Escola de Engenharia Mauá

ECM501 – Teoria dos Grafos, Pesquisa Operacional e ~Métodos de Otimização

Prof. Joyce M Zampirolli
joyce.zampirolli@maua.br

Teoria das Filas

Créditos: prof; Aldo William Medina Garay, 2012
Maio/2019

Você já pegou uma fila?

Você já pegou uma fila?

MOTIVAÇÃO

• As filas estão presentes em nosso cotidiano, no supermercado, no banco, no trânsito, em qualquer situação em que precisamos esperar por um serviço ou oportunidade.
• Um sistema de filas pode ser descrito como clientes que chegam para um determinado serviço em que são atendidos imediatamente ou esperam, saindo após o atendimento.
• O principal motivo de se estudar Teoria de Filas é otimizar o sistema, que se caracteriza por:

- melhor utilização dos serviços disponíveis,
- menor tempo de espera,
- maior rapidez no atendimento.

Estrutura de um sistema de fila de espera

- **Fonte** ou **População**, que gera os clientes que vão chegar ao sistema.
- **Fila**, construída pelos clientes à espera de ser atendidos (Não inclui o(s) cliente(s) em atendimento).
- **Serviço** ou atendimento, que pode ser constituído por um ou mais postos de atendimento.

$$\text{Fila} + \text{Serviço} = \text{Sistema}$$

- Número de clientes no sistema (Em cada instante) = **Estado do sistema**.

Estrutura de um sistema de fila de espera

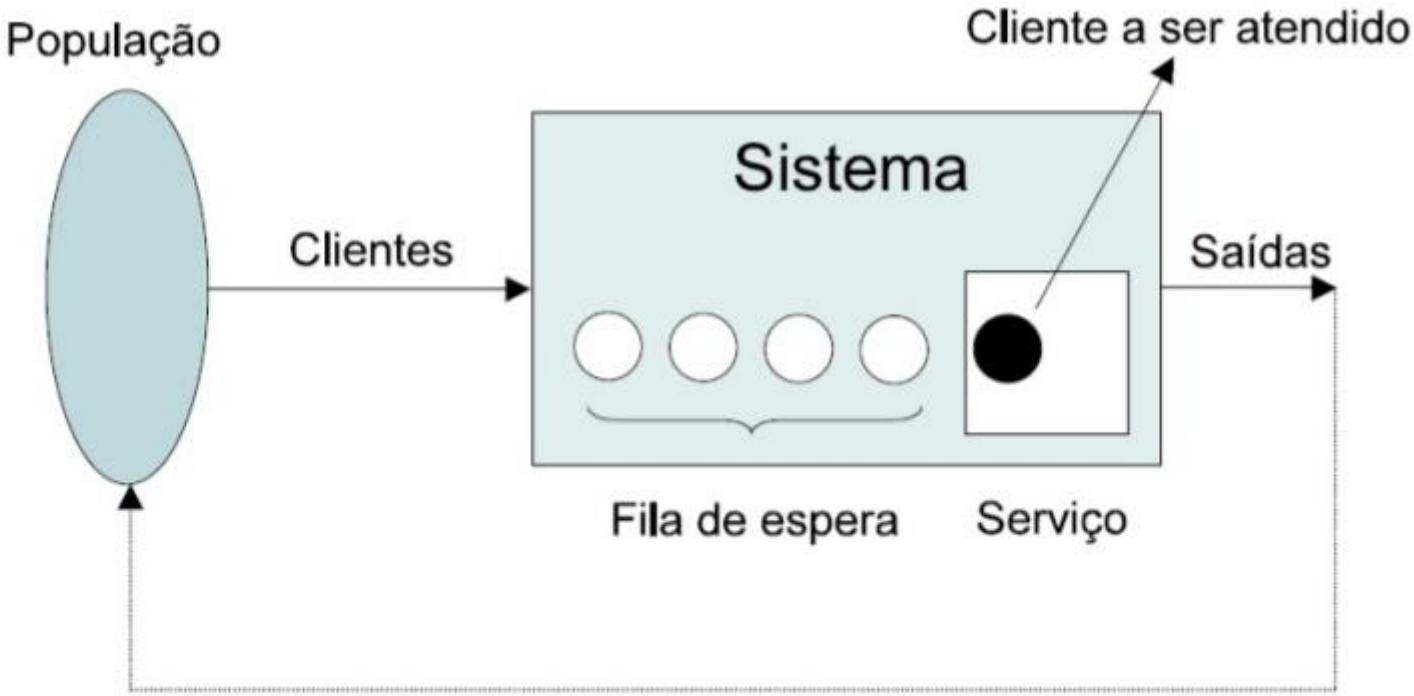

Figura: Estrutura de um sistema de fila de espera

FONTE

- **Dimensão da população:**

  - Infinita: quando a probabilidade de ocorrer uma nova chegada não é influenciada pelo número de clientes que já se encontram no sistema .
  - Finita

- **Dimensão da chegada:**

  - Clientes chegam um a um.
  - Clientes chegam em grupo.

- **Controle das chegadas:**

  - Chegadas controláveis (Por exemplo, inscrições em dias fixos).
  - Chegadas incontroláveis (Por exemplo, urgência de um hospital).

CLIENTE

- **Distribuição das chegadas:**

  - O padrão das chegadas pode ser descrito pelo tempo entre duas chegadas consecutivas (**Tempo entre chegadas**) ou pelo número de chegadas por unidade de tempo. (Distribuião das chegadas) .

- **Taxas das chegadas** (λ)**:**

  -Número médio de clientes que procuram o serviço por unidade de tempo.

- **Atitude dos clientes:**

  - *Paciente*, permanecem na fila até serem atendidos.
  - *Impaciente*, desistem de esperar ou simplemente não se juntam à fila se esta for muito grande.

FILA

- **Número de filas:**

  - Fila simples: uma única fila mesmo que o servidor tenha vários postos de atendimento.
  - Fila múltipla: uma fila por posto de atendimento; cada posto de atendimento constitui um sistema separado de fila de espera.

- **Comprimento da fila:**

  - Infinito: A capacidade máxima da fila é muito grande quando comparada com o número de elementos que habitualmente a constituem.
  - Finito: A fila pode acolher apenas um número determinado (pequeno) de clientes.

- **Disciplina da fila:**

  - FIFO: 'First in First Out'.
  - Prioridades: Reservas, idade, emergência.

SERVIÇO

- **Estagios do serviço:**

  - Um sistema com um único Estágio: Por exemplo uma barberia
  - Um sistema de multi-estágio

- **Dimensão do serviço:**

  - Simples;
  - Em grupo (Por exemplo um elevador atende vários clientes simultaneamente).

- **Distribuição do tempo de serviço:**

  - Constante
  - Aleatório: Distribuição exponencial, Erlang, entre outras.

- **Taxa de serviço** (μ)**:**

  - Número médio de clientes que podem ser atendidos por cada servidor por unidade de tempo. $\frac{1}{\mu}$ é a duração média do serviço.

MEDIDAS DE DESEMPENHO

- Comprimento médio da fila (L_q).
- Número médio de clientes no sistema (L).
- Tempo médio de espera da fila (W_q).
- Tempo médio de espera no sistema (W).
- Tempo médio de ocupação (e desocupação) do serviço (percentagem de tempo durante o qual o serviço está ocupado).

NOTAÇÃO NOS SITEMAS DE FILAS: X/Y/Z/W

- A notação de processos de filas mais utilizada atualmente foi proposta por Kendall, em 1953.

- X, Y representam as distribuições do intervalo de tempo entre chegadas e do tempo de serviço respectivamente, onde:
  - M representa a distribuição exponencial,
  - G representa uma distribuição não especificada,
  - D representa as chegadas ou atendimentos determinísticos.

- Z, representa o número de servidores em paralelo.

- W, representa outras características do sistema, tais como comprimento da fila.

- Entre os tipos de filas mais conhecidos temos o $M/M/1$, $M/M/c$, $M/M/\infty$, entre outros.

TIPO M/M/1

- Suponha que os clientes chegam a uma estação de um único servidor, de acordo com um processo de Poisson com taxa de λ.
- Após a chegada, cada cliente se encaminha diretamente ao serviço se o servidor está livre, caso contrário ele espera na fila.
- Quando o servidor termina de atender um cliente, o cliente deixa o sistema e o próximo cliente da fila será atendido.
- Os tempos sucessivos de serviço de atendimento dos clientes são considerados variáveis aleatórias exponenciais independentes com média $1/\mu$.
- Seja $X(t)$ o número de clientes no sistema no tempo t, então $\{X(t), t \geq 0\}$.

$$\mu_n = \mu \quad n \geq 1$$
$$\lambda_n = \lambda \quad n \geq 0$$

TIPO M/M/1

- **Distribuição Estacionária**

Seja $\pi_n(t) = P(N(t) = n)$, pode-se mostrar que:

$$\pi_{n+1} = -\frac{(\lambda+\mu)}{\mu}\pi_n + \pi_{n+1} + \frac{\lambda}{\mu}\pi_{n-1}, \qquad n \geq 1$$

$$\pi_1 = \frac{\lambda}{\mu}\pi_0$$

Se $\lambda > \mu$ o número médio de chegadas por unidade de tempo é maior que o número médio de saidas por unidade de tempo, e $X_t \to \infty$ quase certamente.

Se $\lambda = \mu$ o processo é recorrente nulo.

Se $\lambda < \mu$, o processo é recorrente positivo, e considerando $\sum_{n=1}^{\infty} \pi_n = 1$, pode-se mostrar por indução que a distribuição estacionária é dada por:

$$\pi_n = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^n \qquad (1)$$

TIPO M/M/1

- **Valor esperado do número de clientes**

Sob condição de equilíbrio, seja N o número total de usuários no sistema, temos que seu valor esperado é:

$$L = E\pi\left(X_t\right) = E\left(N\right) = \sum_{n=0}^{\infty} n\pi_n = \sum_{n=0}^{\infty} n\left(1-\frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^n = \frac{\lambda}{\mu-\lambda} \qquad (2)$$

Sob condição de equilíbrio, seja N_q o número total de usuários na fila, temos que seu valor esperado é:

$$L_q = E\left(N_q\right) = \sum_{n=1}^{\infty}\left(n-1\right)\pi_n = \sum_{n=1}^{\infty} n\pi_n - \sum_{n=1}^{\infty}\pi_n = \frac{\lambda}{\mu}\left(\frac{\lambda}{\mu-\lambda}\right) \qquad (3)$$

TIPO M/M/1

- **Medida de desempenho do sistema**

Relacionar o número médio de usuários na fila ou no sistema com o tempo médio de espera na fila denotadas por W e W_q respectivamente.

Essas relações, que foram apresentadas por Little em 1961, são dadas pelas seguintes expressões:

$$L = \lambda * W \quad \text{e} \quad L_q = \lambda * W_q$$

TIPO M/M/1

II.2. Fórmula de Little: a fórmula geral das filas

Considere *qualquer* sistema de filas em estado estacionário, onde:

- λ é taxa média de chegadas de clientes ao sistema [clientes/unidade de tempo];
- L é número médio de clientes no sistema (tanto em fila quanto em atendimento) [clientes];
- W é tempo médio de permanência de um cliente no sistema [unidades de tempo].

Neste caso, a *fórmula de Little* é definida como:

$$L = \lambda W$$

Essa fórmula também pode ser escrita em função do número esperado de clientes na fila (Lq) e do tempo médio de espera em fila (Wq) ou em função do número médio de clientes em atendimento (Ls) e do tempo médio de atendimento (Ws):

$$L_q = \lambda W_q$$

$$L_s = \lambda W_s$$

Um posto bancário recebe uma média de 30 clientes por hora. O tempo médio de permanência no banco (calculado entre a chegada e a partida de cada cliente) é igual a cinco minutos. Quantos clientes em média encontram-se no banco? Se cada cliente permanece no caixa de atendimento em média por um minuto, qual o número médio de clientes em fila?

Neste exemplo, a taxa de chegadas é de 30 clientes/h, ou seja:

$$\lambda = \frac{30}{60} = 0{,}5 \text{ cliente/min}$$

O tempo médio de espera no sistema é de cinco minutos, ou seja:

$$W = 5 \text{ min}$$

Para se determinar o número médio de elementos no banco, a aplicação da fórmula de Little é imediata:

$$L = \lambda W = 0{,}5 \times 5 = 2{,}5 \text{ clientes}$$

Se cada cliente permanece por um minuto no caixa, então:

$$W_s = 1 \text{ min}$$

TIPO M/M/1

ELSEVIER

E o número médio de clientes em atendimento é obtido diretamente da fórmula de Little:

$$L_s = \lambda W_s = 0,5 \times 1 = 0,5 \text{ cliente}$$

Como a média de clientes no sistema é a soma da média de clientes em fila com a média de clientes em atendimento, temos:

$$L = L_q + L_s \Rightarrow L_q = L - L_s = 2,5 - 0,5$$

$$\therefore L_q = 2,0 \text{ clientes em fila}$$

TIPO M/M/1

Parâmetro	Expressão
Índice de ocupação do sistema	$\rho = \frac{\lambda}{\mu}$
Probabilidade de o sistema estar livre	$p_0 = 1 - \rho$
Probabilidade de j elementos no sistema	$p_j = (1-\rho)\rho^j,\ j = 1,2,\ldots$
Probabilidade de mais do que k elementos no sistema	$P[\geq k \text{ no sistema}] = \rho^k$
Média de elementos em fila	$L_q = \frac{\rho^2}{1-\rho} = \frac{\lambda^2}{\mu(\mu-\lambda)}$
Média de elementos em atendimento	$L_s = \rho$
Média de elementos no sistema	$L = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda}$

TIPO M/M/1

- **Exemplo**

  Seja um aeroporto com uma única pista de pouso/decolagem. Os aviões chegam a uma taxa de 15/hora, e levam em média 3 minutos para aterrisar. Assumindo que as chegadas são um processo de Poisson, e o tempo de aterrisagem é distribuído por uma exponencial.

$$\lambda = 15/hora \quad \text{e} \quad \mu = \frac{60}{3}/hora = 20/hora$$

Intensidade de tráfego: $= \frac{\lambda}{\mu} = \frac{3}{4} = 0.75$

Número médio de aviões aguardando para pousar: $E(N_q) = \frac{\lambda}{\mu} * \frac{\lambda}{(\mu-\lambda)} = 2.25$

Tempo médio de espera para o pouso: $W_q = \frac{\lambda}{\mu(\mu-\lambda)} = \frac{3}{20} = 9$ minutos

TIPO M/M/c

- É uma extensão da fila $M/M/1$, apresentando a mesma distribuição de chegada, a mesma disciplina e a mesma distribuição do tempo de serviço. O que difere de ambas é a quantidade de servidores, agora considerado um sistema com c servidores.

- Agora as taxas de chegadas e saídas são dadas por:

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

$$\mu_n = \{n\mu, \quad 0 \leq n < c \quad \text{e} \quad c\mu, \quad n \geq c\}$$

TIPO M/M/c

- **Distribuição Estacionária**

Consideramos agora que a intensidade de tráfego é dada por: $\frac{\lambda}{c\mu}, \quad n \geq c$

Com o fato do estado se encontrar em equilíbrio, pode-se mostrar que

$$\pi_n = \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n \pi_0 \quad 0 \leq n < c; \quad \text{e} \quad \frac{1}{c!c^{n-c}}\left(\frac{\lambda}{\mu}\right)^n \pi_0 \quad n \geq c \qquad (4)$$

Nesta fila também temos que $\lambda < c\mu$, neste caso o processo também será recorrente positivo.

TIPO M/M/c

- **Valor esperado do número de clientes**

Sob condição de equilíbrio, o valor esperado do número de pessoas na fila é dada por:

$$E(N_q) = \sum_{n=c}^{\infty} (n-c)\,\pi_n = \frac{1}{(c-1)!}\left(\frac{\lambda}{\mu}\right)^c \pi_0 \frac{\lambda\mu}{(c\mu-\lambda)^2} \tag{5}$$

- **Medida de desempenho do sistema**

Utilizando a relação de Little, obtemos que

$$W_q = \frac{E(N_q)}{\lambda} = \frac{1}{(c-1)!}\left(\frac{\lambda}{\mu}\right)^c \pi_0 \frac{\mu}{(c\mu-\lambda)^2}$$

TIPO M/M/c

Parâmetro	Expressão
Taxa efetiva de entrada no sistema	$\lambda_c = \lambda(1 - p_c)$
Índice de ocupação do sistema	$\rho = \frac{\lambda}{\mu}$
Probabilidade do sistema estar livre	$\lambda \neq \mu$: $p_0 = \frac{1-\rho}{1-\rho^{c+1}}$ $\lambda = \mu$: $p_0 = \frac{1}{1+c}$

TIPO M/M/c

Parâmetro	Expressão
Probabilidade de j elementos no sistema	$\lambda \neq \mu$: $p_j = \rho^j p_0,\ j = 1,\ldots,c$; $p_j = 0,\quad j > c$
	$\lambda = \mu$: $p_j = \frac{1}{1+c} \quad j = 0,1,\ldots c$
Probabilidade de mais do que k elementos no sistema	$P[\geq k \text{ no sistema}] = \rho^k$
Média de elementos em fila	$L_q = L - L_s$
Média de elementos em atendimento	$L_s = 1 - p_0$
Média de elementos no sistema	$\lambda \neq \mu$: $L = \frac{\rho\left[1-(c+1)\rho^c + c\rho^{c+1}\right]}{\left(1-\rho^{c+1}\right)(1-\rho)}$
	$\lambda = \mu$: $L = \frac{c}{2}$

TIPO M/M/c

- **Exemplo**

  Reconsiderando o exemplo utilizado para o tipo de filas $M/M/1$, seja um aeroporto, agora com duas pistas de pouso/decolagem. Os aviões chegam a uma taxa de 15/hora, e levam em média 3 minutos para aterrisar. Assumindo que as chegadas são um processo de Poisson, e o tempo de aterrisagem seguem uma distribuíção exponencial.

  $\lambda = 15/hora, \quad \mu = \frac{60}{3}/hora = 20/hora$ e o número de servidores: $c = 2$

  Intensidade de tráfego: $= \frac{\lambda}{c\mu} = \frac{3}{8} = 0.375$

TIPO M/M/c

- **Exemplo (Continuação)**

Número médio de aviões aguardando para pousar:

$$E(N_q) = \frac{(\lambda/c\mu)(\lambda/\mu)^c \pi_0}{c!(1-\lambda/c\mu)^2} = \frac{(3/8) * (3/4)^2 0.4545}{2(5/8)^2} = 0.1227$$

Tempo médio de espera para o pouso:

$$W_q = \frac{(\lambda/\mu)^c \pi_0}{c!c\mu(1-\lambda/c\mu)^2} = \frac{3}{20} = \frac{(3/4)^2 0.4545}{2 * 2 * 20(5/8)^2} = 0.49, \quad \text{minutos}$$

ESTUDO DE CASO: CINEMA

Um grupo de empresários estão reorganizando o atendimento ao público em um cinema que eles abriram há 6 mêses.
Nesta primeira etapa da reorganização eles querem saber quantos caixas deveriam contratar para que maximizem seus ganhos, num total máximo de 6 atendentes, considerando:

- O cinema funciona diariamente no horário das 20:00-24:00 horas. O cinema tem 5 salas, com uma capacidade total de 720 pessoas, mas os proprietários sabem que todos os dias a quantidade máxima de pessoas que chega no local é 700.

ESTUDO DE CASO: CINEMA

- O número de chegadas segue aproximadamente uma distribuição Poisson, de parâmetro $\lambda = 2.50$ pessoas por minuto, e os tempos de atendimento aos clientes seguem uma distribuição exponencial com taxa $\mu = 1.5$ atendimentos por minuto.

- O custo por hora de cada atendente é de $R\$25.00$.
  Para o preço do ingresso foi considerado o seguinte esquema: Ao entrar na fila de atendimento, cada cliente recebe uma senha com a hora da entrada. Ao ser atendido, o preço do ingresso, por pessoa, será de $R\$20.00 - R\$0.25 * T$, onde T representa o tempo de espera na fila, em minutos.

REFERÊNCIAS

Modelagem e Simulação de eventos discretos – Teoria e aplicações.

Chwif e Medina, 4edição

Disponível na Biblioteca da Mauá