



UNIVERSIDADE DO ESTADO DE MINAS GERAIS
FUNDAÇÃO EDUCACIONAL DE ITUIUTABA
CURSO DE SISTEMAS DE INFORMAÇÃO



Apostila de
Modelagem e Avaliação de
Desempenho:

Teoria das Filas e Simulações

Prof. Walteno Martins Parreira Júnior

www.waltenomartins.com.br

waltenomartins@yahoo.com

2010

SUMÁRIO

1	MODELAGEM DE SISTEMAS	2
1.1	Sistemas Balanceados	2
1.2	Filas	2
1.3	As Filas não são Simpáticas	3
1.4	As Filas são Dispendiosas	3
1.5	Aspectos Históricos	4
1.5.1	Teoria das Filas	4
1.5.2	Simulação	4
1.6	Aplicações de Modelagem de Sistemas	4
1.6.1	Linhas de Produção	4
1.6.2	Transportes	5
1.6.3	Comunicações	5
1.6.4	Bancos, Supermercados, Escritórios, etc.	6
1.6.5	Confiabilidade	6
1.6.6	Processamento de Dados	6
2	CONCEITOS BÁSICOS (I)	7
2.1	Elementos de uma Fila	7
2.2	Características de Uma Fila	7
2.2.1	Clientes e Tamanho da População	7
2.2.2	Processo de Chegada	7
2.2.3	Processo de Atendimento	8
2.2.4	Número de Servidores	9
2.2.5	Disciplina da Fila	9
2.2.6	Tamanho Médio da Fila	9
2.2.7	Tamanho Máximo da Fila	9
2.2.8	Tempo Médio de Espera na Fila	9
2.3	Variáveis Randômicas	10
2.4	Observando a Dinâmica de Uma Fila	10
2.5	Sistemas Estáveis	12
2.6	O Tamanho de Amostra	13
2.7	Opções de Dimensionamento: o Tipo da Fila	13
2.8	Gerenciando Filas	13
2.9	Exercícios	14
3	CONCEITOS BÁSICOS (II)	16
3.1	Introdução	16
3.2	Variáveis Randômicas Fundamentais	16
3.2.1	Relações Básicas	17
3.2.2	Taxa de Utilização dos Atendentes	17
3.2.3	Intensidade de Tráfego ou Número Mínimo de Atendentes	17
3.2.4	Fórmulas e Little	18
3.2.5	Resumo das Fórmulas	18
3.3	Exemplos	19
3.4	Postulados Básicos	20
3.5	Exercícios	21
4	Referencias bibliográfica	23

1 MODELAGEM DE SISTEMAS

1.1 Sistemas Balanceados

Ao efetuarmos certos tipos de estudos de planejamento, é comum depararmos com problemas de dimensionamento ou fluxo cuja solução é aparentemente complexa. O cenário pode ser uma fábrica, o trânsito de uma cidade, um escritório, um porto, uma mineração, etc. Geralmente estamos interessados em saber:

- Qual a quantidade correta de equipamentos (sejam eles máquinas, veículos, pessoas, etc);
- Qual o melhor *layout* e o melhor fluxo dentro do sistema que está sendo analisado.

Ou seja, desejamos que nosso sistema tenha um funcionamento eficiente ou otimizado. Por otimizado queremos dizer que teremos um custo adequado e que teremos usuários satisfeitos com o ambiente ou com o serviço oferecido. Chamamos tais estudos de modelagem de sistemas.

Estudos de modelagem de sistemas podem envolver modificações de layout, ampliações de fábricas, troca de equipamentos, reengenharia, automatização, dimensionamento de uma nova fábrica, etc. Assim, dado um determinado objetivo de produção ou de qualidade de atendimento, o estudo vai procurar definir a quantidade de atendentes (equipamentos, veículos, pessoas, etc) que devem ser colocados em cada estação de trabalho, assim como o melhor layout e o melhor fluxo. Para dimensionar gargalos, ou seja, pontos onde ocorrem filas. Dizemos também que um sistema ou processo adequadamente dimensionado está balanceado.

Dentre as técnicas disponíveis para a Modelagem de Sistemas temos a Teoria das Filas e a Simulação, sendo esta última a mais utilizada. Simulação, como o próprio nome indica, é uma técnica que permite imitar o funcionamento de um sistema real. Os modernos programas de computador permitem construir modelos nos quais é possível visualizar na tela o funcionamento de um banco, uma fábrica, um pedágio, um porto, um escritório, etc, tal como se estivéssemos em uma posição privilegiada em cada um destes cenários. Antes de efetuar alterações em uma fábrica real, podemos interagir com uma fábrica virtual. A junção da tradicional teoria da simulação com as técnicas modernas de computação e jogos (tais como “vídeo games”) têm possibilitado estes avanços.

1.2 Filas

Qualquer pessoa sabe exatamente o que são filas em decorrência das experiências que o dia-a-dia nos coloca. Nós entramos em uma fila para descontar um cheque em um banco, para pagar pelas compras em um supermercado, para comprar ingresso em um cinema, para pagar o pedágio em uma estrada e tantas outras situações. Filas existem também em ambientes de produção, tais como de lingotes aquecidos em uma aciaria, esperando junto a uma carregadeira a vez de serem carregados com minério.

Algumas vezes as filas são algo abstrato, tais como uma pilha de papéis referentes a pedidos de manufatura em uma fábrica de geladeira ou uma máquina estragada dentro de uma fábrica e que necessita de reparos, mas tem de aguardar porque o reparador está ocupado efetuando outros consertos. Outras vezes a fila não é vista enfileirada”mas, sim, dispersa, como, por exemplo, pessoas em uma barbearia, esperando pela vez de cortar o cabelo, aviões sobrevoando um aeroporto, esperando pela vez para aterrissar, ou navios parados no mar, esperando pela vez de atracar no porto para descarregar.

Uma área de muita importância surgiu nas últimas décadas: filas em computadores. Aqui temos filas de programas esperando por espaço na memória, ou para serem atendidos pela UCP (Unidade Central de Processamento), ou para buscar um registro de dados em um disco magnético ou para terem acesso a um servidor através da rede.

1.3 As Filas não são Simpáticas

Certamente não é agradável entrar em uma fila e esperar pelo serviço (o ideal é chegar ao local de serviço e ser imediatamente atendido) e, quando a espera é longa, ficamos aborrecidos (algumas pessoas ficam profundamente irritadas). Quando isto acontece, mudamos o nosso estado de espírito. Se existem filas, passamos a comparar o desempenho da nossa fila com o das outras e, geralmente, somos levados a pensar como uma das leis de Murphy: *“a fila que anda é a outra, mas não adianta trocar de fila, pois a fila que anda é a outra”*.

Como consequência de nossas amargas experiências tomamos algumas atitudes, tais como não mais comprar em um determinado supermercado, mudar a conta bancária para outra agência (ou banco), etc. Visto pelo ângulo das empresas, estas atitudes significam perda de negócio.

1.4 As Filas são Dispendiosas

Além de não serem simpáticas, as filas têm ainda o lado desfavorável do custo. Isto é válido em qualquer ambiente, indo de fábricas a um supermercado. Por exemplo, nas fábricas a existência de fila em um equipamento pode implicar espera por peças que necessitam ser processadas, o que ocasiona um aumento nos tempos de produção. As consequências disto são aumento nos custos e atrasos no atendimento aos pedidos dos clientes.

O ideal, conforme dissemos, seria a não existência de filas. Se isto realmente ocorresse, certamente não teríamos clientes aborrecidos, mas isto nem sempre é possível. Se imaginarmos uma barbearia com três barbeiros onde constatamos fila e fizermos esta afirmação ao proprietário, ele certamente nos contestará dizendo que, para não haver filas, seriam necessários dez ou mais barbeiros e que, então, muitos deles ficariam ociosos grande parte do tempo e, portanto, não ganhariam dinheiro suficiente para sobreviver. Talvez até o proprietário diga ainda que os atuais três barbeiros representam o melhor dimensionamento, com o qual os clientes esperam um tempo considerado tolerável. Ele poderá ainda dizer ter observado que alguns dias aparecem muitos clientes, mas que acabam voltando outro dia. Se este fato começar a ficar freqüente, ele pretende contratar mais um ou dois barbeiros e tudo voltará a funcionar de uma forma aceitável.

A situação acima se espelha em muitas outras na vida real: apesar de não serem simpáticas e causarem prejuízos, temos que conviver com filas na vida real visto ser antieconômico superdimensionar um sistema para que nunca existam filas. O que se tenta obter é um balanceamento adequado que permita um atendimento aceitável pelo menor custo.

1.5 Aspectos Históricos

A modelagem de sistemas pode ser feita por duas abordagens inteiramente diferentes entre si: Teoria das Filas e Simulação. A Teoria das Filas é um método analítico que aborda o assunto através de fórmulas matemáticas. Já a simulação é uma técnica que, usando o computador digital, procura montar um modelo que melhor represente o sistema em estudo.

1.5.1 Teoria das Filas

A abordagem matemática de filas se iniciou no princípio deste século (1908) em Copenhague, Dinamarca, através de A. Kendall Erlang, considerado o pai da Teoria das Filas, quando trabalhava em uma companhia telefônica estudando o problema de redimensionamento de centrais telefônicas. Foi somente a partir da segunda guerra mundial que a teoria foi aplicada a outro problema de filas. Apesar dos enormes progressos alcançados pela teoria, inúmeros problemas não são adequadamente resolvidos por causa de complexidades matemáticas.

1.5.2 Simulação

Com o surgimento do computador na década de 50, a modelagem de filas pode ser analisada pelo ângulo da simulação, em que não mais se usam fórmulas matemáticas, mas apenas tenta-se imitar o funcionamento do sistema real. As linguagens de simulação apareceram na década de 60 e hoje, graças aos microcomputadores, podem ser facilmente usadas. A técnica de simulação visual, cujo uso se deu a partir da década de 80, por causa de sua maior capacidade de comunicação, teve uma aceitação surpreendente. Além disso, por apresentar um menor nível de complexidade, seu uso também cresceu enormemente. O ensino desta técnica ainda se concentra em escolas de graduação, mas já tem havido iniciativas em ensino de segundo grau (cursos técnicos). Algumas linguagens são mundialmente conhecidas como Arena, ProModel, Automod, Taylor, Gpss, Gasp, Simscrip, Siman, etc.

1.6 Aplicações de Modelagem de Sistemas

A modelagem de sistemas tem inúmeras aplicações, no mundo atual, nas áreas mais diversas, que vão desde produção em uma manufatura até o movimento de papéis em um escritório. Costuma-se dizer que “tudo que pode ser descrito pode ser simulado”.

1.6.1 Linhas de Produção

Esta é a área que tem apresentado a maior quantidade de aplicações de modelagem. Inúmeros cenários se encaixam neste item, desde empresas manufatureiras até minerações. Os seguintes casos podem ser analisados:

- a) Modificações em sistemas existentes, tais como as produzidas pela expansão da atual produção, pela troca de equipamentos, ou pela adição de novos produtos, que vão afetar a dinâmica do atual processo. Pode-se antecipar onde serão formados os gargalos oriundos de modificações no sistema existente. Pela introdução de modificações apropriadas (tais como modificações no fluxo, alterações na programação das atividades, ou pela adição de novas facilidades), após algumas tentativas, pode-se chegar ao melhor modelo que incorpore as modificações requeridas.
- b) Um setor de produção totalmente novo pode ser planejado, obtendo-se o melhor fluxo dentro dele.
- c) A melhor política de estoques pode ser obtida por meio de simulação. O modelo deve incluir a função “solicitação de material” e a função “atendimento pelos fornecedores”. Como resultado se obtém o “ponto de pedido” e a “quantidade do pedido”.

1.6.2 Transportes

A modelagem de filas tem sido usada no transporte ferroviário, rodoviário, marítimo e no transporte por elevadores.

No transporte ferroviário o pátio de consertos e serviços apresenta problemas interessantes, que incluem o número e localização dos desvios e alocação de máquinas de serviço (com base em uma tabela de trens e carros a serem removidos ou adicionados), além da tabela de horários de trens diretos que passam pelo local. Por outro lado, o sistema ferroviário pode ser analisado como um todo, com o objetivo de minimizar o movimento de carros vazios.

No transporte marítimo e aéreo as aplicações se referem à confecção da tabela de horários e dimensionamento de portos e aeroportos.

No modelo rodoviário é possível dimensionar um pedágio ou estabelecer o melhor esquema do fluxo de veículos pelas ruas de uma cidade, com as durações dos semáforos, de modo a melhorar o serviço, agilizando o sistema e, conseqüentemente, diminuindo os gastos com combustível.

No modelo de elevadores é possível minimizar o tempo de espera e o custo de movimentação dos elevadores, pois quanto mais paradas ocorrerem entre andares, maior o custo. A partir da distribuição de chegada de pessoas aos vários andares, juntamente com seus destinos, é possível utilizar um modelo para determinar o número de elevadores em funcionamento para atender um dado padrão de serviço.

1.6.3 Comunicações

Uma ampla variedade de problemas de comunicação pode ser analisada pela modelagem de filas. A configuração ótima de uma rede de comunicações pode ser modelada. Informações sobre o tempo de resposta e sobre chamadas perdidas podem ser obtidas. Regras para análise de rotas alternadas podem ser comparadas e um estudo econômico pode avaliar o valor de concentradores, canalizadores de linha, etc. Empresas de telefonia podem fazer proveitosos usos desta técnica no estudo de seus complexos sistemas de comunicações.

1.6.4 Bancos, Supermercados, Escritórios, etc.

Pela simulação pode-se dimensionar o número de caixas de modo que as filas se mantenham abaixo de um valor especificado. Pode-se também avaliar o uso de caixas especiais tais como “caixas rápidos” dos supermercados. No caso de bancos o uso de “fila única” pode trazer um melhor atendimento aos clientes, apesar de poder assustar pelo tamanho que geralmente costuma ter.

1.6.5 Confiabilidade

A confiabilidade (ou disponibilidade) de um sistema complexo freqüentemente deve satisfazer rigorosas necessidades. Isto é bastante válido para sistemas militares ou de computadores on line. A simulação é uma excelente ferramenta para se obter uma medição quantitativa da confiabilidade do sistema se as características dos componentes individuais são conhecidas. Especificamente o tempo médio de falha e o tempo médio de reparo de cada componente devem ser conhecidos, em adição ao tempo necessário para substituir cada componente. Um planejamento de manutenção preventiva pode ser também construído através da simulação. Através de diversas tentativas no modelo, os componentes vitais de estoque podem ser determinados. Isto pode ser feito para diferentes requisitos de disponibilidade do sistema e, então, obtém-se a relação entre disponibilidade e custo total. A validade de duplicação de certos componentes do sistema também pode ser testada. Modelos desta natureza têm sido usados para testes dos mais diversos sistemas, desde sistemas de processamento de dados até esquadrões aéreos.

1.6.6 Processamento de Dados

A modelagem de filas tem sido amplamente usada pelas empresas que desenvolvem computadores e pelas universidades de modo a se medir a produtividade ou o tempo de resposta de um certo sistema de computadores e terminais. A área de teleprocessamento possui inúmeras opções de uso. Uma outra área que está se tornando popular dentro da comunidade de informática é o estudo de performance e de capacidade pois através dele pode-se identificar gargalos e indicar opções de crescimento.

2 CONCEITOS BÁSICOS (I)

2.1 Elementos de uma Fila

Na figura 1 ve-se os elementos que compõem uma fila. Nela tem-se que, de uma certa **população**, surgem **clientes** que formam uma **fila** e que aguardam por algum tipo de **serviço**. O termo cliente é usado de uma forma genérica e pode designar tanto uma pessoa, um navio ou um lingote. Como sinônimo de cliente usa-se também o termo “**transação**” ou “**entidade**”. A **atendimento** é constituído de um ou mais servidores (que podem ser chamados de **atendentes** ou **canais de serviço**) e tanto pode designar um barbeiro, um cais de atracação ou uma máquina de lingotamento.

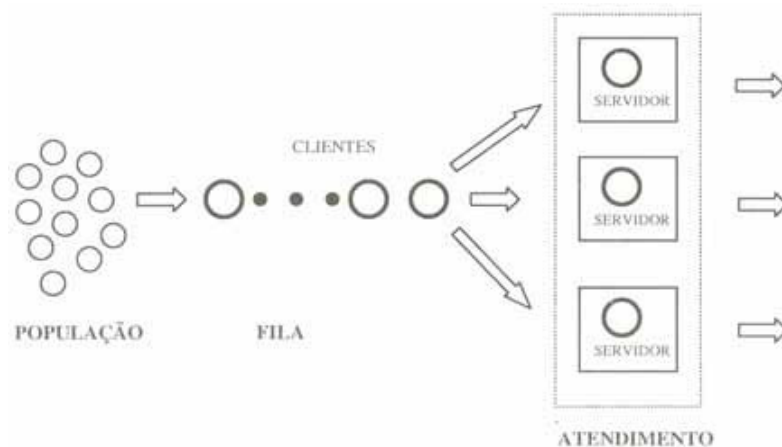


Figura 1 – Elementos de uma fila

2.2 Características de Uma Fila

Antes de observar uma fila em funcionamento, deve-se conceituar melhor alguns termos da Teoria das Filas.

2.2.1 Clientes e Tamanho da População

Um cliente é proveniente de uma população. Quando a população é muito grande (dizemos infinita para efeitos práticos), a chegada de um novo cliente a uma fila não afeta a taxa de chegada de clientes subsequentes e concluímos dizendo que as chegadas são independentes. Como exemplo, citamos o funcionamento de um metrô. Quando a população é pequena o efeito existe e pode ser considerável. Como exemplo extremo, podemos citar uma mineração, na qual uma carregadeira carrega minério em caminhões que chegam. Se existem 3 caminhões e, se ocorrer que todos eles estejam na fila da carregadeira, então não chegará nenhum outro caminhão.

2.2.2 Processo de Chegada

Considerando como exemplo um posto de pedágio com 5 atendentes. Pode-

se constatar, por exemplo, que o processo de chegada entre 7 e 8 horas da manhã pode ser definido por 20 automóveis por minuto ou 1 automóvel a cada 3 segundos. Trata-se de um valor médio, pois não significa que em todo intervalo de 1 minuto chegarão 20 automóveis. Em alguns intervalos de 1 minuto pode-se constatar a chegada de 10, 15, 25 ou até 30 automóveis. Conseqüentemente, o intervalo de 3 segundos entre chegadas não é rígido e podemos constatar valores, por exemplo, desde zero segundos (2 veículos chegando juntos) até 20 segundos. O número fornecido, 3 segundos, representa, assim, o intervalo médio entre chegadas no período de 7 as 8 horas da manhã.

Resumindo as afirmações acima pode-se quantificar o processo de chegada dizendo que a taxa média de chegada é de 20 veículos por minuto ou que o intervalo médio entre chegadas é de 3 segundos.

Poderia ser encontrado um outro sistema de filas que possui exatamente os mesmos valores médios acima citados, mas com diferentes variações no entorno da média (por exemplo, uma situação em que se observa que os intervalos entre chegadas estão entre 0 e 10 segundos). Este sistema, conforme veremos, terá um comportamento diferente do primeiro e concluí-se dizendo que não basta apenas fornecer os valores médios, é necessário também mostrar como os valores se distribuem em torno da média. Deve-se, pois, lançar mão de uma distribuição de freqüência, tal como a distribuição normal, a de Poisson, a exponencial, etc.

Um tipo raro de processo de chegada é o regular, ou seja, aquele em que não existe nenhuma variação entre valores para os intervalos entre chegadas. Nesta situação, se disser que o intervalo entre chegadas é de 10 segundos, tem-se que rigorosamente a cada 10 segundos chega um novo cliente. Esta situação ocorre apenas em processos altamente automatizados.

Quando se estudam filas, o ritmo de chegada é uma importante variável randômica. Para quantificar esta variável se usa a letra grega λ para significar ritmo de chegada e se usa IC para intervalo entre chegadas. Assim, no primeiro exemplo acima teremos:

$$\lambda = 20 \text{ clientes por minuto}$$

$$IC = 3 \text{ segundos}$$

2.2.3 Processo de Atendimento

Continuando no exemplo do pedágio e observando um atendente em serviço, pode-se constatar, por exemplo, que ele atende 6 veículos por minuto ou que gasta 10 segundos para atender um veículo. Estes valores são médios e, para descrevê-los corretamente, deve-se também lançar mão da distribuição de probabilidades. Aqui também é rara a existência prática de atendimento regular, ou seja, existe um único valor (sem variação) para a duração do atendimento.

O processo de atendimento é também quantificado por uma importante variável randômica. A letra grega μ é usada para significar ritmo de atendimento e se usa TA para tempo ou duração do serviço ou atendimento. Assim, no exemplo acima tem-se que:

$$\mu = 6 \text{ clientes por minuto}$$

$$TA = 10 \text{ segundos por cliente}$$

2.2.4 Número de Servidores

O mais simples sistema de filas é aquele de um único servidor que pode atender um único cliente de cada vez. Conforme aumente o ritmo de chegada, pode-se manter a qualidade do serviço aumentando convenientemente o número de servidores. Esta é, portanto, uma das características de uma fila que pode-se utilizar para modelar um sistema de filas.

2.2.5 Disciplina da Fila

Trata-se da regra que define qual o próximo a ser atendido e o comum é que o primeiro da fila é atendido ou, de uma maneira mais ampla, o “primeiro a chegar é o primeiro a ser atendido” (em inglês, diz-se FIFO: First In First Out). Outras disciplinas podem existir tais como “último a chegar primeiro a ser atendido” (em inglês diz-se LIFO: Last In First Out), serviço por ordem de prioridade, serviço randômico, etc.

2.2.6 Tamanho Médio da Fila

Esta é a característica da fila mais considerada ao se defrontar com a opção de escolher uma fila. O ideal é a situação de chegar é ser atendido (fila zero). Quando a fila é de um tamanho razoável (por exemplo, 10 elementos) intuitivamente sabe-se que o tempo de espera na fila será longo. O tamanho da fila não é constante e, posteriormente, ver-se-á que, quando os ritmos médios de chegada e de atendimento são constantes, o tamanho da fila oscila em torno de um valor médio.

2.2.7 Tamanho Máximo da Fila

Quando os clientes devem esperar, alguma área de espera deve existir (por exemplo: as cadeiras de uma barbearia). Observa-se, nas vidas reais, que os sistemas existentes são dimensionados para uma certa quantidade máxima de clientes em espera, sendo este dimensionamento geralmente feito com base em experiência real. Quando existe um crescimento na demanda, se faz uma ampliação também baseada na experiência com o manuseio do referido sistema. Observam-se, também, casos em que um novo cliente que chega pode ser recusado devendo tentar novamente em outro instante (exemplo: tentativa de conseguir uma linha telefônica, recebendo o sinal de “ocupado” ou de que não há linha disponível).

As considerações anteriores se referem ao que chamamos de “tamanho máximo da fila”, importante etapa do projeto de um sistema de filas. Tanto pode se referir a uma área de espera para caminhões que vão se abastecer de combustível em uma refinaria como a um buffer onde transações de consulta a um computador devem esperar antes de serem atendidas.

2.2.8 Tempo Médio de Espera na Fila

Esta é a característica capaz de causar irritação quando se está em uma fila de espera. O ideal é que não exista espera, mas esta nem sempre é a melhor situação do ponto de vista econômico. Se entrar numa fila com 10 pessoas à nossa frente, o tempo de espera será igual ao somatório dos tempos de atendimento de cada um dos clientes da nossa frente ou, possivelmente, será igual a 10 vezes a duração média de atendimento. Tal como o tamanho médio da fila, o tempo médio

de espera depende dos processos de chegada e de atendimento.

2.3 Variáveis Randômicas

Pode-se observar, em referencia a filas, a utilização de variáveis randômicas. Assim, para as principais variáveis existe um valor médio e uma distribuição de probabilidades, que mostra as chances de ocorrências dos valores. Quando se afirma que o tamanho médio de uma fila é de 5 clientes, não está dizendo que o tamanho da fila é sempre de 5 clientes. Em diferentes momentos de observação você pode, por exemplo, constatar que o tamanho é de 10, 7, 3 ou que não existe fila. O valor médio 5 significa uma média aritmética ponderada dos tamanhos da fila durante consecutivos intervalos de tempo.

2.4 Observando a Dinâmica de Uma Fila

Imagine agora, o observador comodamente instalado em uma poltrona dentro de um banco, com a finalidade de observar o funcionamento da fila formada por pessoas que desejam um novo talão de cheques. No período de meia hora verificou-se que chegaram ao sistema 12 pessoas.

Os intervalos entre chegadas foram (valores em minutos):

Cliente	1	2	3	4	5	6	7	8	9	10	11	12
Intervalo	2	3	3	3	5	0	1	5	1	4	1	2
Momento	2	5	8	11	16	16	17	22	23	27	28	30

O valor zero acima significa que o sexto cliente chegou junto com o quinto. O valor médio dos dados acima é 2,5 minutos e, portanto, o sistema acima funcionou com um ritmo médio de 24 chegadas por hora.

Por outro lado, os dados anotados para cada atendimento são os seguintes (valores em minutos):

Cliente	1	2	3	4	5	6	7	8	9	10	11	12
Duração	1	2	1	1	3	2	1	4	2	3	1	3

O valor médio dos dados acima é 2,0 minutos e, portanto, pode-se dizer que o servidor tem uma capacidade de atender 30 clientes por hora. Finalmente, o sistema funcionou conforme a Figura 2 e por ela verifica-se que os clientes de números 6, 7, 9, 10, 11 e 12 tiveram que esperar em fila.

Cliente	1	2	3	4	5	6	7	8	9	10	11	12
Tempo em Fila	0	0	0	0	0	3	4	0	3	1	3	2

Analisando a Figura 2, pode-se concluir que:

Tempo Médio na Fila = $(3 + 4 + 3 + 1 + 3 + 2)/12 = 16/12 = 1,33$ minutos

Número Médio na Fila = $(3 + 4 + 3 + 1 + 3 + 2)/35 = 16/35 = 0,46$ clientes

Vemos, ainda, que o sexto cliente chegou no 16º minuto, ficou na fila durante 3 minutos e seu atendimento durou 2 minutos.



Figura 2 – Funcionamento da Fila do Posto Bancário

Imaginando que o mesmo problema fosse proposto ao aluno da seguinte forma: clientes chegam a um posto de serviço a um ritmo de 24 chegadas por hora (ou 1 cliente a cada 2,5 minutos) e são atendidos por um servidor capaz de atender 30 clientes por hora (ou 2,0 minutos para cada cliente). Pergunta: Haverá fila?

De posse destes dados, a tendência do aluno é inferir que o sistema se comportará conforme a Figura 3, na qual tanto o processo de chegada como o de atendimento é regular e, portanto, não haverá formação de filas. Processos como este são raros na vida real. Como se pode deduzir, a existência de filas ocorre em decorrência do fato de que os processos não são regulares e a aleatoriedade tanto ocasiona filas como ocasiona também longos períodos de inatividade para o servidor.



Figura 3 – Funcionamento Regular de Uma Fila

2.5 Sistemas Estáveis

A abordagem matemática de filas pela Teoria das Filas exige que exista estabilidade no fluxo de chegada e no processo de atendimento, ou seja, os valores de λ e μ se mantêm constantes no tempo. Por exemplo, observando o funcionamento de um banco, poderíamos verificar que o fluxo de chegada de clientes varia durante o dia:

Período	10 às 12	12 às 14	14 às 16
Fluxo	Médio	Alto	Médio

Ou seja, não existe estabilidade para o ritmo de chegada no período de 10:00 às 16:00, e, portanto, não se pode analisar seu funcionamento pela teoria das Filas, a menos que usemos alguns artifícios como, por exemplo, retalhar o período global em períodos parciais. Infelizmente, isto torna ainda mais complexa a abordagem pela Teoria das filas.

Em fábricas que funcionam 24 horas ininterruptamente tem-se geralmente uma situação estável, não sendo necessário à redução do intervalo. Outra exigência para que o processo seja estável é que os atendentes sejam capazes de atender ao fluxo de chegada. No caso de “uma fila e um atendente” isto significa dizer que $\mu > \lambda$ (a capacidade de atendimento é maior que o ritmo de chegada) e, caso isto não ocorra, resulta que o tamanho da fila aumentará infinitamente.

Conforme visto no item anterior, é possível a formação de filas em sistemas nos quais $\mu > \lambda$. As pessoas geralmente têm dificuldades em entender este fato pois seu raciocínio está quase sempre preso a processos regulares, em que não há flutuações em torno da média (ver Figura 3).

Em processos randômicos estáveis, λ e μ representam valores médios e para atender a razão da ocorrência de filas é necessário ter em mente que sempre é possível a ocorrência de fatos “ruins” como, por exemplo:

- Em um determinado instante podem chegar mais clientes que a capacidade de atendimento daquele momento, gerando filas temporárias;
- O atendimento de um determinado cliente pode ser muito mais moroso que a média, obrigando os clientes que chegam em seguida a ficarem em fila.

Em sistemas estáveis, todas as características randômicas das filas se mantêm estáveis o tempo todo, significando que oscilam em torno de um valor médio como, por exemplo, o tamanho médio, o tempo médio de espera, o tempo médio de atendimento, etc.

2.6 O Tamanho de Amostra

Um sistema estável em estudo mostrará sempre os mesmos resultados desde que adequadamente analisado. Para isto, a escolha de um correto tamanho de amostra é fundamental. A não observância deste item pode confundir, por produzir diferentes valores para a mesma variável. Por exemplo, em um sistema estável podemos ter um tempo médio de espera na fila de 5 minutos. Para se chegar a esta conclusão foi necessário observar o funcionamento do sistema durante um longo período, no qual inúmeros clientes foram atendidos. Se for observado apenas o atendimento de uns poucos clientes, pode-se encontrar um valor bastante diferente para o tempo médio de espera na fila.

A análise de sistemas não estáveis é mais bem efetuada por simulação.

2.7 Opções de Dimensionamento: o Tipo da Fila

Quando se deseja dimensionar um sistema, pode-se escolher entre diversas opções para o atendimento:

- Uma única fila e um único servidor;
- Uma única fila e diversos servidores;
- Diversas filas e diversos servidores;
- Filas especiais;
- Alteração dinâmica no sistema de atendimento.

A escolha entre as opções acima vai depender das características do sistema em estudo, pois o que pode ser ótimo em uma situação pode ser péssimo em outra ou, então, inadequado.

2.8 Gerenciando Filas

Estudam-se filas para poder modificar sistemas nos quais existem gargalos, com o objetivo de prestar um melhor atendimento ou uma redução de custos. O conhecimento do comportamento de uma fila, quando alteramos algumas de suas características, pode ser fundamental para atingirmos os objetivos citados de qualidade de atendimento e custos.

Em situações nas quais o tempo de atendimento pode variar dentro de uma larga faixa de valores não recomenda o uso de diversas filas. É o caso de bancos,

correios, etc, em que sempre pode ocorrer que alguns clientes apresentem uma carga de serviço muito grande e, portanto, o tempo de atendimento para eles será exageradamente maior que a média. Aqui uma fila única com diversos atendentes é a melhor solução.

Em outras situações, é conveniente modificar dinamicamente a quantidade de atendentes conforme aumente ou diminua o fluxo de chegada de clientes. Bancos têm usado estes expedientes, tornando disponíveis atendentes extras no horário de pico.

Às vezes a fila é impraticável, como no caso de supermercados. Aqui a existência de “caixas expressos”, para clientes com poucos itens de compra, representa uma maneira de se prestar um bom serviço, além de conquistar clientes que, do contrário, não se sujeitariam a filas morosas para adquirir poucos itens.

2.9 Exercícios

1 - Considere um sistema em que navios chegam a um porto para carregar algum produto. Abaixo estão anotados os valores de intervalos entre chegadas (em horas) para 20 navios:

Cliente	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Intervalo	10	2	13	7	2	8	8	8	10	9	1	14	14	1	10	9	9	9	8	14

A duração da carga (em horas) de cada navio são as seguintes:

Cliente	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Duração	5	5	3	3	6	7	6	8	2	5	8	8	8	3	4	3	3	4	5	5

Pede-se:

- O intervalo médio entre chegadas;
- A duração média da carga;
- Monte o desenho do funcionamento do sistema acima (veja Figura 2);
- Calcule o tamanho médio da fila;
- Calcule o tempo médio de espera na fila.

2 - Considere um sistema em que clientes chegam a uma lanchonete. Abaixo estão anotados os valores de intervalos entre chegadas (em minutos) para 10 clientes:

Cliente	1	2	3	4	5	6	7	8	9	10
Intervalo	4	3	3	3	3	4	2	1	2	5
Duração	2	2	4	3	1	5	1	5	3	4

Pede-se:

- a) O intervalo médio entre chegadas;
- b) A duração média da carga;
- c) Monte o desenho do funcionamento do sistema acima (veja Figura 2);
- d) Calcule o tamanho médio da fila;
- e) Calcule o tempo médio de espera na fila.

3 - Considere um sistema em que clientes chegam a lava-jato. Abaixo estão anotados os valores de intervalos entre chegadas (em minutos) para 10 clientes:

Cliente	1	2	3	4	5	6	7	8	9	10
Intervalo	3	3	3	2	2	1	5	4	4	3
Duração	3	3	2	1	2	4	1	5	5	4

Pede-se:

- a) O intervalo médio entre chegadas;
- b) A duração média da carga;
- c) Monte o desenho do funcionamento do sistema acima (veja Figura 2);
- d) Calcule o tamanho médio da fila;
- e) Calcule o tempo médio de espera na fila.

3 CONCEITOS BÁSICOS (II)

3.1 Introdução

O objetivo deste capítulo é continuar com as considerações conceituais do capítulo anterior, agora com um enfoque matemático, no qual apresentam-se as chamadas “variáveis randômicas fundamentais”.

3.2 Variáveis Randômicas Fundamentais

Considera-se o sistema de filas da Figura 4, em situação estável, na qual clientes chegam e entram em fila, existindo M servidores para atendê-los. Seja λ o ritmo médio de chegada e μ o ritmo médio de atendimento de cada atendente. Dentre as variáveis randômicas que estudaremos, algumas serão freqüentemente citadas e as chamaremos de “variáveis randômicas fundamentais”. É o que mostra a Figura 4:



Figura 4 – Localização das Variáveis

- Variáveis Referentes ao Sistema
 TS = Tempo Médio de Permanência no Sistema
 NS = Número Médio de Clientes no Sistema
- Variáveis Referentes ao Processo de Chegada
 λ = Ritmo Médio de Chegada
 IC = Intervalo Médio entre Chegadas
 Por definição: $IC = 1 / \lambda$
- Variáveis Referentes à Fila

TF = Tempo Médio de Permanência na Fila

NF = Número Médio de Clientes na Fila

- Variáveis Referentes ao Processo de Atendimento

TA = Tempo Médio de Atendimento ou de Serviço

M = Quantidade de atendentes

NA = Número Médio de Clientes que estão sendo atendidos

μ = Ritmo Médio de Atendimento de cada Atendente

Por definição: $TA = 1 / \mu$

3.2.1 Relações Básicas

Existem duas relações óbvias entre as variáveis randômicas mostradas na Figura 4:

$$NS = NF + NA$$

$$TS = TF + TA$$

Pode-se demonstrar também que $NA = \lambda / \mu = TA / IC$. Portanto:

$$NS = NF + NA = NF + \lambda / \mu = NF + TA / IC$$

3.2.2 Taxa de Utilização dos Atendentes

Para o caso de “uma fila / um atendente”, chama-se de taxa de utilização do atendente a expressão:

$$\rho = \lambda / \mu$$

na qual λ = ritmo médio de chegada e μ = ritmo médio de atendimento.

No caso de uma fila / vários atendentes, a expressão se torna:

$$\rho = \lambda / (M \cdot \mu)$$

em que M é o número de atendentes.

Assim, ρ representa a fração média do tempo em que cada servidor está ocupado. Por exemplo, com um atendente, se chegam 4 clientes por hora e se o atendente tem capacidade para atender 10 clientes por hora, logo a taxa de utilização é 0,40 e pode-se também afirmar que o atendente fica 40% do tempo ocupado e 60% do tempo livre (esta afirmação é intuitiva mas pode ser matematicamente demonstrada).

Visto que o estudo é com sistemas estáveis (os atendentes sempre serão capazes de atender ao fluxo de chegada) tem-se sempre que $\rho < 1$. Quando $\rho = 1$ o atendente trabalhará 100% do tempo (e estranhos fatos ocorrerão ...).

3.2.3 Intensidade de Tráfego ou Número Mínimo de Atendentes

Chama-se de intensidade de tráfego a expressão:

$$i = \lceil \lambda / \mu \rceil = \lceil TA / IC \rceil$$

em que i é o próximo valor inteiro que se obtém (ou seja, o valor absoluto) e é medido em “erlangs” em homenagem a A.K. Erlang. Na prática, i representa o

número mínimo de atendentes necessário para atender um dado fluxo de tráfego. Por exemplo, se $\lambda = 10$ clientes / hora e $TA = 3$ minutos (ou $\mu = 20$ clientes / hora) tem-se que $\lambda / \mu = 0,5$, ou $i = 1$, e concluí-se dizendo que 1 atendente é suficiente para o caso. Se o fluxo de chegada aumentar para $\lambda = 50$ clientes / hora, tem-se que $\lambda / \mu = 2,5$, ou $i = 3$, isto é, necessita-se de no mínimo 3 atendente. Na indústria telefônica esta variável é bastante utilizada ao se referir a tráfego em troncos telefônicos.

3.2.4 Fórmulas e Little

J.D.C. Little demonstrou que, para um sistema estável de filas, temos:

$$NF = \lambda \cdot TF$$

$$NS = \lambda \cdot TS$$

Estas fórmulas são muito importantes, pois fazem referências a quatro das mais importantes variáveis randômicas de um sistema de filas: NS, NF, TS e TF. Por exemplo, se além de λ e μ é conhecido TS, pode-se obter as outras variáveis:

$$NS = \lambda \cdot TS$$

Se $TA = 1 / \mu$

Portanto, $TF = TS - TA = TA - 1 / \mu$

Finalmente, $NF = \lambda \cdot TF$

É importante salientar que todas as fórmulas acima independem da quantidade de servidores e do modelo de fila, pois se trata de fórmulas fundamentais básicas.

3.2.5 Resumo das Fórmulas

Nome	Fórmula
Intervalo entre chegadas	$IC = 1 / \lambda$
Tempo de Atendimento	$TA = 1 / \mu$
Taxa de Utilização dos Atendentes	$\rho = \lambda / (M \cdot \mu)$
Intensidade de Tráfego	$i = \lambda / \mu = TA / IC $
Relação entre Fila, Sistema e Atendimento	$NS = NF + NA$ $NA = \lambda / \mu$ $NS = NF + \lambda / \mu = NF + TA / IC$ $TS = TF + TA$ $NA = \rho = \lambda / (M \mu)$
Fórmulas de Little	$NF = \lambda \cdot TF$ $NS = \lambda \cdot TS$
Duração do Ciclo	$Ciclo = Qtdd / \lambda$ $Ciclo = TFS + TS$

Figura 5 – Resumo das fórmulas

3.3 Exemplos

Exemplo 1: Em uma fábrica observou-se o funcionamento de um dado setor, em que Ritmo médio de chegadas (λ) é de 20 clientes / hora, o Ritmo médio de atendimento (μ) é de 25 clientes / hora e o Tempo médio de permanência no sistema (TS) é de 0,3 hora. Pede-se o tamanho médio da fila (NF).

Solução:

$$TA = 1 / \mu = 1 / 25 = 0,04$$

$$TF = TS - TA = 0,3 - 0,04 = 0,26$$

$$NF = \lambda \times TF = 20 \times 0,26 = 5,2 \text{ clientes}$$

Exemplo 2: Para o mesmo sistema acima, calcular o Número médio de clientes no sistema (NS) e o Número médio de clientes sendo atendidos (NA).

Solução:

$$NS = \lambda \times TS = 20 \times 0,3 = 6 \text{ clientes}$$

$$NA = NS - NF = 6 - 5,2 = 0,8 \text{ clientes}$$

Exemplo 3: Em uma mineração verificou-se que o tempo médio de permanência no sistema (TS) dos caminhões junto às carregadeiras é de 3 minutos e que, em média, existem 6 caminhões (NS) no setor. Qual a taxa de chegada de caminhões? Veja Figura abaixo.

Solução:

$$NS = \lambda \times TS \text{ ou } \lambda = NS / TS$$

Pela lei de Little:

$$\lambda = 6 / 3 = 2 \text{ chegadas por minuto}$$

Logo:

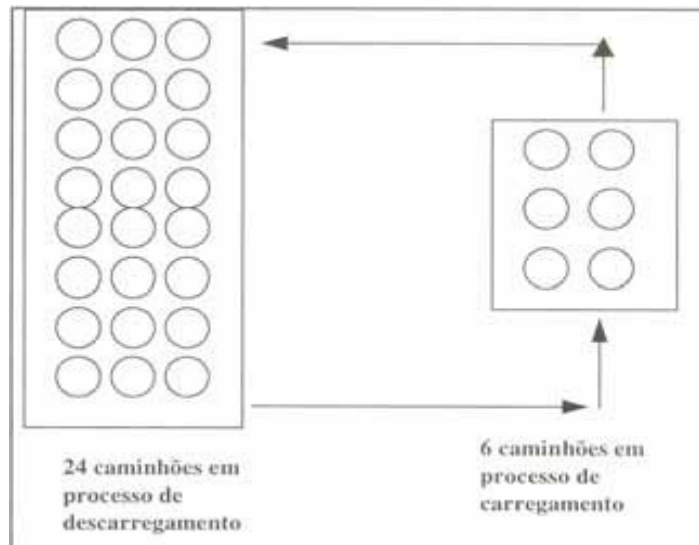


Figura 6 – Caminhões em uma Mineração

Exemplo 4: No mesmo sistema acima, existindo um total de 30 caminhões em serviço, qual a duração de um ciclo?

Solução:

Chamamos de ciclo ao tempo gasto para que todos os caminhões “passem” pela carregadeira uma vez. Ao final de um ciclo o sistema terá atendido uma vez a cada um dos 30 caminhões.

$$\text{Duração do ciclo} = (\text{quantidade de caminhões}) / \lambda$$

$$\text{Duração do ciclo} = 30 / \lambda = 30 / 2 = 15 \text{ minutos}$$

Exemplo 5: No mesmo sistema acima, qual o tempo médio para o processo completo de descarregamento (ou TFS: Tempo Fora do Sistema)?

Solução:

Um ciclo corresponde à soma do tempo dentro do sistema (TS = 3) com o tempo de descarregamento (TFS). Logo:

$$\text{TFS} + \text{TS} = 15 \text{ caminhões}$$

$$\text{TFS} = 15 - 3 = 12 \text{ caminhões}$$

3.4 Postulados Básicos

Apresenta-se a seguir alguns postulados básicos que se aplicam a quaisquer sistema de filas nos quais existe estabilidade ou seja, λ é menor que μ em todas as estações de trabalho (o ritmo médio de chegada é menor que o ritmo médio de atendimento).

- a) Em qualquer sistema estável, o fluxo que entra é igual ao fluxo que sai.
No sistema a seguir, o fluxo de entrada (λ) é igual ao fluxo de saída (λ):



Figura 7 – Sistema estável

- b) Em um sistema estável, o fluxo de entrada se mantém nas diversas seções do sistema.

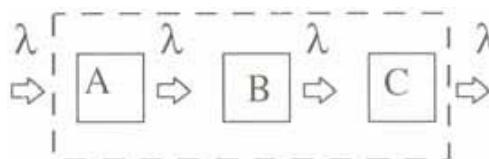


Figura 8 – Fluxo de entrada em sistema estável

c) Em um sistema estável, a junção de fluxos equivale às suas somas.

No sistema abaixo, o fluxo que chega a C é: $\lambda_3 = \lambda_1 + \lambda_2$:

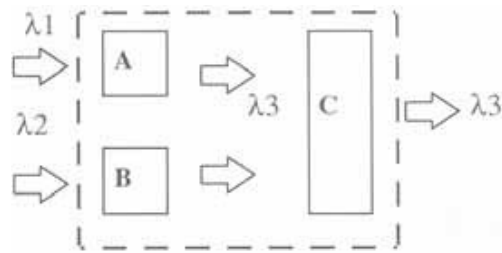


Figura 9 – Sistema estável e junção de fluxos de entrada

d) Em um sistema estável, o fluxo se desdobra aritmeticamente.

No sistema abaixo, se 80% do fluxo que sai de A se dirige para B, podemos afirmar que o fluxo que chega a B é de $\lambda_2 = 0,8 \times 20 = 16$ clientes / minuto.

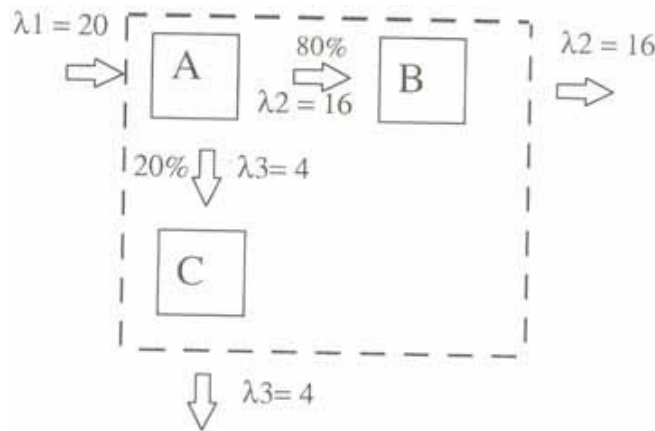


Figura 10 – Fluxo de saída desdobrado em sistema estável

3.5 Exercícios

- 1) Em uma pizzaria que faz entregas em casa, chegam, em média, 4 entregadores por minuto para pegar o produto a ser entregue. Sabe-se, ainda, que o número médio de entregadores dentro da pizzaria é de 6 (NS). Qual o tempo médio no sistema?
- 2) No mesmo sistema anterior, existem 40 entregadores. Qual o tempo médio da entrega (TFS)?
- 3) Em um sistema de computação tem-se:
 - Tempo médio de pensar e fornecer dados (TFS) = 15 segundos;
 - Quantidade de terminais ativos = 40;
 - Taxa de chegada de transações = 2 por segundo;

Pede-se o tempo de resposta do computador (TS).

- 4) Em uma mineração temos 12 caminhões efetuando um ciclo no qual consomem 4 minutos entre fila e carregamento pela escavadeira (TS) e, a seguir, gastam 8 minutos para levar a carga até o britador e voltar (TFS). Calcular λ e NS.
- 5) Em um sistema de computação temos 21 terminais. O tempo médio de resposta do computador (TS) é de 2 segundos e existem, em média, 6 transações (NS) dentro do sistema. Pede-se:
 - a) Qual a taxa de chegada de transações?
 - b) Qual a duração de um ciclo?
 - c) Qual o “tempo médio de pensar e fornecer dados” (TPS)?
- 6) Na figura 11, representativo do fluxo de peças em um setor de uma fábrica, calcule o fluxo de chegada em cada equipamento.

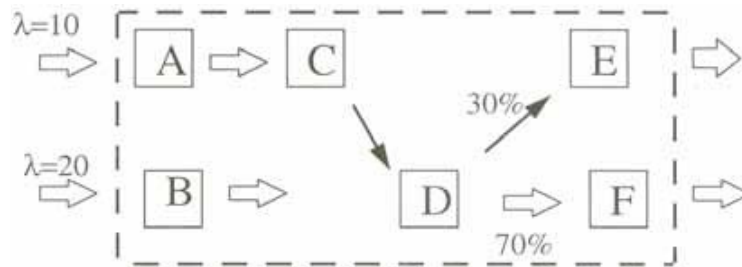


Figura 11 – Fluxos em sistema estável

4 Referencias bibliográfica

AURELIO, Marco. **Teoria das filas e das simulações** (apostila). Santo André: UNIA, 2004.

Apostila de filas.doc