Universidade do Minho
Escola de Engenharia
Departamento de Produção e Sistemas

Filas de Espera

Elsa Silva
elsa@dps.uminho.pt

Índice

Resultados de aprendizagem

- **Caracterizar** um sistema que possa ser modelado através da teoria das filas de espera.
- Obter as **medidas de desempenho** fundamentais com base nas probabilidades de estado e nas fórmulas de Little.
- Obter as **medidas de desempenho fundamentais** com base em fórmulas para as filas de espera mais simples **(M/M/1 e M/M/s)**.
- **Sugerir alterações** a sistemas de filas de espera com base nas medidas de desempenho e nos custos de diferentes configurações / parâmetros.

Caracterização geral

- De forma geral, está-se perante um sistema que pode ser modelado através da **teoria das filas de espera** quando, ao longo do tempo, várias entidades pretendem utilizar um serviço, não sendo garantido que a **capacidade de prestação do serviço** é suficiente para todas as entidades serem atendidas logo que o desejem.
- Em teoria das filas de espera, as entidades que procuram o serviço designam-se por ***clientes*** e as entidades que prestam o serviço por ***servidores***.
- A fila de espera pode ser física ou conceptual.
- Exemplos de sistemas que podem ser modelados através de filas de espera:
  - Caixas multibanco;
  - Urgências de hospitais;
  - Filas de espera para cirurgias;
  - Emergência médica.

Exemplos

Clientes	Serviço	Servidor(es)
Pessoas	Operações bancárias (caixa automática)	Caixas automáticas
Pacientes	Urgência médica (hospital)	Equipas de urgência
Pacientes	Cirurgia	Equipas cirúrgicas
Compradores	Compra de bens em lojas	Vendedores
Telefonemas	Atendimento telefónico (*call center*)	Colaboradores
Mensagens	Encaminhamento em redes de computadores	*Router*
Aviões	Aterragem (aeroporto)	Pistas de aeroporto
Barcos	Descarregamento / Carregamento (porto marítimo)	Guindastes
Carros	Travessia marítima	Ferry boat
Máquinas	Reparação/Preparação	Operadores
Encomendas	Processamento	Linhas de produção
Tarefas	Processamento	Processador (de computador)
Falha de stock	Entrega de produto	Fornecedores
Veículos	Deslocação viária	Cruzamentos / rotundas
Pessoas	Transporte / deslocação	Taxis

Motivação (1)

Abastecimento de combustível numa única bomba

- **Cenário determinístico 1**
  - Chega um cliente exactamente de **15 em 15 minutos** → taxa de chegada é de **4 clientes por hora**
  - Cada cliente demora exactamente **10 minutos** a ser atendido → taxa de serviço é de **6 clientes por hora**
  - **Nunca há fila de espera!**
  - Proporção de tempo que o servidor está ocupado **é 2/3**

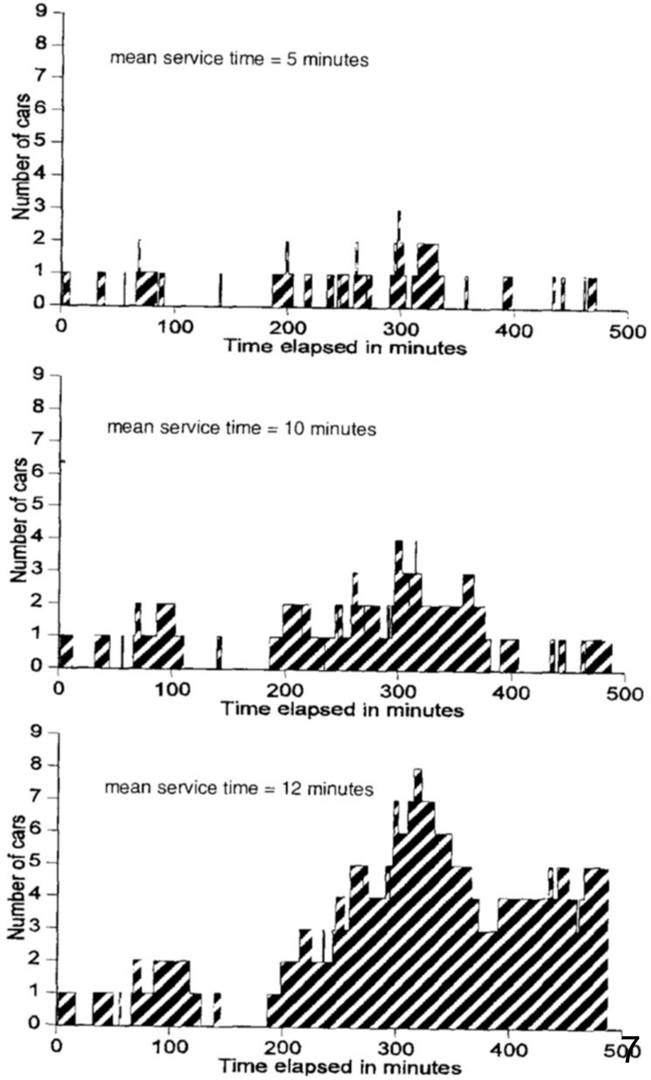
- **Cenário determinístico 2**
  - Chega um cliente exactamente de **15 em 15 minutos** → taxa de chegada é de **4 clientes por hora**
  - Cada cliente demora exactamente **20 minutos** a ser atendido → taxa de serviço é de **3 clientes por hora**
  - **Fila de espera nunca pára de crescer!**
  - Servidor está sempre ocupado

Motivação (2)

- **Cenários estocásticos**
  - Tempo entre chegadas segue uma **distribuição de probablilidade** (exponencial negativa) com **valor esperado de 15 minutos**
  - Tempo de serviço segue uma **distribuição de probablilidade** (exponencial negativa) **com valor esperado de 5 minuto**s (cenário 1), **10 minutos** (cenário 2), **12 minutos** (cenário 3)

- Exemplos de questões relevantes (a serem respondidas no seguimento destas notas):
  - **Qual o comprimento médio da fila?**
  - **Em média, quanto tempo tem de esperar um cliente para ser atendido?**

Simulação do comprimento da fila ao longo de 500 minutos para os cenários 1, 2 e 3 (Daellenbach and McNickle, 2005)



Motivação (3)

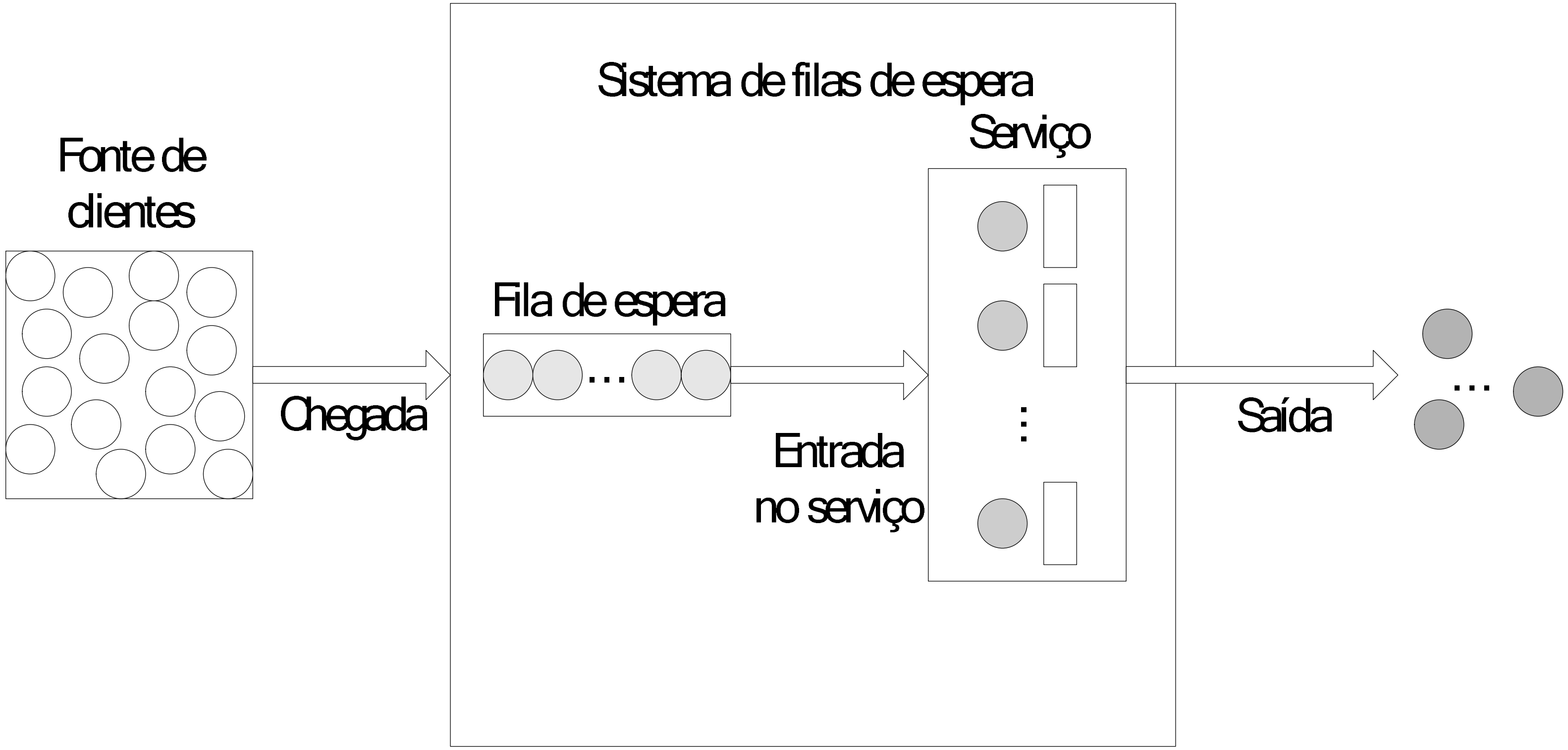
- Sistemas de filas de espera podem ser estudados como uma **sequência de eventos aleatórios** em que a distribuição de probabilidade de cada evento é conhecida / pressuposta / estimada

- **Vantagens**
  - Análise é exclusivamente teórica (esforço dispendido na obtenção de resultados é muito reduzido)
  - Análise é simples

- **Desvantagens**
  - Simplificações do sistema para permitir tratamento por modelos filas de espera podem ser significativas (quando comparados, por exemplo, com modelos de simulação)

- **Cuidados a ter**
  - Pressupostos podem ser abusivos

Caracterização de sistemas de filas de espera (1)



- Questão fundamental: **qualidade do serviço *vs.* custo do serviço**
  - Um extremo: **sistema rarefeito** (servidores muito pouco ocupados)
  - Outro extremo**: sistema congestionado** (servidores muito ocupados)

Caracterização de sistemas de filas de espera (2)

- **Fonte (gera clientes que chegam ao sistema)**
  - **Número de fontes**
    - no caso de haver tratamento para diferentes grupos, a cada grupo deverá ser associada uma fonte
  - **Dimensão da população**
    - finita (importante considerar quando o número de clientes no sistema influencia o padrão das chegadas)
    - Infinita
  - **Dimensão da chegada**
    - simples
    - em grupo
  - **Controlo das chegadas**
    - controláveis
    - incontroláveis

Caracterização de sistemas de filas de espera (3)

- **Padrão das chegadas** (distribuição do tempo entre chegadas)
  - constante
  - aleatória
    - observação experimental
    - distribuição de probabilidade teórica

- **Taxa de chegada** (λ, número médio de clientes que procuram o serviço por unidade de tempo)
  - pode depender ou não do número de clientes no sistema

- **Atitude dos clientes**
  - paciente
  - impaciente

Caracterização de sistemas de filas de espera (4)

- **Fila (clientes que esperam o atendimento)**
  - **Número de filas**
    - nenhuma
    - única
    - uma por servidor
  - **Comprimento máximo da fila**
    - capacidade virtualmente infinita
    - Finita
  - **Disciplina da fila**
    - FIFO – first in first out
    - LIFO – last in first out
    - com prioridades
    - aleatória

Caracterização de sistemas de filas de espera (5)

- **Serviço (postos de atendimento)**
  - **Configuração**
    - servidores em paralelo
      - conjunto de servidores iguais, cada servidor atende um cliente de cada vez, cliente só é atendido por um servidor
    - fases (servidores em sequência)
      - o atendimento de um tipo só pode ser feito depois do atendimento de outro tipo terminar (cada fase pode ter servidores em paralelo)
    - redes de filas de espera
      - "percurso" dos clientes não é estruturado
  - **Dimensão do serviço**
    - simples
    - em grupo

Caracterização de sistemas de filas de espera (6)

- **Padrão do serviço** (distribuição do tempo de serviço)
  - constante
  - aleatória
    - observação experimental
    - distribuição de probabilidade teórica

- **Taxa de serviço** (μ, número médio de clientes que podem ser atendidos num servidor por unidade de tempo)
  - pode depender ou não do número de clientes no sistema

Caracterização de sistemas de filas de espera (7)

- Excepto quando referido algo em contrário considera-se
  - **Chegadas simples, incontroláveis, aleatórias, com taxa independente do número de clientes no sistema e clientes pacientes**
  - **Uma única fase com servidores paralelos e idênticos, tempo de serviço aleatório, com taxa de serviço independente do número de clientes no sistema**
  - O sistema em regime permanente (por oposição a regime transitório): o seu estado é independente do estado inicial

Caracterização de sistemas de filas de espera (8)

- Exemplo: uma **caixa automática**

		Modelo
Fonte	**Clientes**	Todas as pessoas*
	Número de fontes	Uma
	Dimensão da população	Infinita*
	Dimensão da chegada	Simples*
	Controlo das chegadas	Incontroláveis
	Padrão das chegadas	Aleatório
	Taxa de chegada	Não depende do número de clientes no sistema
	Atitude dos clientes	Paciente*
Fila	**Número**	Uma
	Capacidade	Infinita
	Disciplina	FIFO
Serviço	**Configuração**	Uma fase com um servidor
	Dimensão	Simples
	Padrão do serviço	Aleatório
	Taxa de serviço	Não depende do número de clientes no sistema

Conceitos básicos e notação (1)

- **Estado do sistema**: número total de clientes no sistema (número de clientes na fila mais o número de clientes a serem atendidos)

λ - **taxa de chegada (**número médio de clientes que procuram o serviço por unidade de tempo)

$1/\lambda$ - **tempo médio entre chegadas** (tempo médio entre duas chegadas seguidas)

μ - **taxa de serviço** (número médio de clientes que podem ser atendidos por unidade de tempo) de um servidor

$1/\mu$ - **tempo médio de serviço**

s - **número de servidores** (paralelos)

ρ - **intensidade de tráfego** (ou taxa de ocupação) – fracção de tempo que *cada* servidor está ocupado

- $\rho = \lambda / (s\mu)$
- Se $\rho \geq 1$ a fila aumenta indefinidamente, logo não existe regime permanente

Medidas de desempenho fundamentais

- W - tempo médio (de permanência de um cliente) no sistema
- W_q - tempo médio (de espera) na fila
- W_s - tempo médio de atendimento, $W_s = 1/\mu$

$$W = W_q + W_s$$

- L - número médio de clientes no sistema
- L_q - número médio de clientes na fila (comprimento médio da fila)
- L_s - número médio de clientes a serem atendidos

$$L = L_q + L_s$$

- π_n probabilidade de estarem n clientes no sistema (ou fracção de tempo em que estão n clientes no sistema)

Relações fundamentais

- *K*: capacidade do sistema (número máximo de clientes no sistema)
  - se um cliente chega ao sistema e já estão *K* clientes no sistema, o cliente não entra
  - se o sistema não tiver limite de capacidade, $K = \infty$

$$L = \sum_{n=0}^{K} n\pi_n \qquad L_q = \sum_{n=s+1}^{K} (n-s)\pi_n \qquad L_s = \sum_{n=0}^{s-1} n\pi_n + s\left[\sum_{n=s}^{K} \pi_n\right]$$

- **Fórmulas de Little**

$$\boxed{\begin{array}{l} L = \lambda W \\ L_q = \lambda W_q \end{array}}$$

Exemplo (1)

- Uma tabacaria tem um **único funcionário** que demora, em média, **0.5 minutos a atender** um cliente. Em média, a tabacaria **tem 20 clientes por hora**.

1. Quantos clientes podem ser atendidos por hora?
2. Qual o tempo médio entre chegadas?
3. O funcionário trabalha 8 horas por dia. Quanto tempo está ocupado?

1. $1 / \mu = 0.5$ minutos
$\mu = 2$ clientes / minuto = 120 clientes/hora

2. $\lambda = 20$ clientes / hora
$1 / \lambda = 0.05$ horas = 3 minutos

3. $\rho = \lambda / \mu = 20 / 120 = 0.167$
*Tempo ocupado por dia = 0.167*8 horas = 1.333 horas*

Exemplo (2)

- Um centro de atendimento telefónico ao cliente tem **três colaboradores e seis linhas de atendimento**
- Quando um cliente telefona pode
  - ser logo atendido (se algum colaborador estiver disponível),
  - ter de esperar ocupando uma linha (se os colaboradores estiverem todos ocupados mas existirem linhas livres),
  - ou não conseguir a ligação (se não existirem linhas disponíveis).
- Uma estimativa das probabilidades associadas a cada um dos possíveis estados do sistema é dada na tabela seguinte

Estado, n	0	1	2	3	4	5	6
Probabilidade, π_n	0.068	0.170	0.212	0.177	0.147	0.123	0.103

Exemplo (2)

1. Qual a probabilidade de todos os colaboradores estarem desocupados?
2. Qual a probabilidade de um cliente não ter de esperar?
3. Qual a probabilidade de um cliente esperar na fila?
4. Qual a probabilidade de um cliente não conseguir ligação?

1. $\pi_0 = 0.068$

2. $\pi_0+\pi_1+\pi_2 = 0.45$

3. $\pi_3+\pi_4+\pi_5= 0.447$

4. $\pi_6 = 0.103$

Exemplo (3)

- Um estação dos correios tem **três balcões**. O tempo médio entre-chegadas de clientes é de **30 segundos** e estes demoram, em média, **1.25 minutos a serem atendidos**. Em média, um cliente passa **3 minutos na estação dos correios**.

1. Qual o número médio de clientes no sistema?
2. Qual a percentagem de tempo que os colaboradores dos correios estão desocupados?

- *$s = 3$*
- *$1/\lambda = 0.5$ minutos/ cliente*
- *$1/\mu = 1.25$ minutos / cliente*
- *$W = 3$ minutos*

1. *$L = \lambda\ W = 6$ clientes*
2. *$\lambda = 2$ clientes / minuto*
   $\mu = 0.8$ clientes / minuto
   Proporção que um servidor está ocupado é ρ, logo a proporção desocupado é:
   $1 - \rho = 1 - \lambda / (s\ \mu) = 0.167$, logo a percentagem é 16.7%.

Notação de Kendall (1)

- **Notação de Kendall: *a / b / c / d / e / f***
  - ***a*** **: Tempo entre-chegadas dos clientes**
    - M : distribuição exponencial;
    - D : determinístico;
    - G : distribuição geral ou arbitrária;
    - …
  - ***b*** **: Tempo de serviço**
    - M : distribuição exponencial;
    - D : determinístico;
    - G : distribuição geral ou arbitrária;
    - …
  - ***c*** **: Número de servidores**

Notação de Kendall (2)

- ***d*** **: Capacidade do Sistema**

- ***e*** **: Dimensão da população de clients**

- ***f*** **: Disciplina da fila**
  - FIFO (*First In First Out);*
  - LIFO (*Last In First Out);*
  - esquemas de prioridade;
  - …

- Exemplos
  - M / M / 1 / ∞ / ∞ / FIFO
  - M / M / 5
  - M / D / 2 / 7 / ∞ / FIFO

Filas de espera de Markov (1)

- **M / M / ...**

- Se o tempo entre-chegadas e o tempo de serviço seguem a **distribuição exponencial negativa** está-se perante uma fila de espera de **Markov**

- Para várias filas de espera de Markov é possível derivar as **probabilidades de estado** ($\pi_0, \pi_1, \pi_2, \pi_3, \ldots$) e com base nessas probabilidades obter **fórmulas fechadas para** o número esperado de clientes e tempos esperados (em regime permanente)

- **Tempo entre chegadas com distribuição exponencial negativa**
  - Chegadas são **independentes** (clientes não estão relacionados)
  - Tempo até à próxima chegada é **independente** do tempo que decorreu desde a última chegada
  - **Tempos entre chegadas curtos são mais frequentes** que entre chegadas longos

Filas de espera de Markov (2)

- M / M / ...
- **Tempo de serviço com distribuição exponencial negativa**

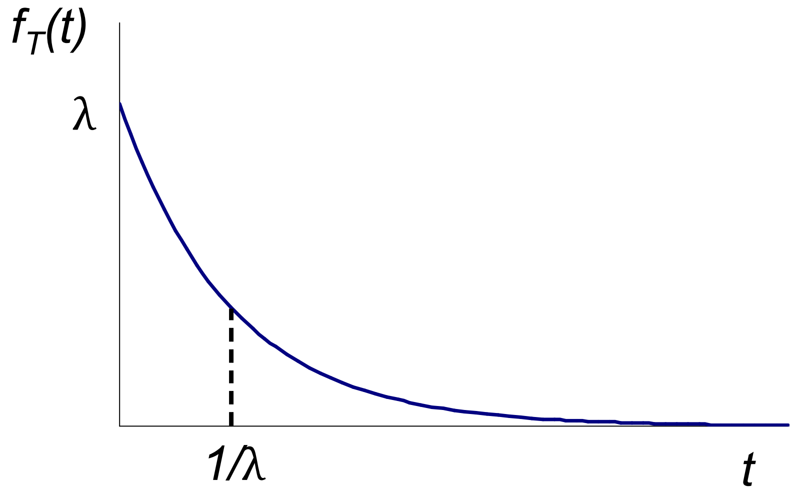
  - Tempo **até à conclusão serviço não depende do tempo que já passou** (se já passou muito tempo, não quer dizer que o atendimento esteja quase a acabar mas sim que o cliente tem alguma característica que o torna o seu atendimento mais demorado)

  - **Muitos clientes são atendidos depressa** e poucos demoram muito tempo

Filas de espera de Markov (3)

- Função densidade de probabilidade da distribuição exponencial com parâmetro λ

- $f_T(t) = \begin{cases} \lambda e^{-\lambda t}, \text{para } t \geq 0 \\ 0, \text{para t} \leq 0 \end{cases}$



- Valor esperado **E(T) = 1/ *λ***

- Probabilidades acumuladas (*t>0*)

$$P\{T \leq t\} = 1 - e^{-\lambda t}$$

$$P\{T > t\} = e^{-\lambda t}$$

Filas de espera de Markov (4)

- A probabilidade de ocorrerem **valores pequenos é maior** do que a probabilidade de ocorrerem valores próximos do valor esperado $E(T)=1/\lambda$

- A distribuição exponencial **não tem memória** (por exemplo, se o tempo entre-chegadas segue uma distribuição exponencial, o tempo até à próxima chegada não é influenciado pelo tempo que já decorreu desde a última chegada)
  - $P\{T > t + \Delta t \mid T > \Delta t\} = P\{T > t\}$

Filas de espera de Markov (6)

- Resultados a serem apresentados para as filas de espera M/M/1 e M/M/s **são válidos** para qualquer disciplina da fila (FIFO, LIFO, ...) desde que
  - Todos os clientes **permaneçam na fila** depois de nela entrarem (clientes são pacientes);
  - O **tempo de serviço médio seja igual** para todos os clientes;
  - O **servidor acabe de atender o cliente antes de começar a atender** o seguinte (não há interrupções);
  - Se houver clientes em espera, um **servidor que acabe de atender um cliente passa imediatamente a atender outro.**

Filas de espera de Markov (7)

- **Resultados para uma fila de espera M / M / 1**

$$\rho = \lambda / \mu$$
$$\pi_0 = 1 - \rho$$
$$\pi_n = \rho^n \pi_0 = \rho^n (1-\rho), n \geq 1$$

$$L_q = \frac{\rho^2}{1-\rho}$$
$$L_s = \rho$$
$$L = \frac{\rho}{1-\rho}$$

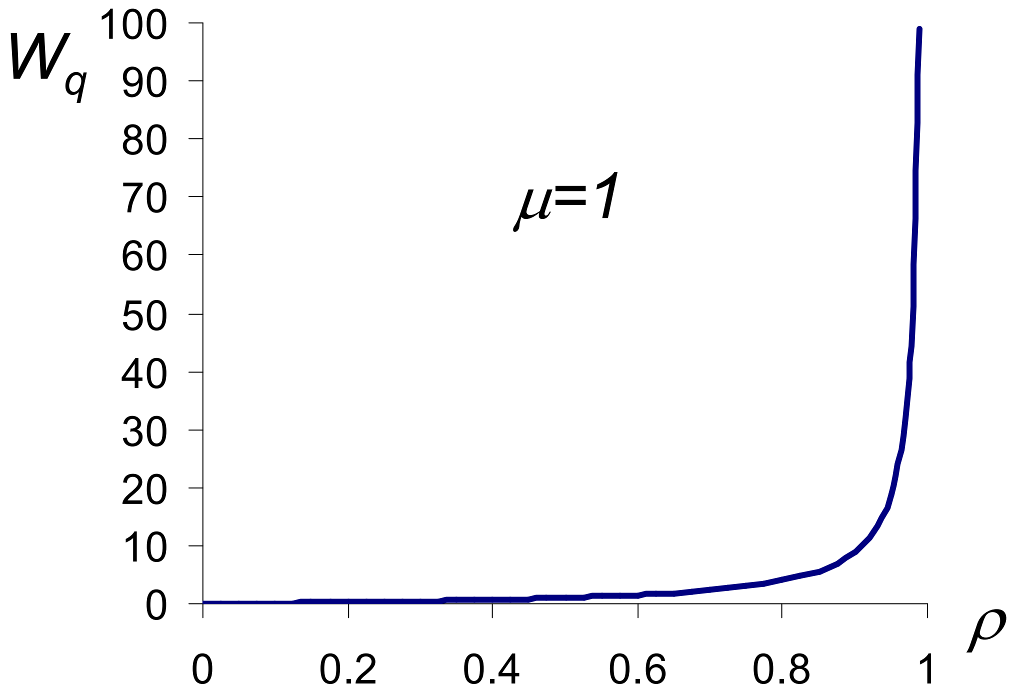
$$W_q = \frac{\rho}{\mu(1-\rho)}$$
$$W_s = 1/\mu$$
$$W = \frac{1}{\mu(1-\rho)}$$

- Probabilidade do tempo de permanência na fila exceder o valor t

$$W_q(t) = \begin{cases} \rho, \, para \; t = 0 \\ \rho e^{-\mu(1-\rho)t}, \, para \; t \geq 0 \end{cases}$$

Filas de espera de Markov (8)

- **Tempo esperado na fila vs. intensidade de tráfego.**



Filas de espera de Markov (9)

- **Resultados para uma fila de espera M / M / s**

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

$$\pi_0 = \left[\frac{(s\rho)^s}{s!(1-\rho)} + \sum_{n=0}^{s-1} \frac{(s\rho)^n}{n!} \right]^{-1}$$

$$\pi_n = \begin{cases} \dfrac{(s\rho)^n \pi_0}{n!}, \, para \; 1 \le \mathrm{n} \le \mathrm{s} \\ \dfrac{s^s \rho^n \pi_0}{s!}, \, para \; \mathrm{n} \ge \mathrm{s} \end{cases}$$

- Probabilidade de todos os servidores estarem ocupados

$$P_B = \frac{\pi_s}{1-\rho}$$

Filas de espera de Markov (10)

- **Resultados para uma fila de espera M / M / s**

$$L_q = \frac{s^s \rho^{s+1} \pi_0}{s!(1-\rho)^2}$$
$$L_s = \lambda / \mu$$
$$L = L_q + L_s$$

$$W_q = L_q / \lambda$$
$$W_s = 1 / \mu$$
$$W = W_q + W_s$$

  - Probabilidade do tempo de espera na fila exceder o valor t

$$W_q(t) = \begin{cases} 1 - \dfrac{(s\rho)^s \pi_0}{s!(1-\rho)}, \text{para } t = 0 \\ \dfrac{(s\rho)^s \pi_0}{s!(1-\rho)} e^{-s\mu(1-\rho)t}, \text{para } t > 0 \end{cases}$$

Filas de espera de Markov (11)

- Outras filas de espera de Markov para as quais há resultados estabelecidos:

  - M/M/1/K
  - M/M/s/K
  - M/M/s/K/N

- Exemplos de filas de espera não-Markovianas para as quais também há resultados estabelecidos:
  - M/G/1
  - M/D/1
  - M/D/s
  - GI/G/1
  - GI/G/s
  - …

Exemplo (4)

- Num determinado processo produtivo, todas as peças têm de passar por uma determinada máquina. Esta **demora, em média, 30 segundos a processar** uma peça. O tempo **médio entre-chegadas de peças à máquina é de 40 segundos**. A gestão da fábrica pretende que, por razões de espaço, em 90% do tempo não haja mais de três peças em espera e que não mais de 10% das peças esperem mais de um minuto para serem processadas pela máquina. Assuma que os tempos entre-chegadas e de serviço seguem a distribuição exponencial.

1. Na situação actual, as metas colocadas pela gestão estão a ser cumpridas?
2. Qual o efeito da compra de uma nova máquina (idêntica) que funcione em paralelo com a existente?

Exemplo (4)

- Situação actual
  - M / M / 1
  - λ = *1.5 peças / minuto*
  - μ = *2 peças / minuto*
  - Probabilidade de estarem mais de três peças em espera =
    $= \pi_5 + \pi_6 + \pi_7 + \pi_8 + \ldots = 1 - (\pi_0 + \pi_1 + \pi_2 + \pi_3 + \pi_4) =$
    $= 0.237$
    Em *76.3%* do tempo não há mais de três peças em espera. A gestão pretendia que esse valor fosse de *90%*.
  - $W_q(1) = 0.455$.
  - *45.5 %* das peças esperam mais de 1 minuto para serem processadas. A gestão pretendia que esse valor fosse de 10%.

Exemplo (4)

- Com uma nova máquina
  - M / M / s
  - s = 2
  - λ *= 1.5 peças / minuto*
  - μ *= 2 peças / minuto*
  - Probabilidade de estarem mais de três peças em espera =
    $= \pi_6 + \pi_7 + \pi_8 + \ldots = 1 - (\pi_0 + \pi_1 + \pi_2 + \pi_3 + \pi_4 + \pi_5) =$
    $= 0.004$
    Em *99.6%* do tempo não há mais de três peças em espera, o que cumpre a meta da gestão.
  - $W_q(1) = 0.0167.$
  - *1.67 %* das peças esperam mais de 1 minuto para serem processadas, o que cumpre a meta da gestão.

Exemplo (5)

- Compare o tempo esperado na fila para as duas situações seguintes para diferentes valores da taxa média de chegada.

1. **Dois servidores cada um deles com uma fila** de espera (como, por exemplo, o que acontece num supermercado com duas caixas abertas).

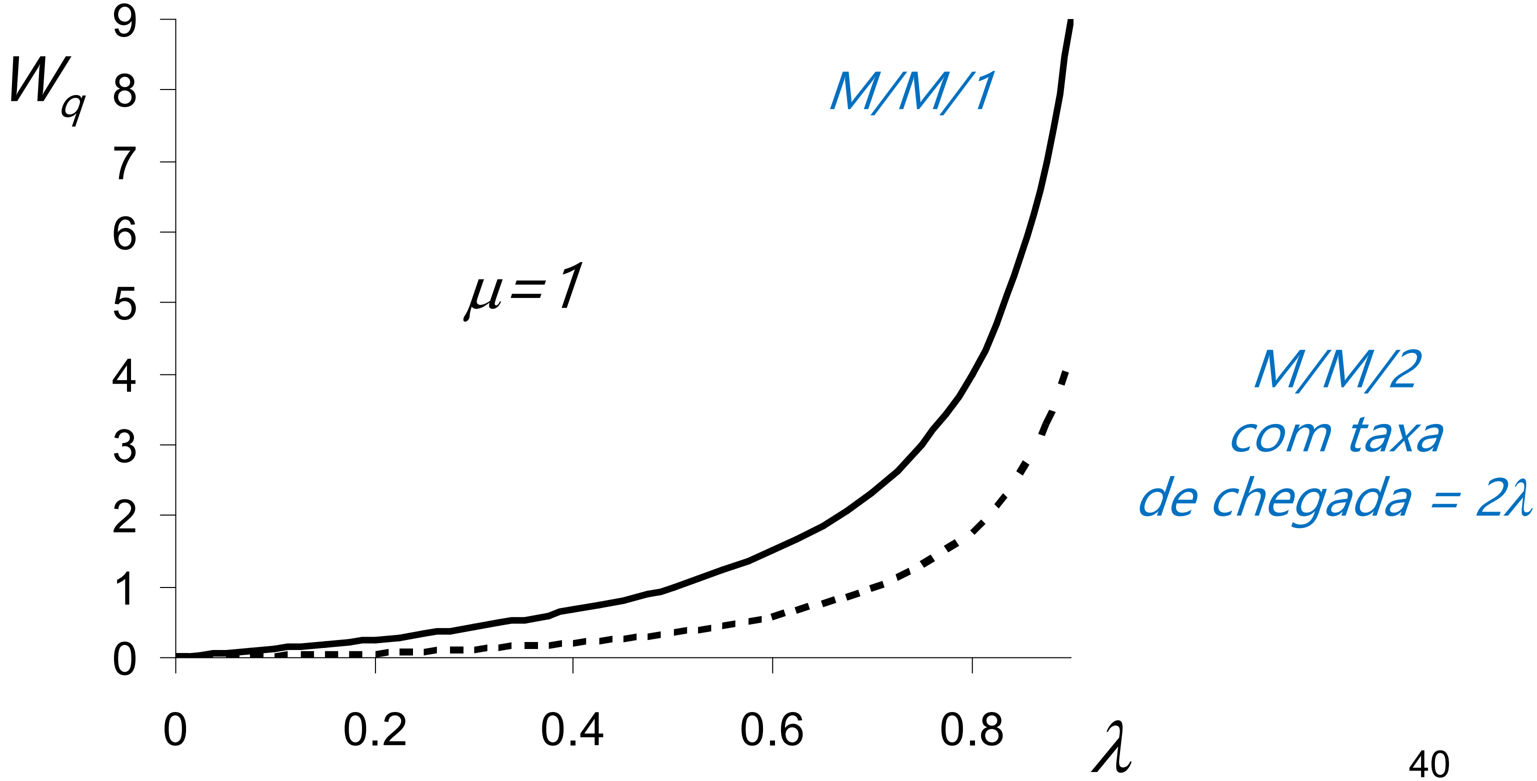
2. **Dois servidores com uma fila única** (como, por exemplo, o que acontece em balcões de atendimento com senhas numeradas).

- Considere que a taxa média de atendimento de um servidor é de 1 cliente / unidade de tempo em ambas as situações.

Exemplo (5)

1. **Dois sistemas M/M/1 em paralelo** (tempo de espera começa a degradar-se muito a partir de uma taxa de chegada de cerca 1.4 que corresponde a 0.7 para cada fila).

2. **Um sistema M/M/2** (tempo de espera começa a degradar-se muito a partir de uma taxa de chegada de cerca 1.6).



Exemplo (6)

- Numa estação de correios de uma pequena vila chegam à estação **4 clientes por hora**. Atualmente a estação tem apenas um funcionário a atender os clientes e estimou-se que o funcionário atende em média **5 clientes por hora**.
  Considere que o pressuposto de se tratar de um sistema de filas de espera de Markov é aceitável.

Dados:

M/M/1

s = 1

λ = 4 clientes/ hora

μ = 5 clientes / hora

1/ λ = ¼ =0,25 entra um cliente na estação a cada 0,25 horas, ou seja a cada 15 minutos

1/ μ = 1/5 =0,2 o funcionário demora em media 0,2 horas a atender, ou seja 12 minutos

Exemplo (6)

- **Medidas de desempenho**

$\rho = \lambda / \mu = 4/5 = 0{,}8 = 80\%$ Taxa de ocupação

$\boldsymbol{L_q} = \frac{\rho^2}{1-\rho} = \frac{{}^4/_5{}^2}{1-\frac{4}{5}} = 3{,}2$ -> clientes em média na fila à esperade serem atendidos

$\boldsymbol{L_s} = \frac{\lambda}{\mu} = \frac{4}{5} = 0{,}8$ -> número médio de clientes a serem atendidos

$\boldsymbol{L} = L_q + L_s = 3{,}2 + 0{,}8 = 4$ -> número médio de clientes na estação dos correios

$\boldsymbol{W_q} = \frac{L_q}{\lambda} = \frac{3{,}2}{4} = 0{,}8$ horas = 48 minutos -> Tempo médio de espera para serem atendidos

Exemplo (7)

- Considere agora que a estação de correios passa a ter **2 funcionários** e se mantém o sistema de **fila única**.

Dados:

- M/M/1
- s = 2
- *λ = 4 clientes/ hora*
- *μ = 5 clientes / hora*
- *1/ λ = ¼ =0,25 entra um cliente na estação a cada 0,25 horas, ou seja a cada 15 minutos*
- *1/ μ = 1/5 =0,2 cada funcionário demora em media 0,2 horas a atender, ou seja 12 minutos*

Exemplo (7)

- **Medidas de desempenho**

$\boldsymbol{\rho} = \frac{\lambda}{s\mu} = \frac{4}{2x5} = 0{,}4 = 40\%$ Taxa de ocupação

$\boldsymbol{\pi_0} = 0{,}4286$ -> Probabilidade de estarem **0** clientes no sistema

$\boldsymbol{L_q} = 0{,}1524$ -> clientes em média na fila à esperade serem atendidos

$\boldsymbol{W_q} = \frac{L_q}{\lambda} = \frac{0{,}1524}{4}$= 0,0381 horas = 2,286 minutos -> Tempo médio de espera para o cliente ser atendido

Exemplo (8)

- Considere agora que a estação de correios passa a ter **2 funcionários, mas cada um com uma fila**

Dados:

- 2 sistemas M/M/1
- s = 1
- λ *= 2 clientes/ hora ->* reduziu-se a taxa de chegada para metade
- μ *= 5 clientes / hora*
- *1/* λ *= 1/2 =0,5 entra um cliente na estação a cada 0,5 horas, ou seja a cada 30 minutos*
- *1/* μ *= 1/5 =0,2 cada funcionário demora em media 0,2 horas a atender, ou seja 12 minutos*

Exemplo (8)

- **Medidas de desempenho**

$\rho = \lambda / \mu = 2/5 = 0{,}4 = 40\%$ Taxa de ocupação

$\boldsymbol{L_q} = \frac{\rho^2}{1-\rho} = \frac{{}^2/_5{}^2}{1-\frac{2}{5}} = 0{,}267$ -> clientes em média na fila à espera de serem atendidos

$\boldsymbol{W_q} = \frac{L_q}{\lambda} = \frac{0{,}267}{2} = 0{,}133$ horas $= 7{,}98$ minutos -> Tempo médio de espera para o cliente ser atendido

Comparação

Medida de desempenho	M/M/1	M/M/2	2 sistemas M/M/1
$\boldsymbol{\rho}$	80%	40%	40%
$\boldsymbol{L_q}$	3,2 clientes	0,15 clientes	0,27 clientes
$\boldsymbol{W_q}$	48 minutos	2,29 minutos	7,98 minutos

Comentários finais

- **Modelos da teoria das filas de espera auxiliam na análise e/ou dimensionamento de sistemas**
- **Sistema pode ser mais eficiente**
  - **Aumentando o número de servidores** (o que, tipicamente, acarreta investimentos significativos)
  - **Aumentando a taxa de serviço**
  - **Formando uma fila para todos os servidores em vez de uma fila por servidor** (para servidores (idênticos) em paralelo)
  - **Reduzindo a variabilidade do tempo de serviço** (por exemplo, tornando o serviço mais repetitivo)
- Medidas para controlar (influenciar) a taxa de chegada (e/ou reduzir a sua variabilidade) podem ser muito benéficas
  - **Por exemplo, descontos nas horas de vazio**

Áreas e tópicos relacionados

- **Redes de filas de espera**
- **Processos estocásticos e processos Markovianos**
  - Um sistema de fila de espera é um caso particular de um processo estocástico.
- **Simulação**
  - Permite analisar sistemas de filas de espera para os quais não existem resultados analíticos (entre muitas outras coisas...).

Bibliografia e *links*

- H. G. Daellenbach and D. C. McNickle, Management Science - Decision Making Through Systems Thinking, Palgrave MacMillan, 2005.
- S. Hillier and G. J. Lieberman, Introduction to Operations Research, McGrawHill, 2005.
- P. A. Jensen and J. F. Bard, Operations Research - Models and Methods, John Wiley and Sons, 2003.
- A. M. Law and W. D. Kelton, Simulation Modeling and Analysis, McGrawHill, 2000.
- L. V. Tavares, R. C. Oliveira, I. H. Themido, F. N. Correia, Investigação Operacional, McGraw Hill, 1996.
- *http://www.usm.maine.edu/math/JPQ/*
- *http://www.me.utexas.edu/~jensen/ORMM/frontpage/jensen.lib/index.html*