Sistemas de Filas: Aula 1

Amedeo R. Odoni

10 de outubro de 2001

Tópicos em Teoria das Filas

9. Introdução a sistemas de filas; lei de Little, M/M/1
10. Filas Markovianas (processo de renovação)
11. Fila M/G/1 e extensões
12. Filas com prioridade de atendimento; representação de estados do sistema
13. Avaliação financeira de congestionamentos
14. Comportamento dinâmico de filas
15. Modelo hipercubo de filas
16. Inferência sobre sistemas de filas; psicologia de sistemas de filas

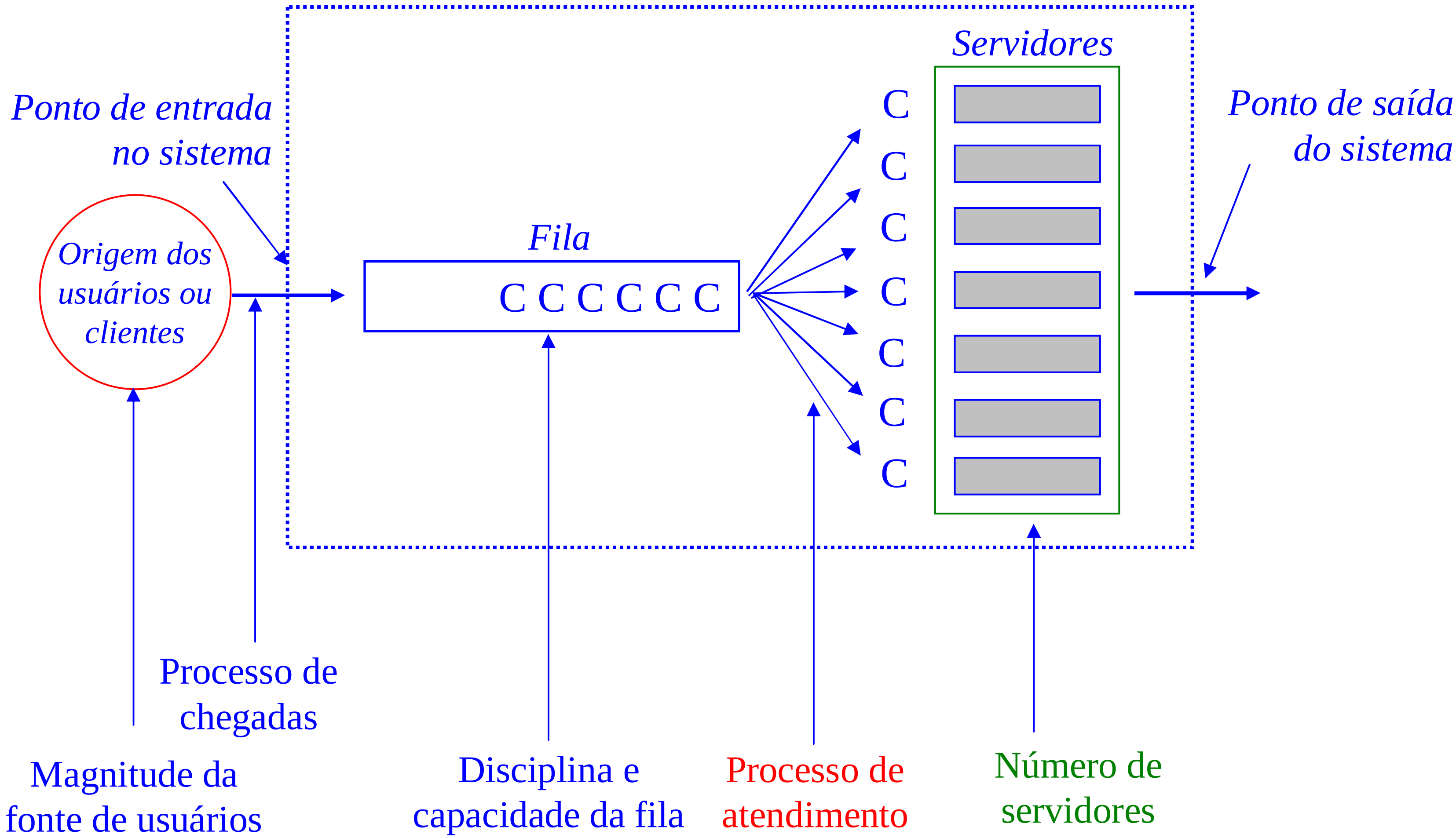
Roteiro da Apresentação

- Introdução a sistemas de filas
- Representação conceitual de sistemas de filas
- Nomenclatura de modelos de filas
- Terminologias e notações
- Lei de Little e relações básicas
- Processos de renovação
- Sistemas de filas M/M/1
- Diagramas de transição entre estados
- Probabilidades para o sistema em equilíbrio

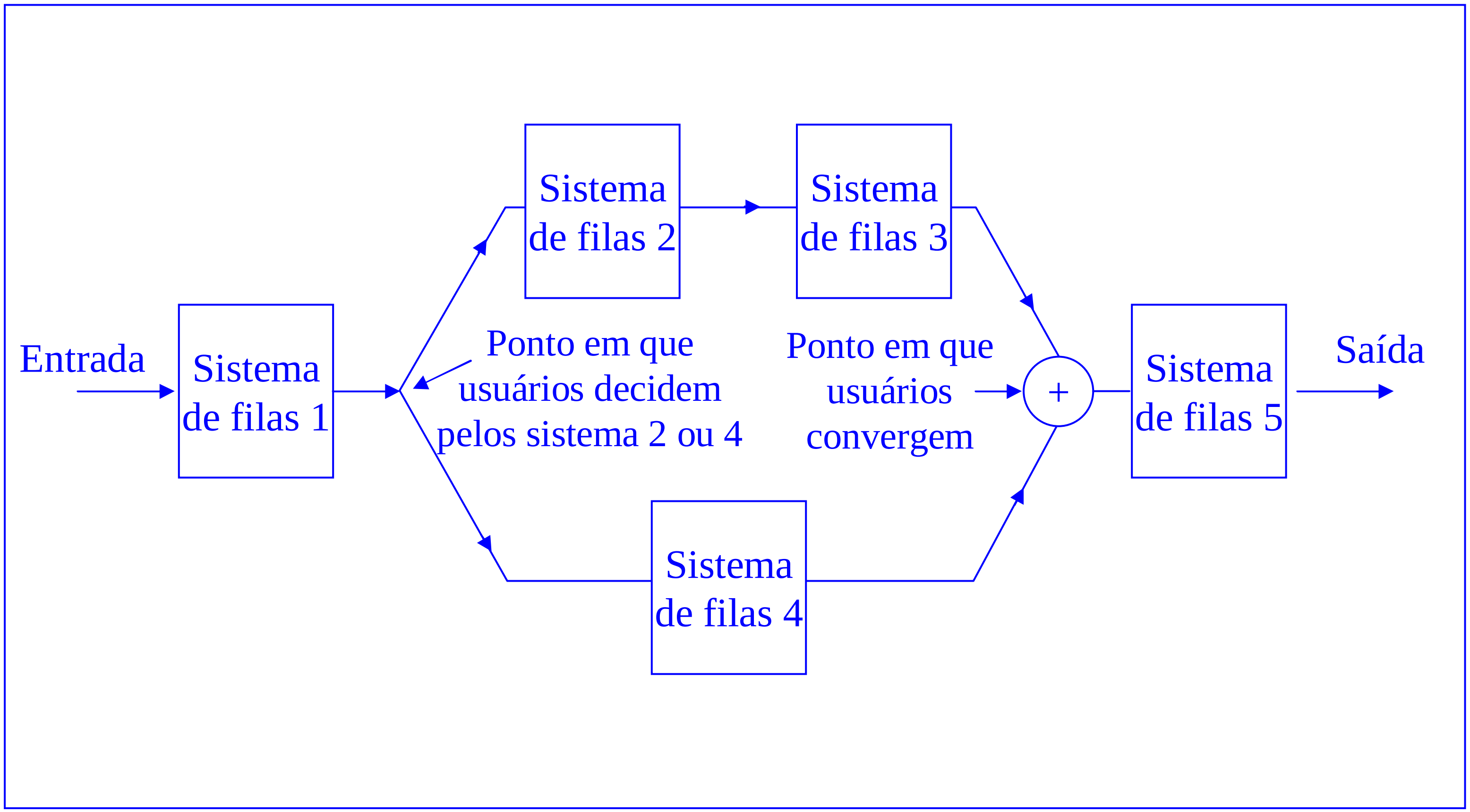
Filas

- A Teoria das Filas é a área da Pesquisa Operacional que trata de sistemas de filas (atrasos ou congestionamentos)
- Um sistema de filas é formado por uma demanda de elementos, uma fila e um ou mais canais de atendimento
- Uma rede de filas é formada por vários sub-sistemas de filas interligados
- Parâmetros fundamentais de um sistema de filas:
  - Taxa de chegadas
  - Capacidade (taxa de atendimento)
  - Tempo entre chegadas sucessivas
  - Tempos de atendimento
  - Capacidade da fila (“finito” vs. “infinita”)
  - Disciplina (FIFO/FCFS, SIRO, LIFO, prioridades)
  - Outros fatores (efeitos de “feedback”, etc.)

Um sistema genérico de filas



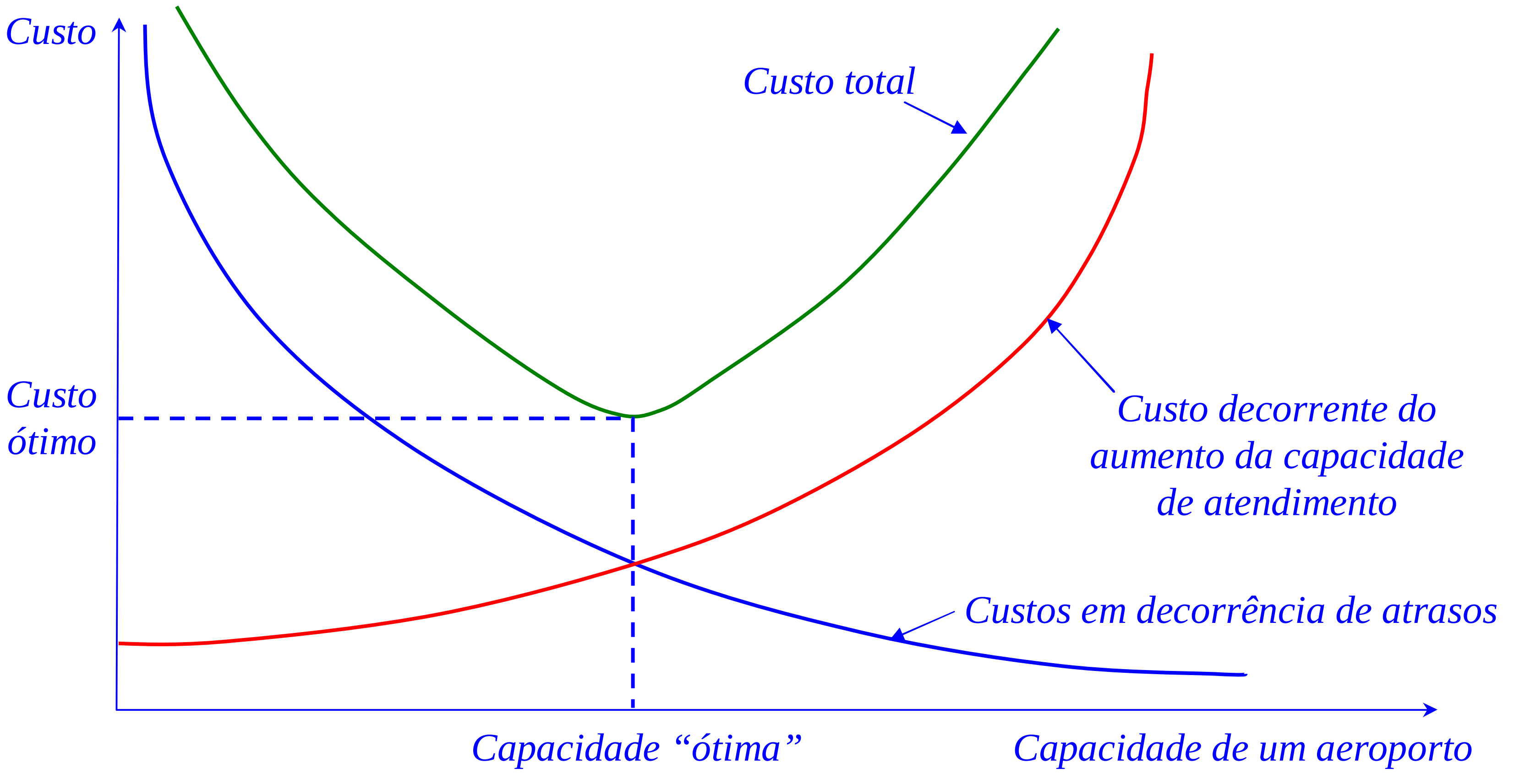
Rede de filas formada por cinco sub-sistemas de filas



Aplicações da Teoria das Filas

- Sistemas tradicionais
  - Balcão de check-in em um aeroporto
  - Caixas automáticos
  - Restaurantes self-service
  - Espera numa ligação 0800
  - Interseção viária
  - Cabines de pedágio
  - Chamados a polícia, bombeiro ou empresas prestadoras de serviços
- Critérios para definição de níveis de serviço (NS)
- Análise econômica da relação entre custos operacionais, investimentos de capital e níveis de serviço desejados

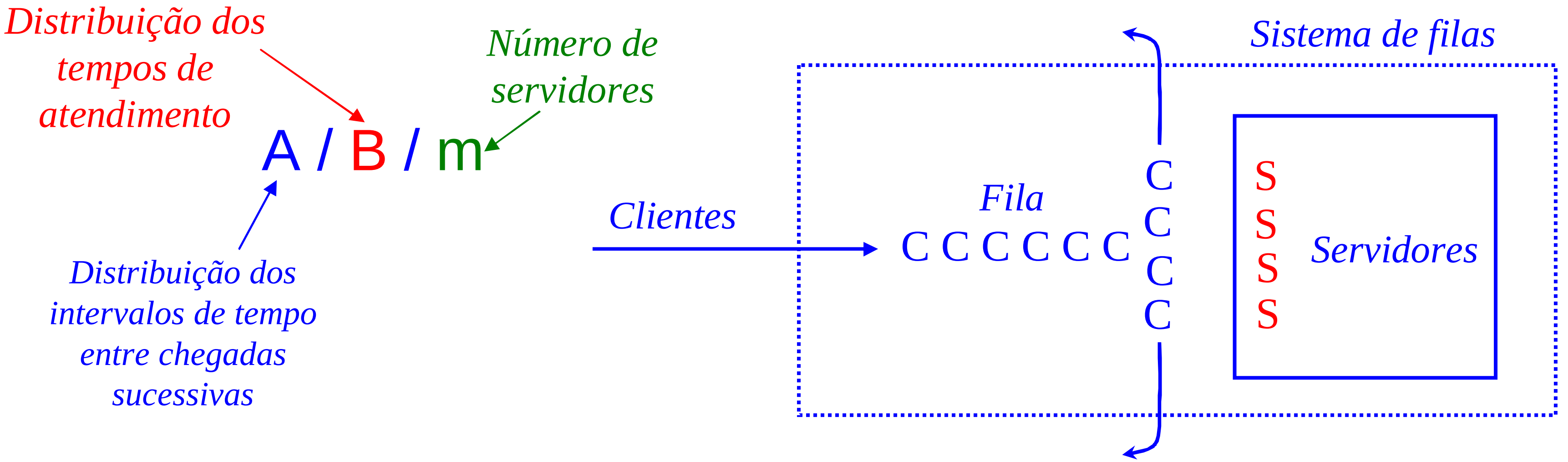
Os modelos de filas podem ser essenciais na análise de investimentos de capital



Vantagens e limitações da teoria das filas

- Os modelos de filas sempre envolvem aproximações e simplificações do sistema real
- Os resultados podem ser úteis para:
  - estimativas a respeito da grandeza de medidas de desempenho do sistema
  - análises de sensibilidade a respeito do impacto de mudanças operacionais
  - tomada de decisão sobre melhorias no sistema
- Os resultados são limitados a “condições de equilíbrio”, e derivados principalmente a partir de suposições estabelecidas para processos de renovação e sistemas de “fases”
- Definição de alguns limites úteis para sistemas mais genéricos em condições de equilíbrio
- Soluções numéricas cada vez mais viáveis para sistemas dinâmicos

Nomenclatura de modelos de filas A/B/m



- Códigos típicos para *A* e *B*:
  - *M*: exponencial negativa (M vem de *memoryless*, ou perda de memória)
  - *D*: determinística
  - E_k: distribuição Erlang de ordem *k*
  - *G*: distribuição genérica
- O modelo abordado nesta apresentação é o M/M/1

Terminologia e notação

- *Estado do sistema*: número de clientes no sistema de filas
- *Comprimento da fila:* número de clientes aguardando atendimento
- $N(t)$ = número de clientes no sistema no instante t
- $P_n(t)$ = probabilidade de $N(t)$ ser igual a n
- λ_n: taxa média de chegadas quando $N(t) = n$
- μ_n: taxa média (combinada) de atendimentos quando $N(t) = n$

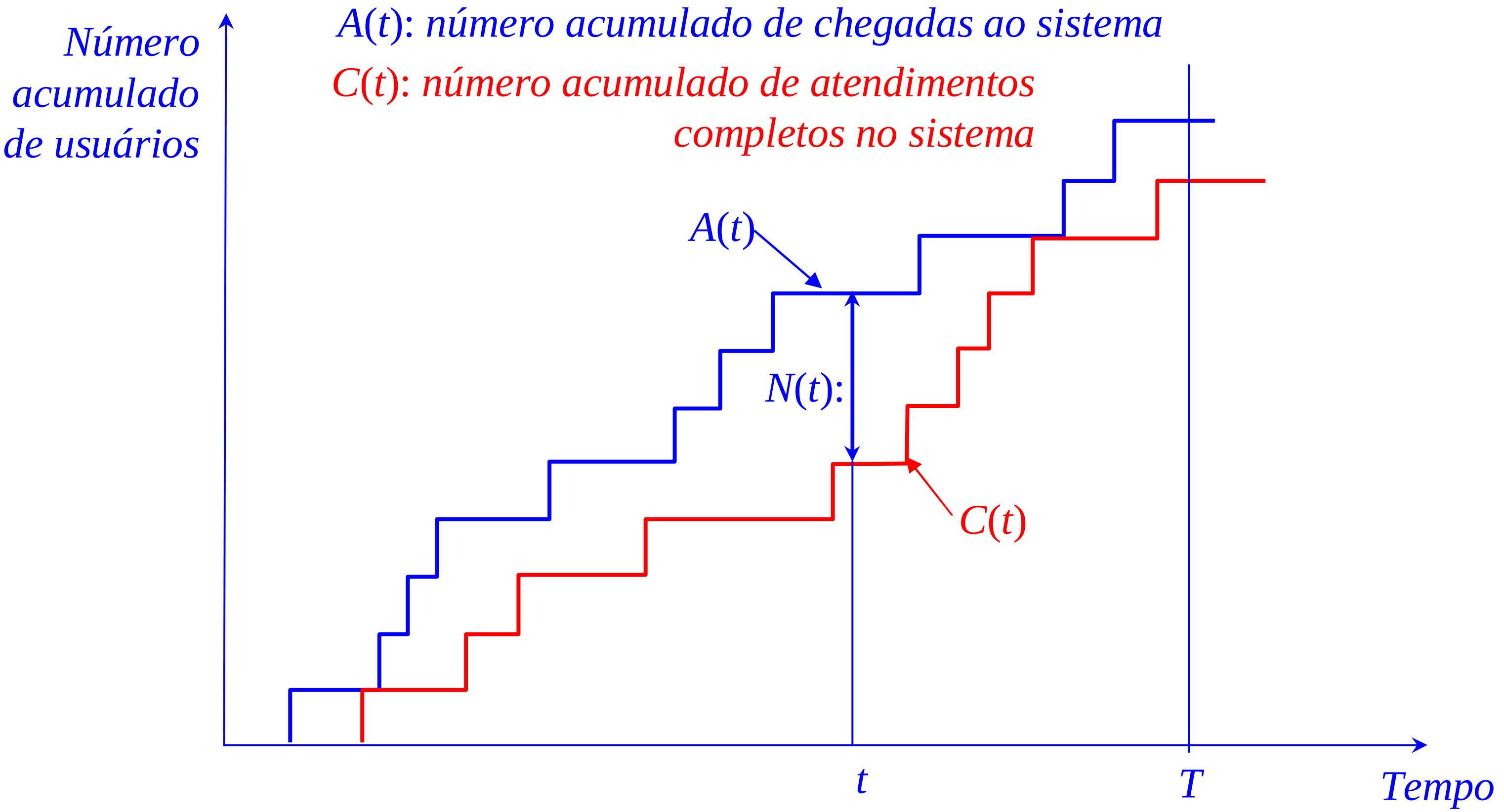
Terminologia e notação (2)

- *Condição transiente*: estado do sistema no instante t depende do estado do sistema quando $t = 0$ e de t
- *Condição de equilíbrio*: o sistema independe do estado inicial e do instante t
- m: número de servidores (canais paralelos de atendimento
- Se λ_n e a taxa de atendimento são constantes, então $\lambda_n = \lambda$, $\mu_n = \min(n\mu, m\mu)$
- $1/\lambda$ = tempo médio entre chegadas sucessivas
- $1/\mu$ = tempo médio de atendimento

Medidas de desempenho de interesse para condição de equilíbrio

- Dados
  - λ = taxa de chegadas
  - μ = taxa de atendimento por canal de atendimento
- Incógnitas
  - L = número médio de clientes no sistema
  - L_q = número médio de clientes em fila
  - W = tempo médio de um cliente no sistema ($W = E(w)$)
  - W_q = tempo médio de um cliente em fila ($W_q = E(w_q)$)
- 4 incógnitas $\Rightarrow$ 4 equações

Lei de Little



$$L_T = \frac{\int_0^T N(t)dt}{T} = \frac{A(T)}{T} \cdot \frac{\int_0^T N(t)dt}{A(T)} = \lambda_T W_T$$

Relações entre L, L_q, W, W_q

- 4 incógnitas: L, W, L_q, W_q
- São necessárias 4 equações, mas só temos 3:
  - $L = \lambda W$ (teorema de Little)
  - $L_q = \lambda W_q$
  - $W = W_q + \dfrac{1}{\mu}$
- Conhecendo uma das quatro incógnitas, determinamos as outras três
- A determinação de L pode ser simples ou complicada, dependendo do tipo de sistema considerado
- $L = \sum_{n=0}^{\infty} n P_n$ (Pn : probabilidade que n clientes estejam no sistema)

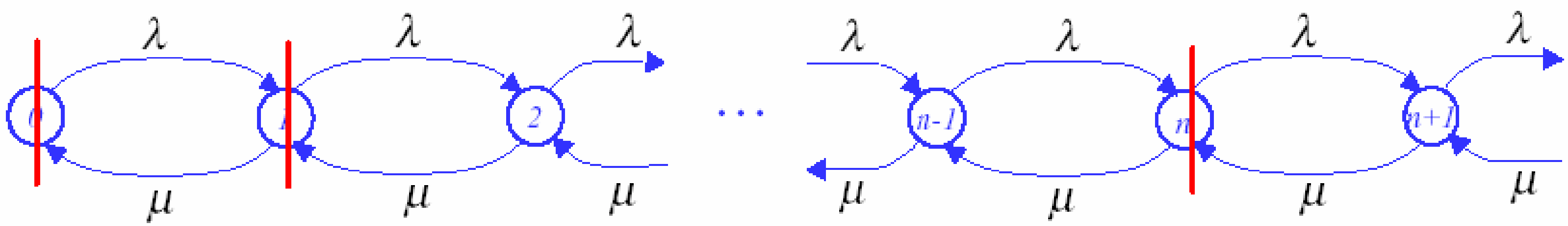
Processos de renovação em sistemas de filas

1. Existem *m* servidores idênticos e paralelos
2. Fila com capacidade infinita
3. Sempre que *n* elementos estejam no sistema (em fila ou no servidor), as chegadas são Poissonianas com taxa λ_n por unidade de tempo
4. Sempre que *n* elementos estão no sistema, os atendimentos são realizados segundo uma distribuição de Poisson com taxa μ_n por unidade de tempo
5. A disciplina da fila é FCFS (o primeiro a chegar é o primeiro a ser atendido)

M/M/1: Diagrama de transição entre estados para dois pontos

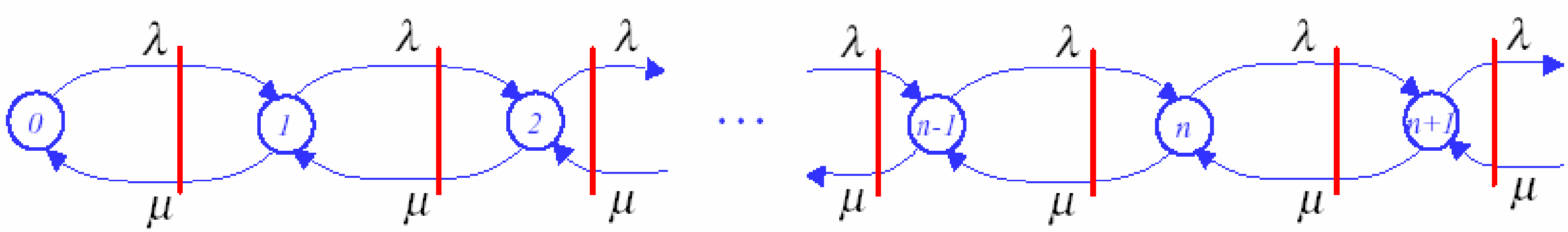
- Ponto 1

$?P_0 = \mu P_1 \quad (\lambda + \mu)P_1 = \lambda P_0 + \mu P_2 \qquad (\lambda + \mu)P_n = \lambda P_{n-1} + \mu P_{n+1}$



- Ponto 2

$?P_0 = \mu P_1 \qquad \lambda P_1 = \mu P_2 \qquad \lambda P_n = \mu P_{n+1}$



M/M/1: derivando P_0 e P_n

Etapa 1 $$P_1 = \frac{\lambda}{\mu} P_0, \quad P_2 = \left(\frac{\lambda}{\mu}\right)^2 P_0, \ \cdots, \quad P_n = \left(\frac{\lambda}{\mu}\right)^n P_0$$

Etapa 2 $$\sum_{n=0}^{\infty} P_n = 1, \quad \Rightarrow \quad P_0 \sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^n = 1 \quad \Rightarrow \quad P_0 = \frac{1}{\sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^n}$$

Etapa 3 $$\rho = \frac{\lambda}{\mu}, \text{ então } \sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^n = \sum_{n=0}^{\infty} \rho^n = \frac{1-\rho^{\infty}}{1-\rho} = \frac{1}{1-\rho} \quad (\because \rho < 1)$$

Etapa 4 $$P_0 = \frac{1}{\sum_{n=0}^{\infty} \rho^n} = 1 - \rho \quad \text{e} \qquad P_n = \rho^n (1-\rho)$$

M/M/1: derivando L, L_q, W, W_q

- $L = \sum_{n=0}^{\infty} nP_n = \sum_{n=0}^{\infty} n\rho^n (1-\rho) = (1-\rho)\sum_{n=0}^{\infty} n\rho^n = (1-\rho)\rho\sum_{n=1}^{\infty} n\rho^{n-1}$

$$= (1-\rho)\rho \frac{d}{d\rho}\left(\sum_{n=0}^{\infty} \rho^n\right) = (1-\rho)\rho \frac{d}{d\rho}\left(\frac{1}{1-\rho}\right)$$

$$= (1-\rho)\rho\left(\frac{1}{(1-\rho)^2}\right) = \frac{\rho}{(1-\rho)} = \frac{\lambda/\mu}{1-\lambda/\mu} = \frac{\lambda}{\mu-\lambda}$$

- $W = \frac{L}{\lambda} = \frac{\lambda}{\mu-\lambda}\cdot\frac{1}{\lambda} = \frac{1}{\mu-\lambda}$

- $W_q = W - \frac{1}{\mu} = \frac{1}{\mu-\lambda} - \frac{1}{\mu} = \frac{\lambda}{\mu(\mu-\lambda)}$

- $L_q = \lambda W_q = \lambda \cdot \frac{\lambda}{\mu(\mu-\lambda)} = \frac{\lambda^2}{\mu(\mu-\lambda)}$