podemos explorar modelos que não tenham distribuições e padrões associados.

REFERÊNCIAS

- CORTES, CHAVES, ROMERO, SALES, VILAÇA. **Aplicação da Teoria das Filas na maximização do fluxo de paletes em um Indústria Química.** UENF - Rio de Janeiro, 2010.
- COSTA, L.C. **Teoria das Filas.** Universidade Federal do Maranhão - UFMA, 2006.
- COX, D. e SMITH, W. **Queues.** Methuen, 1961.
- CUYABANO, JUNG. **Teoria de Filas.** Instituto de Matemática, Estatística e Computação Científica, UNICAMP Campinas, 2009.
- FRAGA K. P. **Análise de um processo de atendimento de emergência usando Teoria de Filas.** UFES - São Mateus, 2012.
- GNEDENKO, B.V., KOVALENKO, I.N. **Introduction to Queueing Theory.** Second Edition, Birkhauser, Boston, 1989.
- HOEL, S.C. Port, Charles J. Stone, **Introduction to stochastic Processes.**
- KENDALL, D. **Some problems in the theory of queues.** J.R. Statist. Soc., Vol. 13, 1951.
- MAGALHAES, M.N. **Introdução à Rede de Filas.** ABE - São Paulo, 1996.
- MÜLLER, D. **Processos Estocásticos e Aplicações.** Almedina, 2007.
- PEREIRA. **Uma Introdução às Filas de Espera.** Dissertação de Mestrado, Departamento de Matemática e Engenharias, Universidade de Madeira, 2009.
- ROSS, **Introduction to Probability Models.** Ninty Edition, University of California Berkeley, 2007.
- SAATY, T.L. **Elements of Queueing Theory With Applications.** McGraw-Hill Book Company, New York, 1961.
- TAKACS, L. **Introduction to the Theory of Queues.** Oxford University Press, New York, 1962.

Apêndices

Nesse apêndice temos os códigos em contexto R para a seção 6 (Simulação para o modelo $M/M/1$). As nomenclaturas utilizadas têm o seguintes significados

1. T: período total de simulação;
2. tc_i : instante de tempo em que ocorre a i -ésima chegada;
3. ta_i : instante de tempo em que ocorre o i -ésimo atendimento;
4. to: tempo total de ociosidade do servidor;
5. te: tempo total de espera na fila;
6. tme: tempo médio de espera na fila;
7. i : contador;
8. t: variável de controle (tempo).

Script:

```

sistema<- function(T,l,u)
{ #cria o vetor dos tempos de chegada dos clientes (tc)
#a partir de uma exponencial de média 1/l
#referente aos tempos entre as chegadas (tec).
t=0
i=1
tc<- as.vector(rep(0,(1.3*T*1))) #criando vetor de zeros de tamanho 30% maior
que o número médio de clientes = T*1.
while (t<T){
tec<- rexp(1,l)
tc[i]<- t+tec
t<- tc[i]
i=i+1 }

```

56

```
#arrumando o vetor tc.
```

```
tc<- subset(tc,tc!=0) #retirando os zeros excedentes do vetor.
```

```
n<- length(tc)-1
```

```
tc<- tc[1:(n)]
```

```
#criando o vetor de tempo de atendimento, ta, para os n clientes a partir de uma  
exponencial de média 1/u
```

```
#e acumulando o tempo ocioso ou o tempo de espera na fila do sistema.
```

```
i=1
```

```
to=tc[1]
```

```
te=0
```

```
t=tc[1]
```

```
ta<- as.vector(rep(0,n))
```

```
while (i<n) {
```

```
ta[i]<- rexp(1,u)
```

```
if (t+ta[i]<tc[(i+1)])
```

```
{ to<- to + tc[i+1]-(t+ta[i])
```

```
t<- tc[i+1] }
```

```
else
```

```
{ te<- te + (t+ta[i])-tc[i+1]
```

```
t<- t+ta[i] }
```

```
i=i+1}
```

```
#fazendo o caso particular de i=n (pois não existe n+1).
```

```
ta[i]<- rexp(1,u)
```

```
if(t+ta[i]<T)
```

```
{ to<- to+(T-t-ta[i]) }
```

```

#cálculo teórico e empírico do tempo de espera médio do sistema
#e da taxa de ocupação do sistema.
te_medio<- te/n
te_medio_teorico<- 1/(u*(u-1))
tx_ocup<- 1-(to/T)
tx_ocup_teorico<- 1/u

#print(tc)
#print(ta)

#criando um vetor que indica em que tempo o cliente saiu do sistema,
#levando em consideração duas possibilidades:
#1) o cliente chega, é atendido e sai.
#2) o cliente chega, espera na fila, é atendido e sai.
ts<- as.vector(rep(0,n))
ts[1]<- tc[1]+ta[1]

for(i in 2:n){
  if(tc[i]<ts[i-1])
  { ts[i]<- ts[i-1]+ta[i] }
  else
  { ts[i]<- tc[i]+ta[i] } }
#print(ts)

#criando um data.frame com as variáveis:
#"CLIENTE": indicando o número do cliente no sistema;
#"SITUAÇÃO": indicando se o cliente está chegando ou saindo do sistema;
#"CÓDIGO_SITUAÇÃO": igual à 1 quando o cliente entra no sistema e -1 quando sai;

```

58

#"TEMPO": indicando em que momento do tempo está a observação.

```
x<-as.data.frame(matrix(c(rep(1:n,2),
rep("chegada",n),
rep("saida",n),
rep(1,n),
rep(-1,n),
tc,
ts),nrow=(2*n),ncol=4,dimnames=list(c(1:(2*n)),c("CLIENTE",
"SITUACAO",
"CODIGO_SITUACAO",
"TEMPO"))))
```

```
x$CODIGO_SITUACAO=as.numeric(as.character(x$CODIGO_SITUACAO)) #transformando o
CÓDIGO_SITUAÇÃO
```

```
x$TEMPO=as.numeric(as.character(x$TEMPO)) #e o TEMPO em numérico para
manipulação.
```

```
x<-x[order(x$TEMPO),] #ordenando os dados em relação ao TEMPO.
```

```
#calcular o tamanho da fila como a soma dos códigos até o momento.
```

```
for(k in 1:(2*n))
```

```
{ x$FILA[k]<- sum(x$CODIGO_SITUACAO[1:k])-1 } #(-1) que representa ociosidade no
sistema.
```

```
#calcular o tempo em que o sistema ficou em cada tamanho de fila.
```

```
for(k in 2:(2*n))
```

```
{ x$PROPORCAO[1]<- x$TEMPO[2]
```

```
x$PROPORCAO[k]<- x$TEMPO[k+1]-x$TEMPO[k] }
```

```
#print(tail(x))
```

```

#calculando o tamanho médio da fila ponderado pelo tempo:
y<-subset(x,x$FILA>0) #separando os números positivos para o tamanho da fila,
pois quando 0, a multiplicação é nula.
y<-y[order(y$TEMPO),]

teste<- sum(x$PROPORCAO[1:(length(x$PROPORCAO)-1)])
#print(teste)

tamanho_medio_fila<- (sum(y$FILA * y$PROPORCAO))/teste
tamanho_medio_da_fila_teorico<- ((1/u)^2)/(1-(1/u))

#imprimir os resultados!
final<-list("Tempo médio de espera na fila"= te_medio,
"Tempo médio de espera na fila (teorico)"= te_medio_teorico,
"Taxa de ocupação"= tx_ocup,
"Taxa de ocupação (teorico)"= tx_ocup_teorico,
"Tamanho médio da fila ponderado"= tamanho_medio_fila,
"Tamanho médio da fila (teórico)"= tamanho_medio_da_fila_teorico)
return(final)}

# quanto maior o número médio de clientes (T*1) mais próximas serão as medidas
empíricas das teóricas.

(n<-(qnorm(.95)/(4*0.02))^2) #Definindo o tamanho de n
te_medio<-numeric(n)
tx_ocup<-numeric(n)
tamanho_medio_fila<-numeric(n)

```

60

```
for (i in 1:n){ #Gerando as simulações
sim<-sistema(500,2,3)
te_medio[i]<-sim[[1]]
tx_ocup[i]<-sim[[3]]
tamanho_medio_fila[i]<-sim[[5]]}

#Cálculo das médias
(te_medio_medio<- mean(te_medio))
(tx_ocup_medio<-mean(tx_ocup))
(tamanho_medio_medio<-mean(tamanho_medio_fila))

#Cálculo dos desvios padrões
(te_medio_sd<- sd(te_medio))
(tx_ocup_sd<- sd(tx_ocup))
(tamanho_medio_sd<- sd(tamanho_medio_fila))

#Intervalos de confiança
(ICinf1<- te_medio_medio - (qnorm(.975))*te_medio_sd)
(ICsup1<- te_medio_medio + (qnorm(.975))*te_medio_sd)

(ICinf2<- tx_ocup_medio - (qnorm(.975))*tx_ocup_sd)
(ICsup2<- tx_ocup_medio + (qnorm(.975))*tx_ocup_sd)

(ICinf3<- tamanho_medio_medio - (qnorm(.975))*tamanho_medio_sd)
(ICsup3<- tamanho_medio_medio + (qnorm(.975))*tamanho_medio_sd)
```