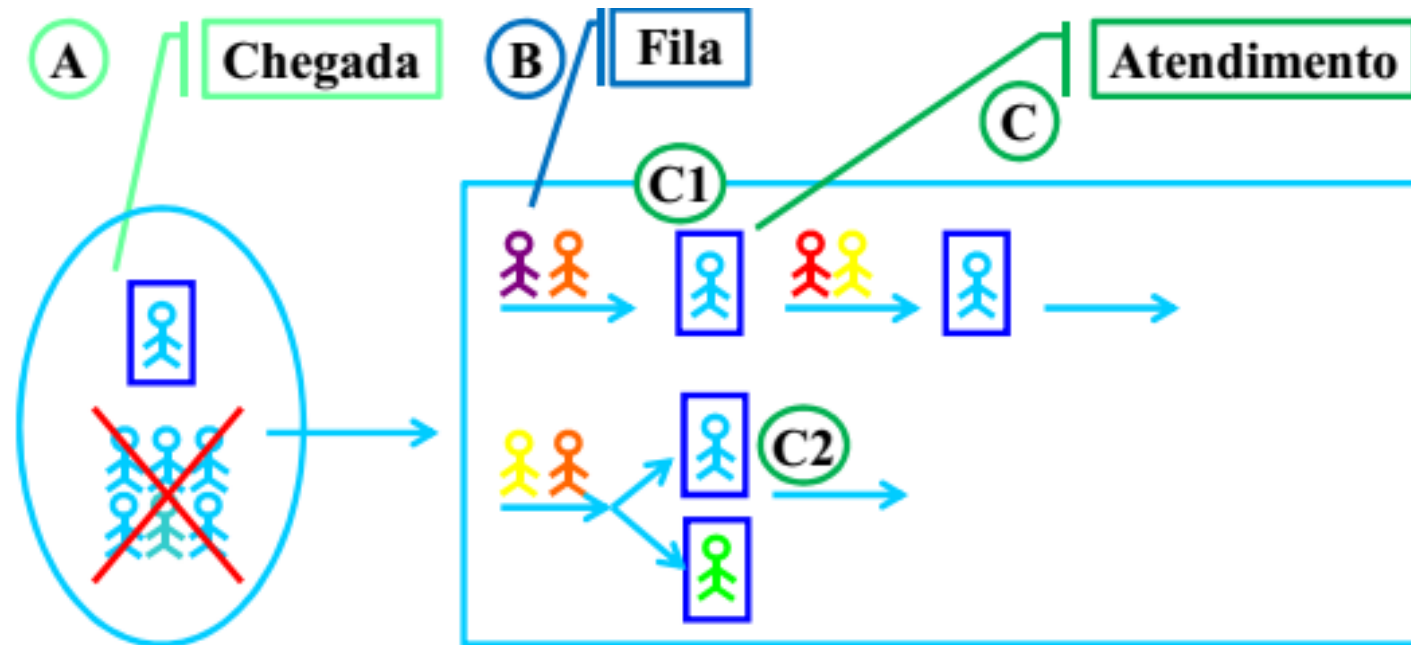


# Teoria de Filas

# Teoria de Filas



- A) Supõe a chegada de apenas um usuário por instante. O tempo para ocorrer uma chegada é função de probabilidade.
- B) Organização da fila: primeiro a entrar, primeiro a sair
- C) Pode ter servidores em paralelo (C2) ou em série (C1). O tempo de serviço é associado a função de distribuição.

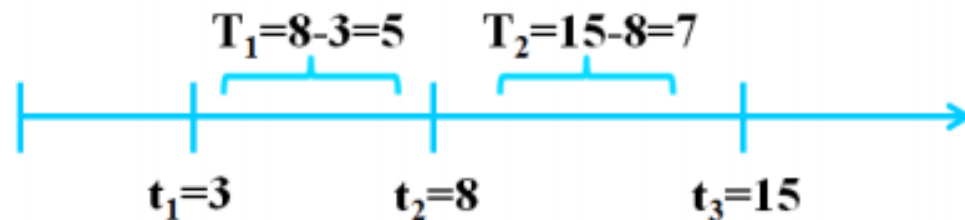
# Teoria de Filas

Notação de Kendall-Lee (1)/(2)/(3)/(4)/(5)/(6) em que:

(1) representa a natureza do processo de chegada. Ex.: M – variáveis aleatórias iid como função de distribuição exponencial

Seja  $t_i$  o tempo que o  $i$ -ésimo cliente chega e

$T_i = t_{i+1} - t_i$  o  $i$ -ésimo intervalo entre chegadas tal como abaixo:



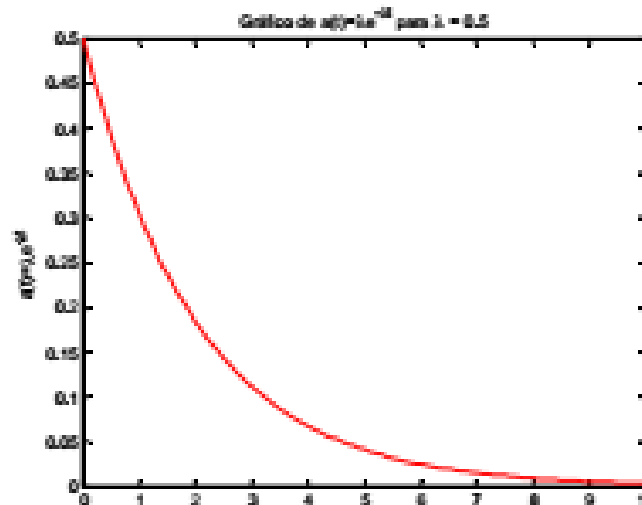
Na modelagem do processo de chegada é suposto que as  $T_i$ s são variáveis aleatórias contínuas e independentes descritas pela variável aleatória A. A independência diz que o valor de  $T_2$  não afeta o valor de  $T_3$ , e assim por diante.

# Teoria de Filas

Na maioria das aplicações uma questão importante é escolher A de modo a refletir a realidade, mas ainda ser computacionalmente tratável. A escolha mais comum para A é a função de distribuição exponencial, ou seja:

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

onde:  $\lambda$ - é a taxa de chegada  
(no. chegadas/hora,  
p.ex.).



O principal motivo para escolha da exponencial é devido a propriedade de que ela não tem memória. Ou seja:

$$P(A > t + h \mid A \geq t) = P(A > h)$$

# Teoria de Filas

Se o tempo entre as chegadas é uma exponencial, a função de distribuição de probabilidade do número de chegadas que ocorrem em qualquer intervalo de tempo de tamanho  $T$  pode ser obtido através do seguinte Teorema:

**Teorema 1:** O tempo entre as chegadas é exponencial com parâmetro  $\lambda$  se e somente se o número de chegadas que ocorrem em um intervalo de tamanho  $t$  seguir uma função de distribuição de probabilidade Poisson com parâmetro  $\lambda t$ .

Uma variável aleatória discreta  $N$  segue a distribuição de Poisson com parâmetro  $\lambda$  se, para  $n = 0, 1, 2, \dots$

$$P(N = n) = \frac{e^{-\lambda} \lambda^n}{n!}$$

# Teoria de Filas

Se  $N$  é uma variável aleatória de Poisson, então, pode ser provado que  **$E(N) = \text{var}(N) = \lambda$** . Definindo que  $N_t$  é o número de chegadas que ocorrem durante o intervalo de tamanho  $t$ , então, o Teorema 1 diz que:

$$P(N_t = n) = \frac{e^{-\lambda t} (\lambda t)^n}{n!} \quad (n = 0, 1, 2, \dots)$$

Se  $N_t$  é Poisson com parâmetro  $\lambda t$ , então, ocorre uma média de  $\lambda t$  chegadas durante o intervalo de tempo de tamanho  $t$  e o **valor  $\lambda$  pode ser visto como o número médio de chegadas por unidade de tempo, ou a taxa de chegada.**

# Teoria de Filas

**Teorema 2:** Para que  $N_t$  seja uma Poisson com parâmetro  $\lambda t$  e o tempo entre as chegadas seja uma exponencial com parâmetro  $\lambda$ , isto é,  $a(t) = \lambda e^{-\lambda t}$ , as seguintes condições devem ser satisfeitas:

- (1) As chegadas definidas em intervalos de tempo que não se sobrepõem são independentes (p.ex. o número de chegadas entre 1 e 10 não fornece nenhuma informação para o intervalo de 30 à 50).
- (2) Para  $\Delta t$  pequeno a probabilidade de que ocorra uma chegada entre  $t$  e  $t+\Delta t$  é dada por:

$$\lim_{\Delta t \rightarrow 0} O(\Delta t)/\Delta t = 0$$

Assim, a probabilidade de que mais de uma chegada ocorra entre  $t$  e  $t+\Delta t$  é  $O(\Delta t)$ .

O que as condições anteriores dizem é que não pode ocorrer chegada em massa.

# Teoria de Filas

Exemplo: O número pedido de copos de cerveja por hora no bar do Zé segue uma distribuição de Poisson com média de 30 pedidos por hora.

(A) Encontrar a probabilidade de que exatamente 60 cervejas sejam pedidas entre 22:00 e 24:00.



# Teoria de Filas

Exemplo: O número pedido de copos de cerveja por hora no bar do Zé segue uma distribuição de Poisson com média de 30 pedidos por hora.

(A) Encontrar a probabilidade de que exatamente 60 cervejas sejam pedidas entre 22:00 e 24:00.

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

Achar Poisson com parâmetro:  
 $\lambda * t = 30 * 2 = 60$

$$P(N_t = 60) = \frac{e^{-60} (60)^{60}}{60!} = 0.051$$

# Teoria de Filas

B) Encontrar a média e o desvio padrão do número de pedidos entre 21:00 e 1:00.

Se  $\lambda = 30$  pedidos por hora e  $t = 4$  horas, então a média é de 120 pedidos. O desvio padrão é:

$$(120)^{1/2} = 10.98$$

(C) Encontrar a probabilidade de que o tempo entre dois pedidos esteja entre 1 e 3 minutos.

Seja  $X$  o tempo (em minutos) entre pedidos sucessivos. A média de pedidos por minuto é exponencial com parâmetro  $30/60 = 0,5$ . Assim, a função de probabilidade do tempo entre dois pedidos é:

$0,5e^{-0,5t}$ . Então:

$$P(1 \leq X \leq 3) = \int_1^3 (0.5e^{-0.5t}) dt = e^{-0.5} - e^{-1.5} = 0.38$$

lembrando que:  $\int e^{at} dt = \frac{1}{a} e^{at} + C$

# Teoria de Filas

Exercício: O número pedido de copos de cerveja por hora no bar do Zé segue uma distribuição de Poisson com média de 30 pedidos por hora.

(A) Encontrar a probabilidade de que exatamente 61 cervejas sejam pedidas entre 22:00 e 24:00.

# Teoria de Filas

Exemplo: O número pedido de copos de cerveja por hora no bar do Zé segue uma distribuição de Poisson com média de 30 pedidos por hora.

(A) Encontrar a probabilidade de que exatamente 61 cervejas sejam pedidas entre 22:00 e 24:00.

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

Achar Poisson com parâmetro:  
 $\lambda * t = 2 * 30 = 60$

$$P(N_t = 61) = \frac{e^{-60} (60)^{61}}{61!} = 0.018$$

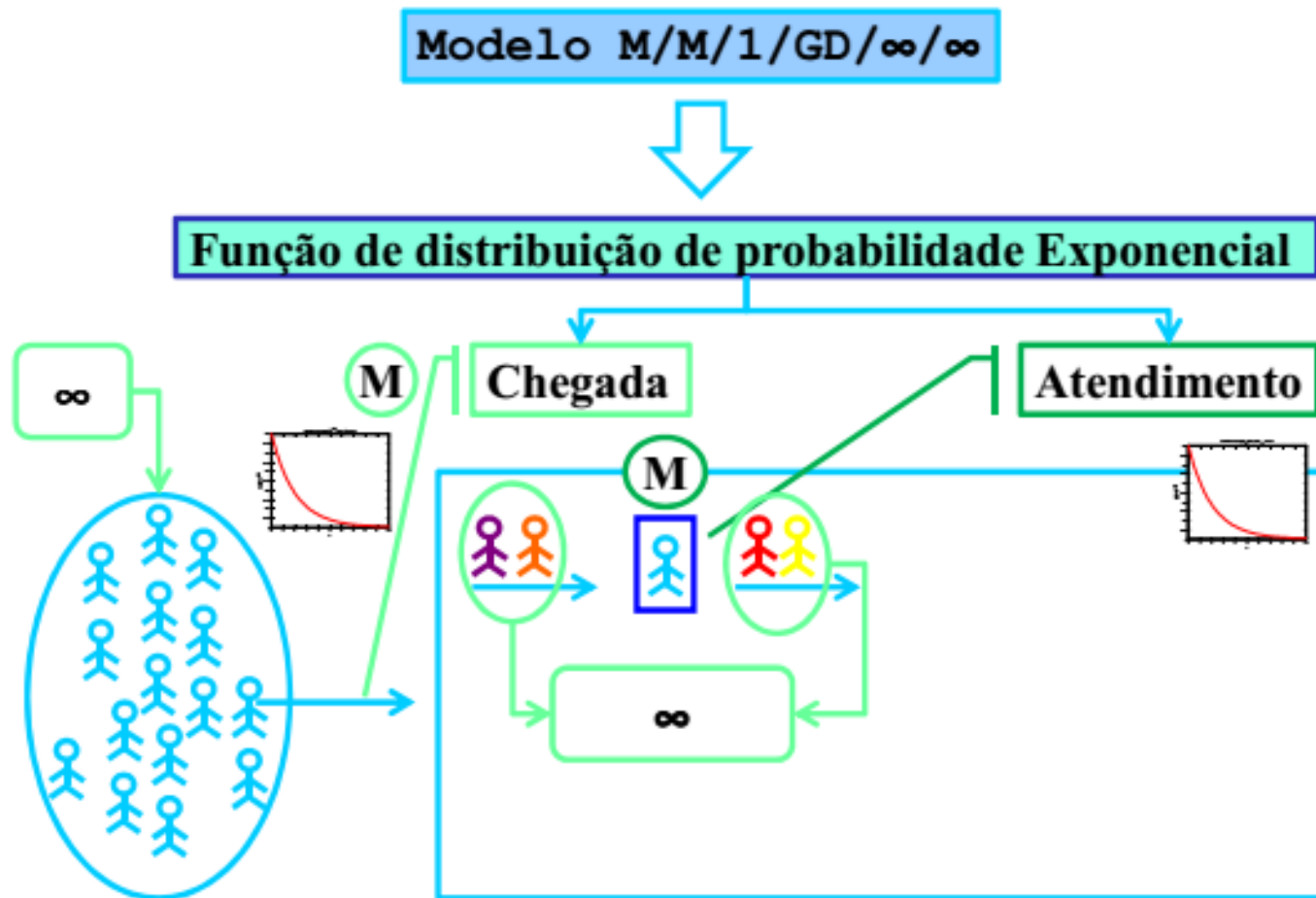
# Teoria de Filas: Resumo

A maioria dos sistemas de filas com tempos exponenciais de intervalos de chegadas e tempos de serviço podem ser modelados por um processo de nascimento e morte.



A propriedade de ausência de memória da exponencial garante que a probabilidade de ocorrer uma chegada no intervalo  $[t, t+\Delta t]$  não depende de há quanto tempo o sistema está no estado  $j$  e só depende da taxa  $\lambda$ . Idem p/ serviço.

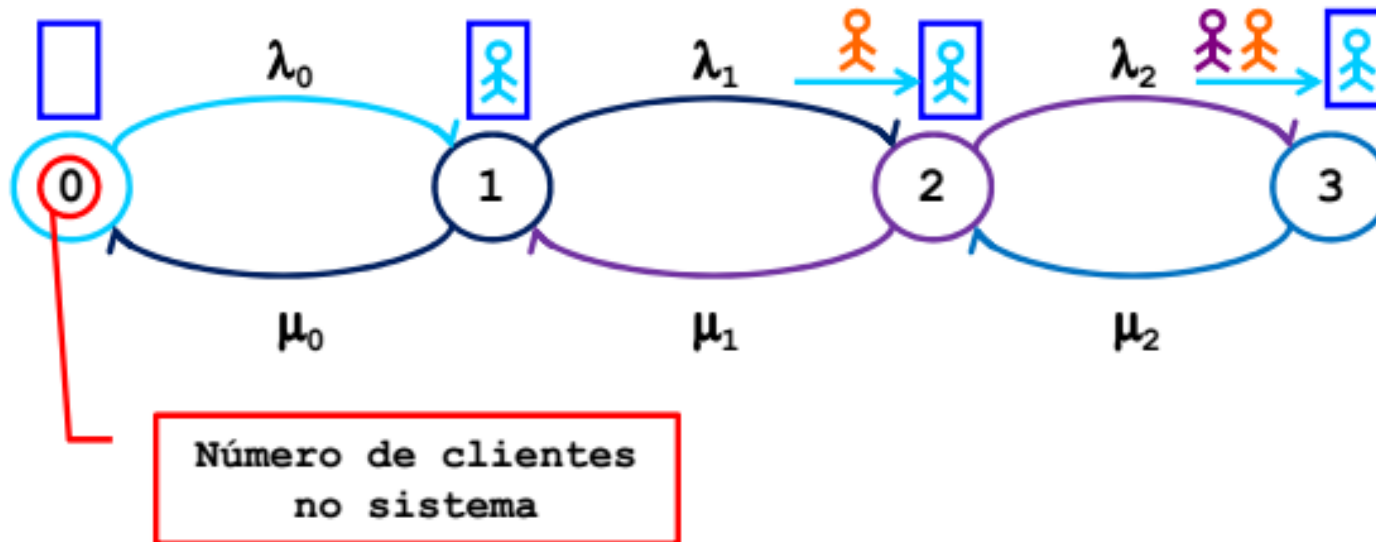
# Teoria de Filas



# Teoria de Filas

Modelo de Fila M/M/1/GD/ $\infty/\infty$ :

A metodologia de nascimento-morte pode ser empregada para analisar as propriedades de um modelo de fila com intervalo exponencial entre as chegadas (taxa  $\lambda$ ) e um único servidor com tempo de atendimento exponencial (taxa  $\mu$ ). Assim:



# Teoria de Filas

Modelo de Fila M/M/1/GD/ $\infty/\infty$ :

A metodologia de nascimento-morte pode ser empregada para analisar as propriedades de um modelo de fila com intervalo exponencial entre as chegadas (taxa  $\lambda$ ) e um único servidor com tempo de atendimento exponencial (taxa  $\mu$ ). Assim:

$$\lambda_j = \lambda, (j=0, 1, 2, \dots)$$

$$\pi_i = \pi_{i-1} \lambda_{i-1} / \mu_i$$

$$\mu_0 = 0 \quad \mu_j = \mu, (j=0, 1, 2, \dots)$$

Então, se:

$$\pi_1 = \lambda_0 \pi_0 / \mu_1 \Rightarrow \pi_1 = \lambda \pi_0 / \mu$$

$$\pi_2 = \lambda_1 \pi_1 / \mu_2 \Rightarrow \pi_2 = \lambda \pi_1 / \mu \Rightarrow \pi_2 = \lambda^2 \pi_0 / \mu^2$$



# Teoria de Filas

Logo:

$$\pi_j = \lambda^j \pi_0 / \mu^j$$

(7)

Lembrar:

$$\sum_{j=0}^{\infty} \pi_j = 1 \quad (3)$$

Definindo que a intensidade de tráfego do sistema de filas é dada por:  $\rho = \lambda/\mu$ . Então, aplicando (7) em (3):

$$\sum_{j=0}^{\infty} \pi_j = 1 \Rightarrow \pi_0 (1 + \rho + \rho^2 + \dots) = 1 \quad (8)$$

Assumindo que  $0 \leq \rho < 1$ . Então, calcula-se a soma S:

$$S = \begin{aligned} & s = 1 + \rho + \rho^2 + \dots & \rho s &= \rho + \rho^2 + \rho^3 + \dots \\ & s - \rho s = 1 \Rightarrow s = 1/(1 - \rho) \end{aligned} \quad (9)$$

# Teoria de Filas

Aplicando (9) em (8):

$$\pi_0 (1 + \rho + \rho^2 + \dots) = 1 \Rightarrow \pi_0 / (1 - \rho) = 1$$
$$\pi_0 = (1 - \rho) \quad (10)$$

Aplicando (10) em (7):

$$\pi_j = \lambda^j \pi_0 / \mu^j \Rightarrow \pi_j = \lambda^j (1 - \rho) / \mu^j \Rightarrow$$
$$\pi_j = \rho^j (1 - \rho) \quad (11)$$

É importante observar que **se  $\rho \geq 1$** , a soma dada em (8) irá crescer tanto quanto se queira e **não há estado estacionário**. Assim, se a taxa de chegada é maior ou igual que a taxa de atendimento ( $\lambda \geq \mu$ ) não existe estado estacionário.

# Teoria de Filas

A afirmação anterior pode ser verificada para um exemplo simples. Imagine que chegam  $\lambda = 6$  clientes por hora e que  $\mu = 4$  clientes são atendidos por hora. Mesmo que todos os servidores trabalhem o tempo todo ainda assim o número de clientes irá, em média, aumentar  $6 - 4 = 2$  clientes por hora. Ou seja, o número de clientes “explode” e não existe estado estacionário.

# Teoria de Filas

**Derivação de L:** Empregando a distribuição de estado estacionário dada por (11) é possível obter, por exemplo, a média do número de clientes presentes no modelo de filas (L):

$$L = \sum_{j=0}^{\infty} j \pi_j = \sum_{j=0}^{\infty} j \rho^j (1 - \rho) = (1 - \rho) \sum_{j=0}^{\infty} j \rho^j$$

Tendo que:

$$S' = \sum_{j=0}^{\infty} j \rho^j = \rho + 2\rho^2 + 3\rho^3 + \dots \quad \text{e} \quad \rho S' = \rho^2 + 2\rho^3 + 3\rho^4 + \dots$$

Então:

$$S' - \rho S' = \rho + \rho^2 + \dots = \rho / (1 - \rho)$$

$$S' = 1 / (1 - \rho)^2$$

# Teoria de Filas

**Derivação de L:** Empregando a distribuição de estado estacionário dada por (11) é possível obter, por exemplo, a média do número de clientes presentes no modelo de filas (L):

$$L = (1 - \rho) \sum_{j=0}^{\infty} j \rho^j = (1 - \rho) \frac{\rho}{(1 - \rho)^2} = \frac{\rho}{(1 - \rho)} = \frac{\lambda}{\mu - \lambda} \quad (12)$$

# Teoria de Filas

**Derivação de  $L_q$ :** Para calcular o número esperado de clientes esperando atendimento (ou na fila) ( $L_q$ ) é necessário observar que se o sistema tem 0 ou 1 clientes, então, não existe fila. Se existirem  $j$  ( $j \geq 1$ ) clientes, então,  $j-1$  clientes estarão esperando atendimento. Assim:

The diagram illustrates the derivation of the expected number of customers in the queue ( $L_q$ ) from the total expected number of customers in the system ( $L$ ). The equation  $L_q = \sum_{j=1}^{\infty} (j-1)\pi_j = \sum_{j=1}^{\infty} j\pi_j - \sum_{j=1}^{\infty} \pi_j$  is shown. The first sum is identified as  $L$  using equation (12). The second sum is identified as  $(1-\pi_0)$  using equation (8). The probability  $\pi_0$  is also shown as  $\sum_{j=0}^{\infty} \pi_j = 1$ .

$$L_q = \sum_{j=1}^{\infty} (j-1)\pi_j = \sum_{j=1}^{\infty} j\pi_j - \sum_{j=1}^{\infty} \pi_j$$

Usando (12) :  $L$

Usando (8) :  $(1-\pi_0)$

$$\sum_{j=0}^{\infty} \pi_j = 1$$

$$L_q = L - \rho = \frac{\rho^2}{1-\rho} \quad (13)$$

# Teoria de Filas

**Derivação de  $L_s$ :** Para calcular o número esperado de clientes em atendimento ( $L_s$ ) em um modelo de filas M/M/1/GD/ $\infty/\infty$  usa-se:

$$L_s = 0\pi_0 + 1(\pi_1 + \pi_2 + \dots) = 1 - \pi_0$$

Usando (8)

Assim:

$$L_s = 1 - (1 - \rho) = \rho \quad (14)$$

# Teoria de Filas

**A fórmula de Little para Filas:** Para calcular o tempo total que um cliente gasta no sistema, 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  **$Wq$** . **Tanto  $W$  como  $Wq$  assumem que o estado estacionário foi atingido.** A fórmula de Little relaciona  **$W$**  e  **$Wq$**  como  **$L$**  e  **$Lq$** , respectivamente:

$$L = \lambda W \quad (15)$$

$$Lq = \lambda Wq \quad (16)$$

$$Ls = \lambda Ws \quad (17)$$