Teoria de Filas

Modelo de Fila M/M/S

Teoria de Filas

- Definição de L: Empregando a distribuição de estado estacionário dada e usando $\rho = \lambda/\mu$ é possível obter a média do número de clientes presentes no modelo de filas (L).
- Definição de Lq: O número esperado de clientes esperando atendimento (ou na fila) é Lq.
- Definição de Ls: O número esperado de clientes em atendimento é Ls.
- Definição de W, Wq e Ws: Define-se W como o tempo esperado gasto pelo cliente no sistema, incluindo o tempo na fila mais o tempo de atendimento. O tempo gasto na fila é Wq e o tempo em serviço é Ws.

Teoria de Filas

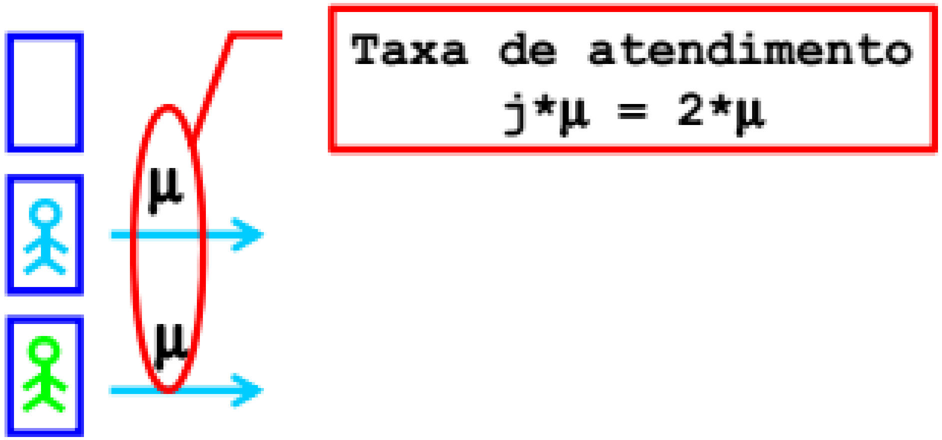
- Modelo de Fila M(1)/M(2)/s(3)/GD(4)/∞(5)/∞(6):
- Este modelo supõe:

1. Natureza do processo de chegada. Ex.: M – variáveis aleatórias iid como função de distribuição exponencial.
2. Natureza do processo de serviço. Ex.: M – variáveis aleatórias iid como função de distribuição exponencial.
3. Número de servidores em paralelo é s ao invés de 1.
4. Organização da fila: FCFS – Primeiro a entrar, primeiro a sair (por exemplo, First come/first served).
5. Número máximo de clientes no sistema (totalizando clientes na fila e em atendimento) não limitado
6. Tamanho da população de clientes não limitado.

Teoria de Filas

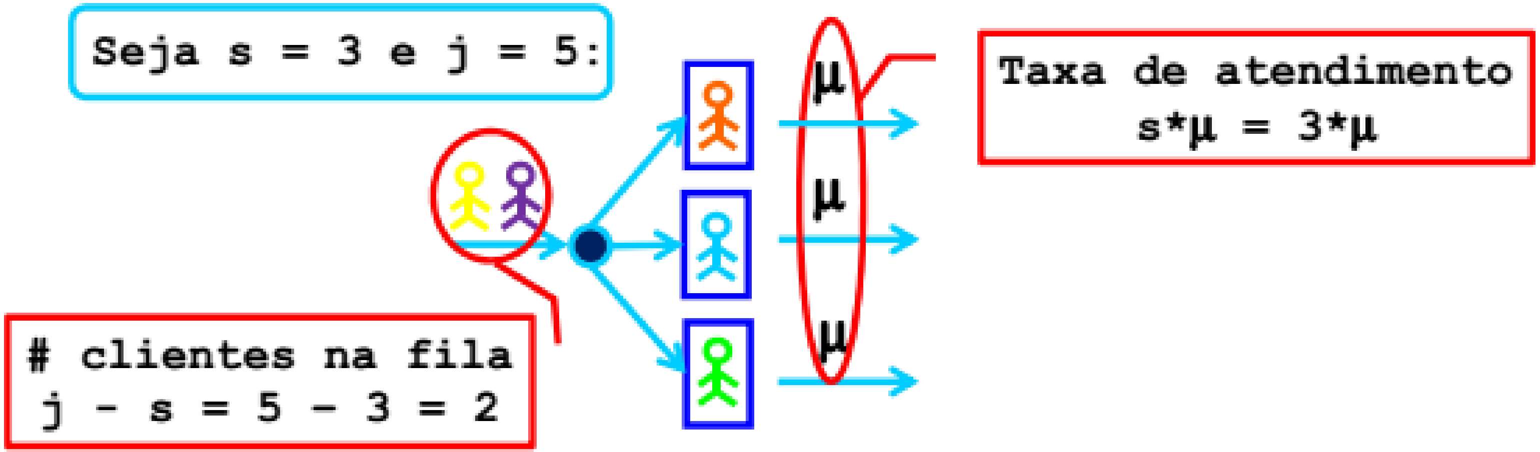
- Modelo de Fila M/M/s/GD/∞/∞: Neste modelo existem s servidores em paralelo tal que se existem j clientes dois casos podem ocorrer:
  - Caso 1: Se j ≤s Todos os clientes presentes estão em atendimento, pois o número de servidores é maior que o de clientes. Se j servidores estão ocupados a taxa de fim de serviço será j*µ.

Exemplo: Seja s = 3 e j = 2:



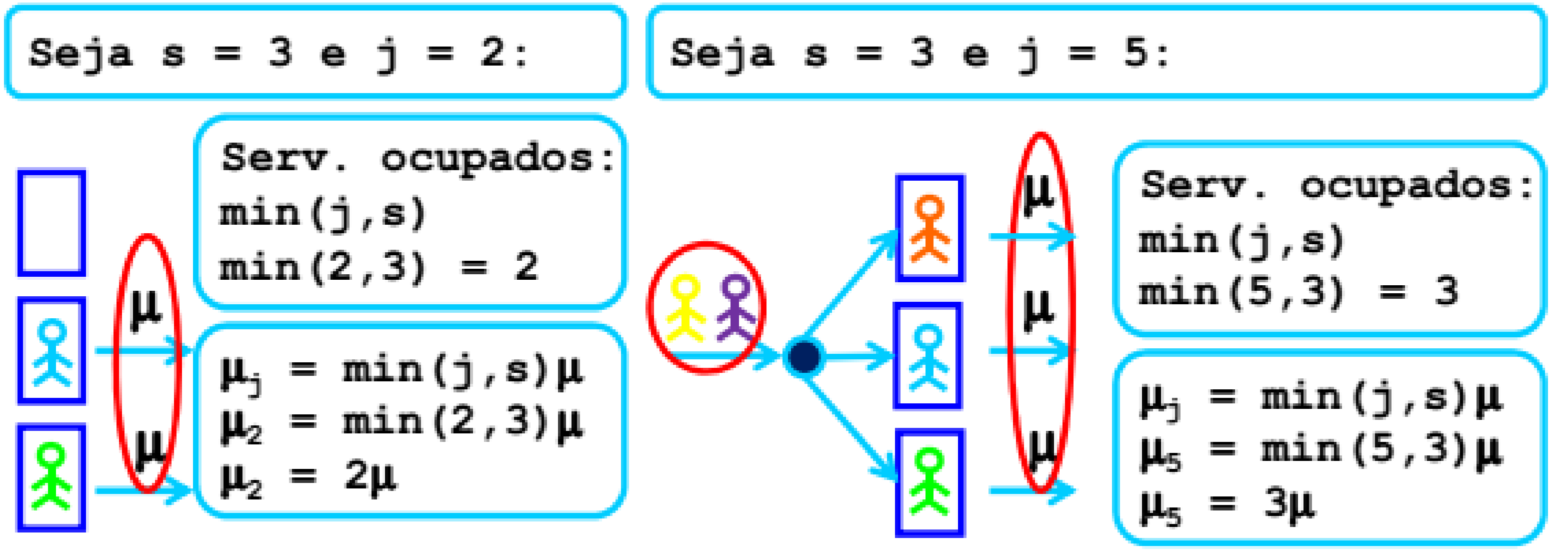
Teoria de Filas

- Modelo de Fila M/M/s/GD/∞/∞: Neste modelo existem s servidores em paralelo tal que se existem j clientes dois casos podem ocorrer:
  - Caso 2: Se j > s. Neste caso s servidores estão ocupados enquanto j-s clientes aguardam atendimento. A taxa de fim de serviço será s*μ.



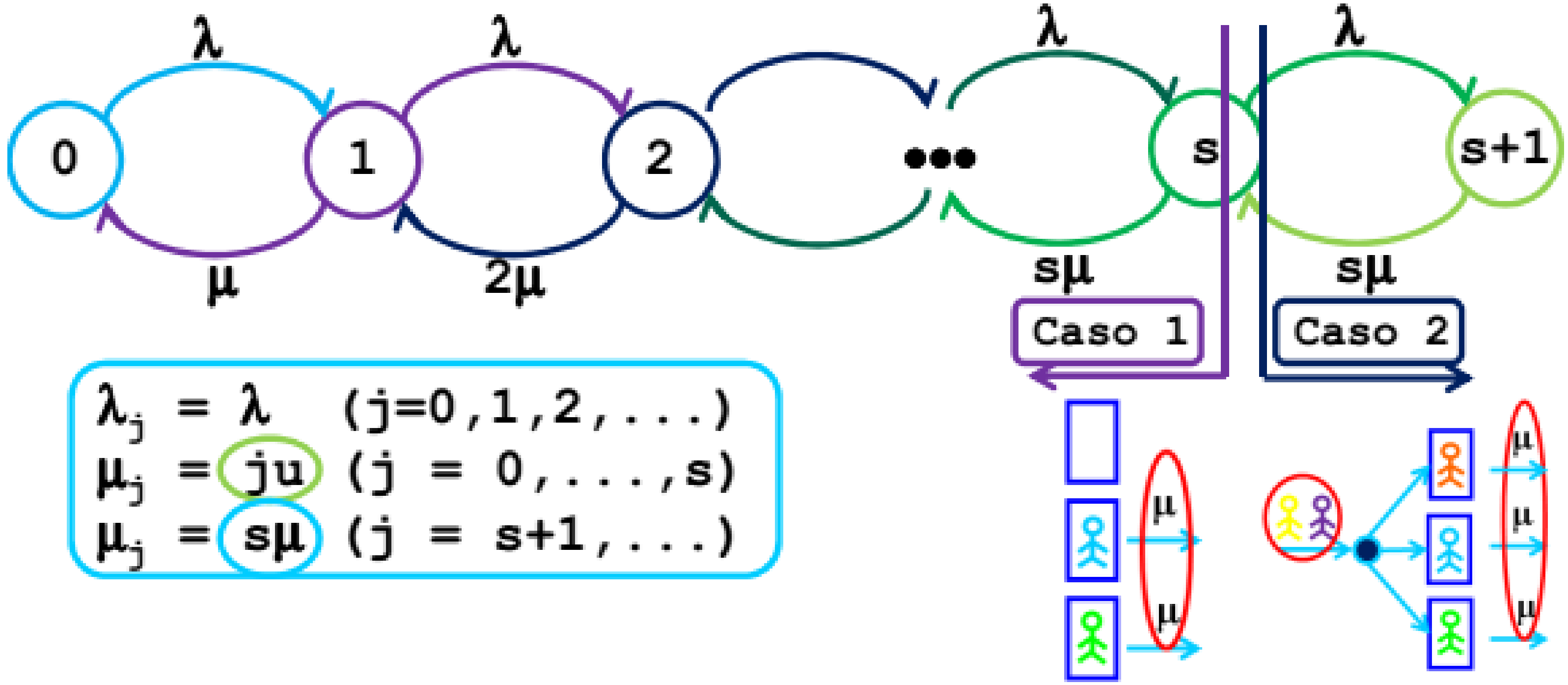
Teoria de Filas

- Modelo de Fila M/M/s/GD/∞/∞: Neste modelo existem s servidores em paralelo tal que se existem j clientes dois casos podem ocorrer:
  - Casos 1 e 2: Se j clientes estão presentes, então, min(j,s) servidores estarão ocupados e a taxa de atendimento será de $\mu_j = \min(j,s)\mu$.



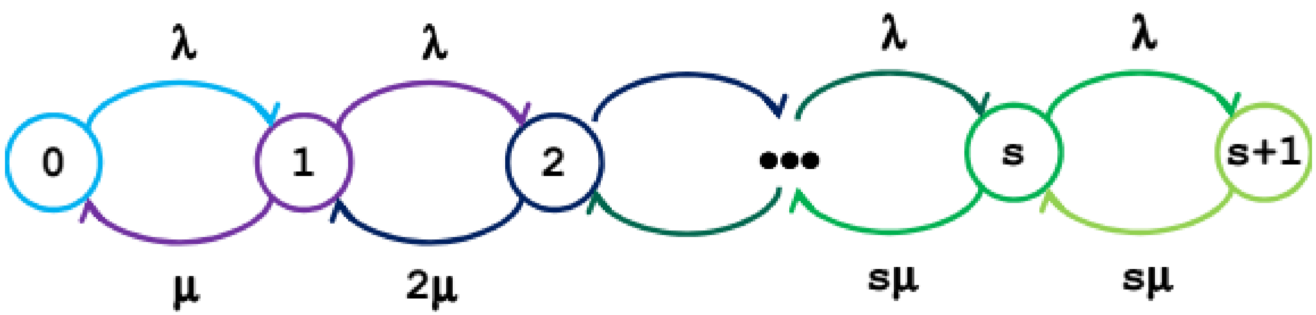
Teoria de Filas

- Modelo de Fila M/M/s/GD/∞/∞: os dois casos anteriores podem ser modelados por um processo de nascimento-morte tal como dado a seguir:



Teoria de Filas

- Modelo de Fila M/M/s/GD/∞/∞: Seja ρ= λ/sμ e ρ<1, então, para este modelo:



$$\pi_0 = \frac{1}{\sum_{i=0}^{s-1} \frac{(s\rho)^i}{i!} + \frac{(s\rho)^s}{s!(1-\rho)}}$$

$$\pi_j = \frac{(s\rho)^j \pi_0}{j!} \qquad \text{(j=1,...,s)}$$

Se ρ ≥ 1, então, não existe estado estacionário

$$\pi_j = \frac{(s\rho)^j \pi_0}{s! s^{j-s}} \qquad \text{(j=s,s+1,...)}$$

Teoria de Filas

- A probabilidade de estado estacionário de que todos os servidores estejam ocupados é dada por:

$$P(j \geq s) = \frac{(s\rho)^s \pi_0}{s!(1-\rho)}$$

Pode ser mostrado também que:

$$L_q = \frac{P(j \geq s)\rho}{1-\rho} \qquad W_q = \frac{L_q}{\lambda} = \frac{P(j \geq s)}{s\mu - \lambda}$$

Para se obter L, usa-se L = Lq + Ls e que Ls = λ/μ:

$$L = L_q + \frac{\lambda}{\mu} \qquad W = \frac{L}{\lambda} = \frac{L_q}{\lambda} + \frac{1}{\mu} = W_q + \frac{1}{\mu} = \boxed{\frac{P(j \geq s)}{s\mu - \lambda} + \frac{1}{\mu}}$$

Teoria de Filas

- Exemplo : Considere um banco com dois atendentes. Um média de 80 clientes por hora chegam ao banco e esperam em uma única fila por um caixa vazio. O tempo médio de atendimento de um cliente é de 1,2 minutos. Assumindo que o tempo entre as chegadas e o tempo de serviços são exponenciais, determinar:

(A)A fração de tempo que um servidor está vazio.

(B)O número esperado de clientes no banco

(C)O tempo médio de espera que um cliente gasta no banco

Teoria de Filas

(A)A fração de tempo que um servidor está vazio.

Para determinar a fração de tempo que um servidor em particular está ocioso é necessário observar que quando j = 0 o servidor está totalmente ocioso, mas quando j = 1, somente 50% do tempo um servidor estará ocioso (são dois), isto é:

Tempo ocioso de 1 servidor = $\pi_0 + 0{,}5\pi_1$

Seja $\rho = \lambda/s\mu = 80/2*50 = 0{,}8$. Calculando π_0, e depois π_1:

$$\pi_0 = \frac{1}{\sum_{i=0}^{s-1}\frac{(s\rho)^i}{i!} + \frac{(s\rho)^s}{s!(1-\rho)}} = \frac{1}{\frac{(2*0{,}8)^0}{0!} + \frac{(2*0{,}8)^1}{1!} + \frac{(2*0{,}8)^2}{2!(1-0{,}8)}}$$

$$\pi_0 = \frac{1}{1+1{,}6+6{,}4} = \frac{1}{9}$$

Teoria de Filas

(A)A fração de tempo que um servidor está vazio.

Sabendo-se que π_0 = 1/9 calcula-se π_1:

$$\pi_j = \frac{(s\rho)^j \pi_0}{j!} \quad \texttt{(j=1,...,s)}$$

$$\pi_1 = \frac{(2*0,8)^1 0,11}{1!} = 0,176$$

A fração de tempo que 1 servidor estará ocioso é:

Tempo ocioso de 1 servidor = $\pi_0 + 0,5\pi_1$ = 0,11 + 0,088 = 0,198

Teoria de Filas

(B) O número esperado de clientes no banco.

Sejam λ= 80 clientes por hora e μ= 50 clientes por hora. Então: ρ= 80/(2*50) = 0,80 < 1 e o estado estacionário existe. Calcula-se P(j ≥s) = P(j ≥2). Então:

$$P(j \geq s) = \frac{(s\rho)^s \pi_0}{s!(1-\rho)}$$

$$P(j \geq 2) = \frac{(2*0{,}8)^2 *0{,}11}{2!(1-0{,}8)} = 0{,}71$$

Teoria de Filas

(B) O número esperado de clientes no banco.

Se P(j ≥2) = 0,71. Então:

$$L_q = \frac{P(j \geq s)\rho}{1-\rho} = \frac{0{,}71 * 0{,}80}{1-0{,}80} = 2{,}84$$

Aplicando a Equação relativa a L:

$$L = L_q + \frac{\lambda}{\mu} = 2{,}84 + \frac{80}{50} = 4{,}44$$

Teoria de Filas

(C)O tempo médio de espera que um cliente gasta no banco.

$$W = \frac{L}{\lambda} = \frac{4{,}44}{80} = 0{,}055$$

Teoria de Filas

<u>Exercício</u>: O mesmo banco do exercício anterior sabe que no início do mês a taxa média de clientes por hora passa de 80 para 95. Sabendo-se que o atendimento a um cliente não deve demorar mais que 20 minutos será necessário aumentar o número de atendentes para 3 ou mais?

(A) Será necessário calcular o tempo médio de espera que um cliente gasta no banco. Se este for maior que 20 minutos, então, verificar se 3 fornece um tempo menor, senão 4 e assim por diante.

Teoria de Filas

Algumas equações para os cálculos: $\rho = \lambda / s\mu$

Se $\rho \geq 1$, então, não existe estado estacionário.

$$\pi_0 = \frac{1}{\sum_{i=0}^{s-1} \frac{(s\rho)^i}{i!} + \frac{(s\rho)^s}{s!(1-\rho)}}$$

$$\pi_j = \frac{(s\rho)^j \pi_0}{j!} \quad \texttt{(j=1,...,s)}$$

$$P(j \geq s) = \frac{(s\rho)^s \pi_0}{s!(1-\rho)}$$

$$L_q = \frac{P(j \geq s)\rho}{1-\rho}$$

$$L = L_q + \frac{\lambda}{\mu}$$

$$W = \frac{L}{\lambda}$$

Teoria de Filas

O tempo no banco é menor que 20 minutos, se λ= 95?

Sejam λ= 95 clientes por hora e μ= 50 clientes por hora e

ρ= $\lambda/s\mu \rightarrow$ ρ=95/(2*50) = 0,95. Então:

$$\pi_0 = \frac{1}{\sum_{i=0}^{s-1}\frac{(s\rho)^i}{i!} + \frac{(s\rho)^s}{s!(1-\rho)}} = \frac{1}{\frac{(2*0{,}95)^0}{0!} + \frac{(2*0{,}95)^1}{1!} + \frac{(2*0{,}95)^2}{2!(1-0{,}95)}}$$

$$\pi_0 = \frac{1}{1+1{,}9+36{,}1} = \frac{1}{39} = 2{,}5641\%$$

Teoria de Filas

O tempo no banco é menor que 20 minutos, se λ= 95?

Sabendo-se que π_0 = 1/39, calcula-se P(j≥s):

$$P(j \geq s) = \frac{(s\rho)^s \pi_0}{s!(1-\rho)}$$

$$P(j \geq 2) = \frac{(2*0{,}95)^2 * (1/39)}{2!(1-0{,}95)} = 0{,}9256$$

$$L_q = \frac{P(j \geq s)\rho}{1-\rho} = \frac{0{,}9256 * 0{,}95}{1-0{,}95} = 17{,}5864$$

$$L = L_q + \frac{\lambda}{\mu} = 17{,}58 + \frac{95}{50} = 19{,}48$$

$$W = \frac{L}{\lambda} = \frac{19{,}48}{95} = 0{,}2051$$

Teoria de Filas

Exercício: O gerente de um banco deve determinar quantos atendentes devem trabalhar na Sexta. Cada minuto que um cliente permanece na fila, o gerente acredita que custa R$ 0,05. Em média 2 clientes por minuto chegam ao banco. Em média são necessários 2 minutos para o atendente completar o serviço. O tempo entre as chegadas e o de serviço são exponenciais. O custo de um atendente por hora é de R$ 9. Para minimizar a soma dos custos de serviço e os custos de atraso, quantos atendentes deverão trabalhar na sexta?

Teoria de Filas

Exercício: Para qual valor do custo de permanência por minuto na fila passa a ser vantajoso a contratação de 6 atendentes (no exemplo c = 0,05)? Lembrando que λ= 2 clientes por minuto e μ= 0,5 clientes porminuto.