

Filas: Conceitos Básicos (I)

2.1 - Introdução

Neste capítulo iremos iniciar a apresentação de conceitos básicos de filas.

2.2 - Elementos de Uma Fila

Na Figura 2.1 vemos os elementos que compõem uma fila. Nela temos 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**". O **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.

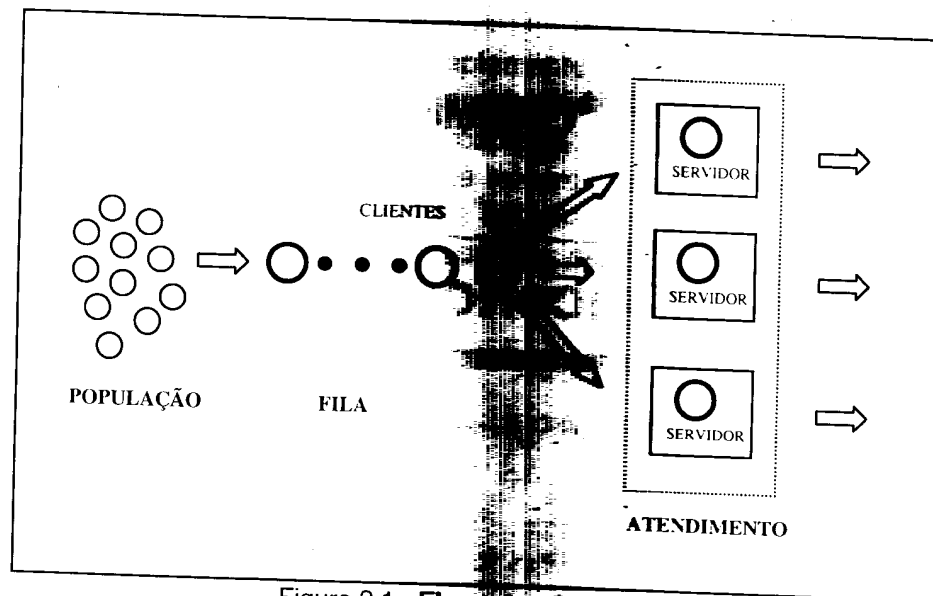


Figura 2.1 - Elementos de uma fila

2.3 - Características de Uma Fila

Antes de observar uma fila em **funcionamento**, iremos conceituar melhor alguns termos da Teoria das Filas.

2.3.1 - 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.

2.3.2 - Processo de Chegada

Consideremos como exemplo um posto de pedágio com 5 atendentes. Podemos 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 mesmo 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 podemos 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.

Poderíamos encontrar 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 concluímos dizendo que não basta apenas fornecer os valores médios, é necessário também mostrar como os valores se distribuem em torno da média. Devemos, pois, lançar mão de uma distribuição de frequência, 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 os valores para os intervalos entre chegadas. Nesta situação, se dissermos que o intervalo entre chegadas é de 10 segundos, teremos 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:

$$\begin{aligned}\lambda &= 20 \text{ clientes por minuto} \\ \text{IC} &= 3 \text{ segundos}\end{aligned}$$

2.3.3 - Processo de Atendimento

Continuando no exemplo do **pedágio** e observando um atendente em serviço, podemos 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**, devemos 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** teremos:

$$\mu = 6 \text{ clientes por minuto}$$
$$TA = 10 \text{ segundos por cliente}$$

2.3.4 - 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, podemos manter a **qualidade** do serviço aumentando convenientemente o número de **servidores**. Esta é, portanto, uma das características de uma fila que **podemos utilizar** para modelar um sistema de filas.

2.3.5 - 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.

2.3.6 - Tamanho Médio da Fila

Esta é, certamente, a característica da fila que mais consideramos ao nos defrontarmos com a opção de escolher uma fila. O ideal é a situação de chegar e ser atendido (fila zero). Quando a fila é de um tamanho razoável (digamos 10 elementos) intuitivamente sabemos que o tempo de espera na

fila será longo. O tamanho da fila não é constante e, posteriormente, veremos que, quando os ritmos médio de chegada e de atendimento são constantes, o tamanho da fila oscila em torno de um valor médio.

2.3.7 - Tamanho Máximo da Fila

Quando os clientes devem esperar, alguma área de espera deve existir (por exemplo: as cadeiras de uma barbearia). Observa-se, na vida real, 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.

2.3.8 - Tempo Médio de Espera na Fila

Esta é a característica capaz de nos causar irritação quando estamos 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 entrarmos 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 na 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.

2.4 - Variáveis Randômicas

Conforme você pôde observar, quando nos referimos a filas utilizamos 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 afirmamos que o tamanho médio de uma fila é de 5 clientes, não estamos **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.

2.5 - Observando a Dinâmica de Uma Fila

Imagine-se agora 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 você verificou 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, podemos dizer que o servidor tem uma capacidade de atender 30 clientes por hora. Finalmente, o sistema funcionou conforme a Figura 2.2 e por ela