TEORIA DAS FILAS (*Queueing Theory*)

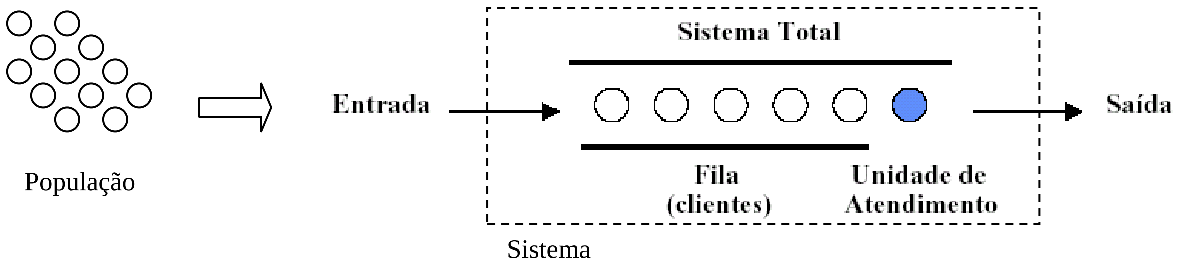
1. INTRODUÇÃO

A abordagem matemática das filas se iniciou em 1908, na cidade de Copenhague, Dinamarca. O pioneiro da investigação foi o matemático Agner Krarup Erlang (1909), quando trabalhava numa companhia telefônica, estudando o problema de redimensionamento de centrais telefônicas. Somente a partir da Segunda Guerra Mundial que a teoria foi aplicada a outros problemas de filas. Seu trabalho foi difundido por outros pesquisadores em diversos países europeus. Na década de 30, dentre as pesquisas nesta área, Andrey Kolmogorov, na Rússia, estudava um sistema com entrada de probabilidade de Poisson (Siméon Denis Poisson) e saída arbitrária em único ou múltiplo atendente.

A Teoria das Filas é uma das técnicas da Pesquisa Operacional, que trata de problemas de congestionamentos de sistemas, onde clientes solicitam alguns tipos de serviços. Esses serviços são limitados por restrições intrínsecas do sistema, que, devido a isso, podem causar filas.

Para melhor entendimento de um sistema de filas e seus componentes pode-se visualizar a figura 1 a seguir.

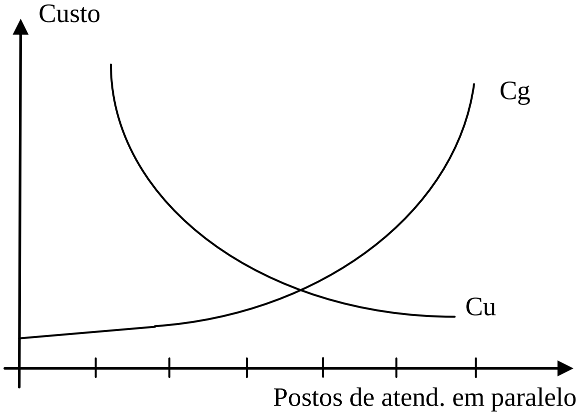
Fig.1. Exemplo de fila com seus componentes

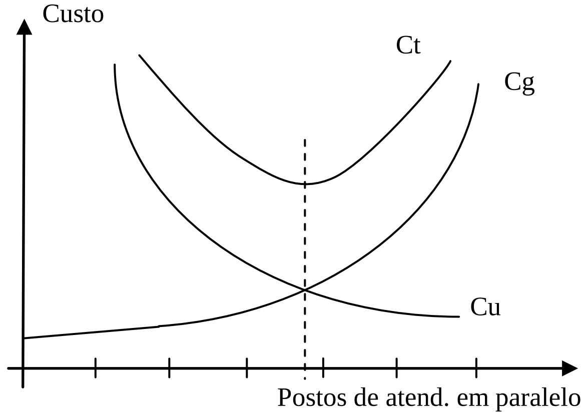


Existem vários tipos de configurações de filas. Por isso, a identificação do modelo que mais se adeque a realidade é fundamental para que a análise do desempenho do sistema seja correta. Para Fogliatti *et al.* (2007), ressalta que as medidas de desempenho têm duas abordagens: a do usuário e da gerência do sistema.

Quanto à visão do usuário, é fundamental a avaliação do tamanho médio da fila e os tempos médios na fila e no sistema. Para o gerente do sistema, compete avaliar os tempos médios do serviço prestado e de ociosidade do servidor. Sendo assim, para que esses atores estejam sendo observados na avaliação do desempenho do sistema, deve-se incluí-los em uma única função, principalmente quanto ao custo da configuração ideal da fila.

O gráfico adiante, exposto por Fogliatti *et al.* (2007) denota os custos associados à configuração pelas visões do usuário (Cu) e do gerente (Cg).





O gráfico da esquerda denota que o custo gerencial se eleva quando se configura um sistema com postos de atendimento em excesso. Em contraposição, poucos postos de atendimento provoca insatisfação do usuário e, em conseqüência, aumento do seu custo. Para esta análise em conjunto, necessita-se formalizar uma equação do custo total (Ct) que represente estes dois custos, a saber: Ct = aCu + bCg, onde *a* e *b* são constantes representativas de cada caso. A curva que representa o Ct está exposta no gráfico da esquerda e o seu ponto de vale indica a melhor configuração para ambos os atores, ou seja, entre 3 e 4 postos de atendimento.

Existem vários exemplos reais de sistemas de filas. Como forma de ilustração a tabela 1 lista quatro exemplos.

Tab.1. Exemplos de Sistemas de Filas

Situação	Processo de Entrada	Processo de Saída
Banco	Usuários chegando ao banco	Usuário atendido pelo caixa
Atendimento em pizzaria	**Pedido para entrega de pizza para o cliente**	**Pizzaria envia pizzas**
Banco de Sangue	Chegada de bolsa com sangue	Bolsa usada por paciente
Estaleiro de Navios	**Navio necessitando reparo é enviado para o estaleiro**	**Navio reparado volta para o mar**

2. DEFINIÇÕES IMPORTANTES

A seguir serão definidos alguns componentes e variáveis importantes para compreensão sobre os sistemas de filas.

- Tamanho da população - Tamanho do grupo que fornece os clientes. Para tamanhos maiores que 30, geralmente, considera-se que a população é infinita, ou ainda, que a chegada de um cliente não afetará significativamente a probabilidade da chegada de outro cliente. Quando a população for pequena, ou seja, menor que 30, o efeito existe e pode ser considerável.
- Clientes – São unidades da população que chegam para o atendimento, como por exemplo, pessoas, peças, máquinas, navios, automóveis etc..
- Fila (linha de espera) - Número de clientes esperando atendimento. Não inclui o cliente que está sendo atendido;

- Unidade de atendimento - Processo ou sistema que realiza o atendimento do cliente. Pode ser unidade única ou múltipla;
- Taxa de chegada dos clientes - Taxa (número de clientes / unid. tempo) segundo a qual os clientes chegam para serem atendidos. O valor médio da taxa de chegada é representado por λ (lambda). Como é raro um processo onde taxa de chegada dos clientes seja regular, ou seja, não existe nenhuma variação entre os valores para os intervalos entre chegadas, são adotadas distribuições de frequência (normal, Poisson, exponencial etc.) para representar o processo. O mesmo modelo com distribuição normal pode diferir significativamente em termos de resultado do que com uma distribuição de Poisson;
- Taxa de atendimento dos clientes - Taxa (número de clientes / unid. tempo) segundo a qual um servidor pode efetuar o atendimento de um cliente. O valor médio da taxa de atendimento é μ (mu). É importante ressaltar que o valor desta taxa é considerado como se o servidor estivesse ocupado 100% do seu tempo. Como há tempo ocioso, a distribuição de frequência (normal, Poisson, exponencial etc.) deste valor é igualmente importante na determinação do grau de complexidade matemática. O pressuposto mais comum é a distribuição de Poisson, porém exige que os eventos de chegada e atendimento sejam completamente independentes. Em todos os casos, os resultados são valores médios ou esperados e supõe-se que as taxas se mantêm constantes ao longo do tempo. De fato, isto pode não ser verdade, uma vez que podem ocorrer alterações no processo tão logo a fila assuma um valor muito alto;
- Disciplina da Fila - Método de decidir qual o próximo cliente a ser atendido. (exemplo: FIFO-primeiro a chegar/ primeiro a ser atendido).
- Número Médio de Clientes na Fila não Vazia (NF) - Número médio de clientes que aguardam o atendimento, ou seja, é o que determina o tamanho da fila. É a característica mais relevante ao se defrontar com a opção de escolher uma fila. A meta é não ter fila, chegar e ser atendido. Supondo que os ritmos médios de chegada e atendimento sejam constantes, o tamanho da fila irá oscilar em torno de um valor médio.
- Número Médio de Clientes no Sistema (NS) - Número de clientes aguardando na fila mais os que estão sendo atendidos. Pode ser entendido também como sendo o tamanho *médio* na fila mais o número *médio* de clientes no atendimento.
- Tempo Médio que o Cliente Fica na Fila (TF) - Tempo médio de espera pelo cliente na fila esperando para ser atendido.
- Tempo Médio que o Cliente Fica no Sistema (TS) - Tempo médio de espera pelo cliente na fila esperando para ser atendido mais o tempo de atendimento. A partir do número médio de clientes no sistema ou na fila, é possível calcular o tempo médio de permanência do cliente no sistema (TS) e na fila (TF).
- A razão ρ (rho) é chamada de “Fator de Utilização do Servidor”, o qual representa a fração média do tempo em que o servidor está ocupado. Este fator é a base de cálculo da probabilidade de haver um número *K* de clientes no sistema, o qual definirá o tamanho da fila e o tempo médio que os clientes permanecem nela e no sistema ($\rho = \lambda / \mu$.).

Observação Importante

Considerando-se que um observador esteja analisando um sistema de atendimento e conclua que $\mu > \lambda$, provavelmente o mesmo concluirá que não deveria haver fila naquele sistema, pois a taxa média de atendimentos do sistema (μ) é maior que a taxa média de chegadas (λ) nele. Vale lembrar que este tipo de análise seria correta se os processos de chegada e de atendimento fossem regulares. Mas, sabendo-se que esses processos são raros na vida real, chega-se a conclusão que existe um fator de aleatoriedade no sistema.
A abordagem matemática da teoria das filas exige que exista estabilidade no sistema (chegada e atendimento), ou seja $\mu > \lambda$, considerando-se com isso que μ e λ devem se manter constantes em relação ao tempo. Do contrário, devem-se utilizar modelos de simulação por computador para efetuar tais análises do sistema.

3. CONDIÇÕES DE OPERAÇÃO DO SISTEMA

Quando essas filas ultrapassam o valor estimado ou normal, pode-se concluir que o sistema está na fase de congestionamento. Nesta fase a qualidade e a produtividade do sistema decresce e o custo operacional tende a subir.

Existem diversos fatores que podem interferir no desempenho de um sistema, tais como:

- ✓ a forma de atendimento aos clientes;
- ✓ a forma da chegada dos clientes;
- ✓ a disciplina da fila e
- ✓ a estrutura do sistema.

3.1. Forma do Atendimento aos Clientes

O primeiro passo para a análise de um sistema de filas é o levantamento estatístico do número de clientes atendidos por unidade de tempo, ou do tempo gasto em cada atendimento. Este procedimento viabiliza a determinação da distribuição de probabilidade do número de atendimentos ou a duração de cada atendimento.

Por exemplo, observando-se a tabela a seguir onde está expresso o tempo de atendimento a 100 clientes, em segundos, de um certo atendente, pode-se chegar ao valor de μ.

20	22	23	18	17	15	21	20	25	26
19	20	18	17	23	22	21	21	22	23
20	23	25	17	14	15	22	20	23	21
25	18	18	17	17	25	26	23	25	24
21	15	17	18	19	22	15	14	15	17
18	20	19	18	20	22	23	24	25	22
22	21	23	20	21	20	23	22	21	20
24	24	25	21	23	20	19	18	17	17
18	15	14	17	13	18	19	18	19	20
18	20	18	22	24	14	24	24	23	25

A tabela a seguir expõe alguns dados importantes para a elaboração da distribuição de frequência.

Dados Importantes		Fórmula no Excel
Menor Valor (segundos)	13	=MÍNIMO(Dados)
Maior Valor (segundos)	26	=MÁXIMO(Dados)
Quant. de Atendimentos	100	=CONT.VALORES(Dados)
Média (segundos / cliente)	20,19	=MÉDIA(Dados)

A média aritmética resulta no tempo médio de atendimento por cliente, ou seja, 20,19 segundos para cada cliente.

Convertendo para minutos 20,19 seg. ÷ 60 min. = 0,3365 min./cliente

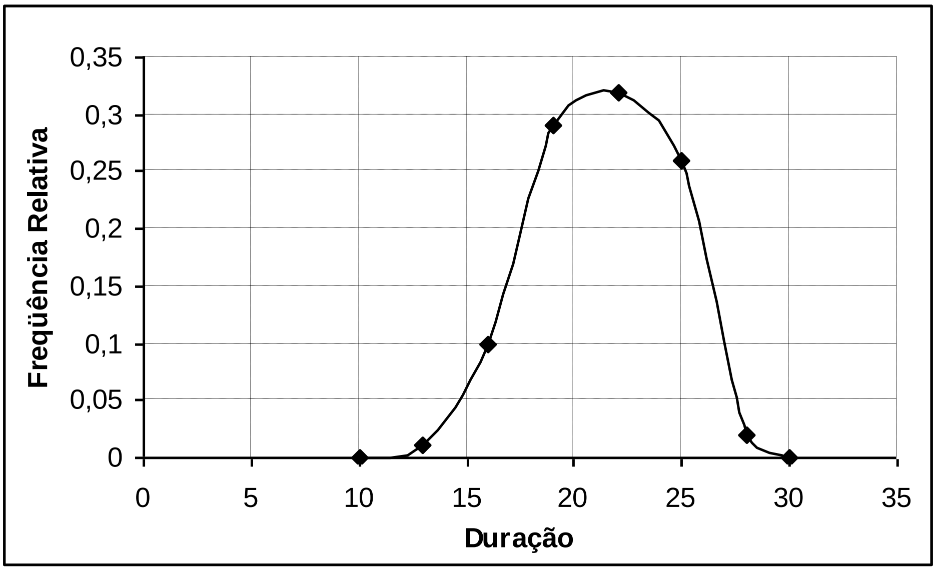
A taxa média de atendimentos pode ser então calculada: μ = 1 cliente / 0,3365 min. = 2,97 clientes/minutos.

Os dados a seguir estão agrupados de forma que se possa avaliá-los em relação a sua distribuição em relação à média. As faixas são determinadas pela Regra de Sturges.

Faixas	Frequência Absoluta	Fórmula no Excel	Frequência Relativa (Probabilidade)
>10	0	={FREQUÊNCIA(Dados;Limite Máx.Tempo)}	0
11-13	1	“	0,01
14-16	10	“	0,10
17-19	29	“	0,29
20-22	32	“	0,32
23-25	26	“	0,26
26-28	2	“	0,02
29>	0	“	0
	$\sum$ = 100		1

Observação: Para determinar a frequência absoluta, localiza-se a célula à direita do primeiro valor (>10) para utilização do assistente de fórmulas. Marcar todas as células, incluindo-se a primeira, incluindo somente o limite máximo (10,13,16,19,22,25,28,29), selecionando-se até a última referência (29>). Pressionar “F2” para editar a fórmula e "CTRL+SHIFT+ENTER" para formação da fórmula em matriz. Pressionar “ENTER” para finalizar a edição.

Para saber-se como os valores se distribuem em torno da média necessita-se plotar os dados em um gráfico.



3.2 Forma da Chegada dos Clientes

Geralmente a chegada dos clientes a um sistema ocorre de forma aleatória. Sendo assim, necessita-se realizar um levantamento estatístico para caracterizar se o processo de chegadas pode ser representado por uma distribuição de probabilidades.

Para efetuar-se este levantamento necessita-se identificar se o processo de chegadas está no estado estacionário, sinalizando que o processo poderá ser sempre representado por este levantamento. Se o levantamento for efetuado no estado não-estacionário, ele não servirá como representante de uma situação normal.

Por exemplo, os usuários de uma agência bancária utilizam-na em um processo estacionário, mas quando da existência de uma greve bancária prolongada, o sistema poderia ser classificado como não-estacionário, pois haveria uma corrida ao banco. Essas situações seriam diferentes e influenciariam nas características da fila, o que poderia implicar em distribuições de probabilidades diferentes.

Por exemplo, observando-se a tabela a seguir onde está expresso a quantidade de veículos que chegaram a um posto de pedágio, em períodos de 1 minuto, em uma hora, pode-se chegar ao valor de λ.

2	4	5	3	3	2	1	4	4	5
2	2	1	3	4	3	4	2	3	4
1	2	4	4	3	2	2	1	1	2
3	2	5	6	6	6	3	3	5	5
5	4	5	5	2	1	1	1	2	1
1	2	2	1	3	3	2	1	3	1

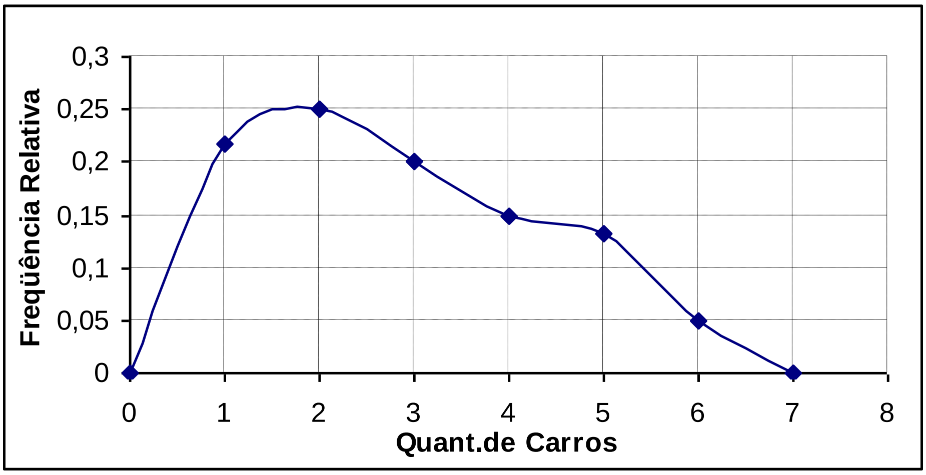
A tabela a seguir expõe alguns dados importantes para a elaboração da distribuição de frequência.

Dados Importantes		Fórmula no Excel
Menor Valor (quant.carros)	1	=MÍNIMO(Dados)
Maior Valor (quant.carros)	6	=MÁXIMO(Dados)
Quant. Total de Veículos	173	=CONT.VALORES(Dados)
Período Total de Análise (minutos)	60	
Quant. de Carros / minuto (λ)	2,88	Total Veíc./ Total Tempo

Os dados a seguir estão agrupados de forma que se possa avaliá-los em relação a sua distribuição em relação a média.

Quant. de Carros	Frequência Absoluta	Fórmula no Excel	Frequência Relativa (Probabilidade)
0	0	={FREQUÊNCIA(Dados;Quant.Carros)}	0
1	13	"	0,22
2	15	"	0,25
3	12	"	0,20
4	9	"	0,15
5	8	"	0,13
6	3	"	0,05
7	0	"	0
	$\sum = 60$		1

Para saber-se como os valores se distribuem em torno da média necessita-se plotar os dados em um gráfico.



Observa-se que esta distribuição se assemelha com a de Poisson.

3.3. Disciplina da Fila

É um conjunto de regras que impõem a ordem em que os clientes serão atendidos. O atendimento pode ser pela ordem de chegada, ou seja, o primeiro a chegar é o primeiro a ser atendido, pela ordem inversa de chegada, ou seja, o último a chegar é o primeiro a ser atendido, por prioridade para certas características etc..

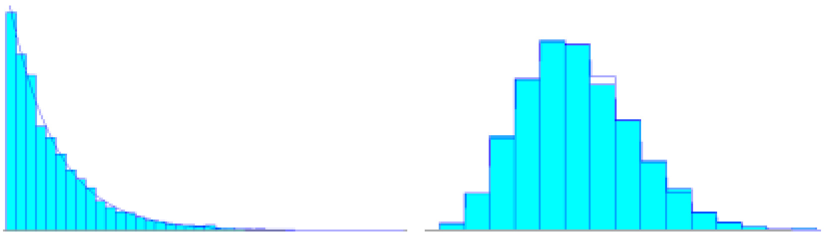
Exemplos:

- ✓ FIFO (*First-In-First-Out*) ou FCFS (*first come, first served*): primeiro cliente a chegar à fila será o primeiro a ser atendido.
- ✓ LIFO (*Last-In-First-Out*) ou LCFS (*last come, first served*): o último cliente a chegar à fila é o primeiro a ser atendido.
- ✓ SIRO (*Service-In-Random-Order*): o atendimento dos clientes faz-se por ordem aleatória.
- ✓ SPT (*Shortest-Processing-Time first*): o cliente a ser atendido em primeiro lugar será aquele cujo tempo de atendimento é menor.
- ✓ PR (*Priority Rules*): o atendimento faz-se de acordo com as regras de prioridades pré-estabelecidas.

Cada Sistema de Filas pode ser descrito segundo a notação de Kendall[1] por seis características (A / B / c / K / m / Z).

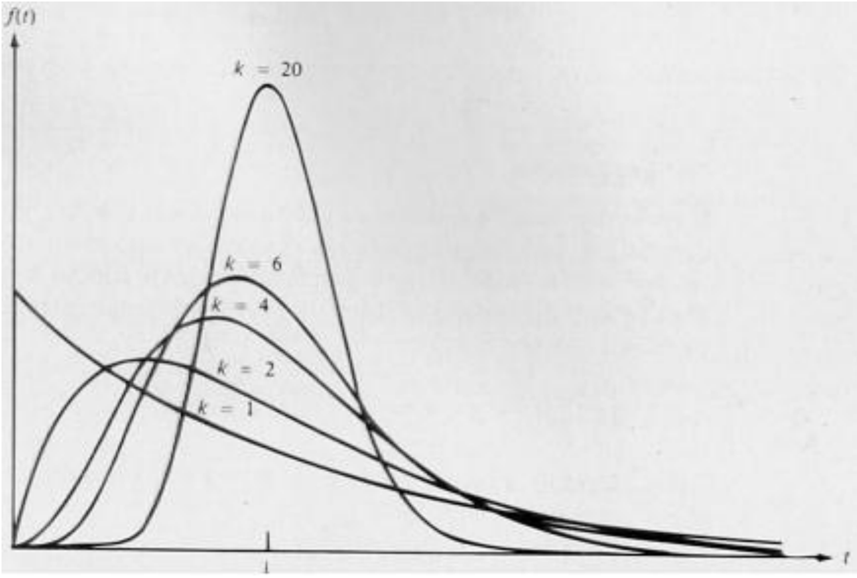
A primeira característica **(A)** especifica a <u>distribuição dos intervalos entre chegadas</u> e a segunda característica **(B)** especifica a <u>distribuição do tempo de serviço.</u> Podem-se utilizar as seguintes abreviações padrões:

- M - intervalos de tempo entre chegadas são independentes, identicamente distribuídos e variáveis aleatórias, seguindo o modelo Marcoviano (distribuição exponencial negativa – distribuição contínua - ou distribuição de Poisson – distribuição discreta);



- D - intervalos de tempo entre chegadas são independentes, identicamente distribuídos e determinístico (distribuição determinística) – os tempos são constantes;
- E_k - intervalos de tempo entre chegadas são independentes, identicamente distribuídos e variáveis aleatórias tendo distribuição de Erlang de ordem "k";

[1] KENDALL, D.G., **Stochastic processes occurring in the Theory of Queues and their analysis by the method of imbedded Markov chains**, p. 338-354 Ann. Math. Statist. 24, 1953.



Observações:
* Para k = 1 a distribuição de Erlang = distribuição exponencial;
* Para k muito grande se aproxima da distribuição normal.

- Hm - Hiper-exponencial de estágio "m" – utiliza-se quando o tempo de serviço apresentar um desvio grande em relação à média;
- G - intervalos de tempo entre chegadas são independentes, identicamente distribuídos e tendo distribuição genérica. Neste caso não é especificada uma distribuição de probabilidade para os tempos de chegada e de atendimento. Os resultados são válidos para todas as distribuições.

A terceira característica **(c)** é a quantidade de servidores em paralelo.

A quarta característica **(K)** especifica o número máximo (capacidade máxima) de usuários no sistema. Se esta capacidade for finita, quando esta for atingida, os usuários que chegam até o instante da próxima liberação são rejeitados.

A quinta característica **(m)** dá o tamanho da população que usa o sistema.

A sexta característica **(Z)** descreve a disciplina da fila.

Observações:

1. Existe uma notação condensada, A/B/c, onde se supõe que não há limite para o tamanho da fila, a população é infinita e a disciplina da fila é FIFO. Para o caso de capacidade limitada, a notação utilizada é A/B/c/K.

2. Os modelos Marcovianos ou de distribuição de Poisson possuem uma grande aplicação teórica uma vez que permitem desenvolver uma teoria sobre filas. Através dele, é possível calcular todas as principais características da fila, sem necessitar efetuar dimensionamentos e estudos financeiros com base em análises mais demoradas com simulação ou uma abordagem matemática complexa. Modelos de filas com distribuições exponenciais levam a dimensionamento de sistemas com mais segurança.

Exemplos:

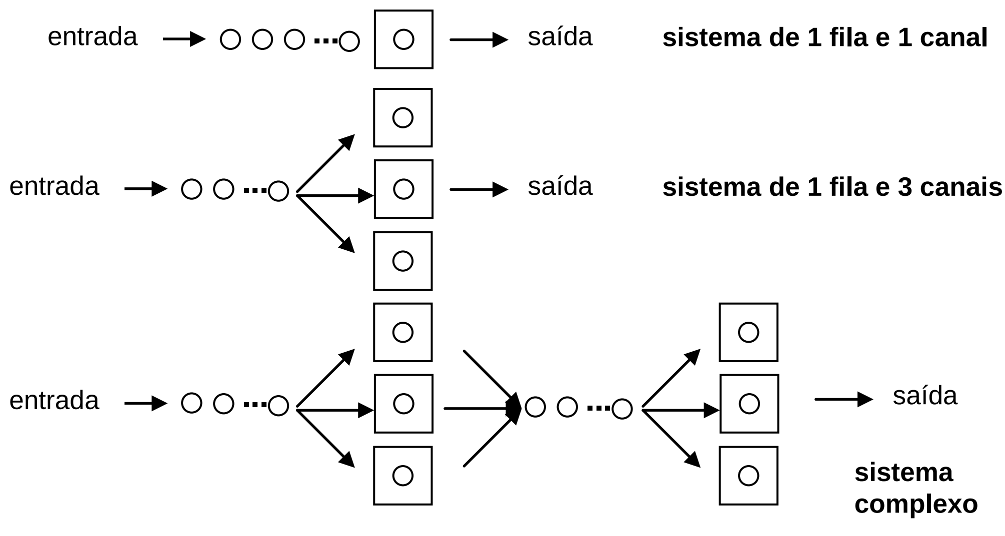
a) M / E2 / 8 / 10 / ∞ / FCFS (*first come, first served*) - pode ser uma clínica com 8 médicos, intervalo entre chegada de clientes representado por distribuição exponencial, tempo de atendimento representado por distribuição de Erlang de ordem 2, disciplina da fila de atendimento por ordem de chegada, com capacidade total do sistema para 10 clientes e população infinita.

b) M / M / 3 / 20 / 1500 / FCFS (*first come, first served*) - O intervalo entre chegadas sucessivas é distribuído exponencialmente, os tempos de serviço são exponencialmente distribuídos, há três servidores, a fila possui *buffers* para 20 usuários, isto é, 3 usuários em atendimento e 17 esperando por serviço, enquanto a quantidade de usuários estiver em seu valor máximo (20) todos os usuários que chegarem serão perdidos até que o comprimento da fila diminua, há um total de 1500 usuários que podem ser atendidos, a disciplina de atendimento é primeiro cliente a chegar à fila será o primeiro a ser atendido.

c) M/M/1 é conhecido como modelo de Poisson. Ele é mais utilizado em estudos teóricos, pois permite, facilmente, calcular todos os atributos de uma fila, facilitando, inclusive, a sua análise financeira. Considera-se que não há limite para o tamanho da fila, a população é infinita e a disciplina da fila é FIFO

3.4. Estrutura do Sistema

Existem vários tipos de estruturas do sistema, e por isso, necessitam-se ser estudados caso a caso. A seguir estão relacionados três exemplos de configurações.



4. MODELO D / D / 1 / k / FIFO

Considerações do Sistema

Neste modelo, os tempos entre chegadas sucessivas são iguais a $1/\lambda$, os usuários são atendidos individualmente e na ordem de chegada por um único servidor em tempos iguais a $1/\mu$. O sistema está limitado quanto à sua capacidade a k usuários.

Neste modelo, se $\lambda \leq \mu$ não haverá formação de fila; para $\lambda > \mu$, a limitação de capacidade do sistema deverá ser imposta, caso contrário, a fila crescerá sem limite. Considerar ainda o seguinte:

- t é o instante de tempo em que se avalia a quantidade de usuários. Entre t = 0 e t = 1/λ o sistema está vazio;
- t_0^* é o instante de tempo em que acontece a primeira rejeição devido à capacidade *k* do sistema. Entre 1/λ e t_0^* a quantidade média de clientes no sistema é avaliada por:
  NS(t) = [qtd de chegadas entre 0 e *t*] – [qtd de serviços completados até *t*]
  Onde [x], na expressão anterior, representa a parte inteira do conteúdo de *x*.
  NS (t_0^*) = *k* + 1
- Entre t_0^* e ∞ há chegadas de clientes no sistema com efetivo ingresso e outras que são rejeitadas devido à capacidade *k*.

Por isso, as medidas de desempenho do sistema são:

✓ Quantidade Média de Clientes no Sistema:

➢ Quando existir *m* inteiro positivo tal que 1/μ = m 1/λ, tem-se:

$$NS\ (t) = \begin{cases} 0, & 0 \leq t < \frac{1}{\lambda}\ (sistema\ vazio) \\ [\lambda t] - \left[\left(t - \frac{1}{\lambda}\right)\mu\right], & \frac{1}{\lambda} \leq t < t_0^*\ (fase\ antes\ da\ 1^{\text{a}}\ rejeição) \\ K, & t \geq t_0^*\ (fase\ após\ a\ 1^{\text{a}}\ rejeição) \end{cases}$$

Obs.: K define a capacidade do sistema.

➢ Quando <u>não</u> existir *m* inteiro positivo tal que 1/μ ≠ m 1/λ, então:

$$NS\ (t) = \begin{cases} 0, & 0 \leq t < \frac{1}{\lambda} \\ [\lambda t] - \left[\left(t - \frac{1}{\lambda}\right)\mu\right], & \frac{1}{\lambda} \leq t < t_0^* \\ K - 1\ ou\ K, & t \geq t_0^* \end{cases}$$

✓ Tempo de espera do *n-ésimo* cliente na fila :

➢ Quando existir *m* inteiro positivo tal que 1/μ = m 1/λ, tem-se:

$$TF = \begin{cases} \left(\frac{1}{\mu} - \frac{1}{\lambda}\right)(n-1), & n < \lambda t_0^* \\ (k-1)\frac{1}{\mu}, & n \geq \lambda t_0^* \end{cases}$$

➢ Quando <u>não</u> existir *m* inteiro positivo tal que 1/μ ≠ m 1/λ, então:

$$\text{TF} = \begin{cases} \left(\frac{1}{\mu} - \frac{1}{\lambda}\right)(n-1), & n < \lambda t_0^* \\ \textit{não existe expressão geral}, & n \geq \lambda t_0^* \end{cases}$$

Sendo λt_0^* a ordem do primeiro usuário rejeitado, devido à capacidade *k* do sistema.

Exemplo 1

Analisar o sistema representado por D/D/1/6/FIFO com $1/\lambda$=3 segundos e $1/\mu$ = 6 segundos.

m = $1/\mu \div 1/\lambda$ = 6/3 = 2

$$\text{NS} = \begin{cases} 0, & 0 \leq t < 3 \\ [t/3] - \left[(t-3)\frac{1}{6}\right], & 3 \leq t < t_0^* \\ 6, & t \geq t_0^* \end{cases}$$

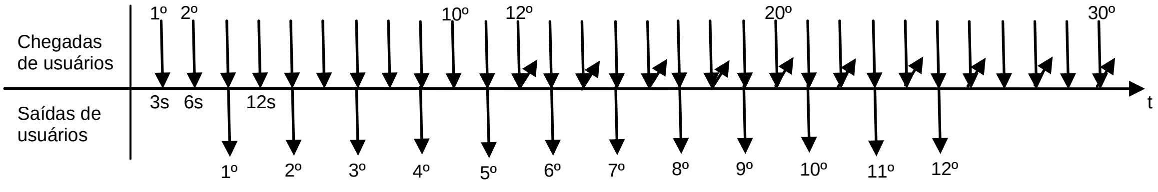
Como NS (t_0^*) = *k* + 1 = 6 + 1 = 7 clientes então: NS (t_0^*) = $[t_0^*/3] - \left[(t_0^* - 3)\frac{1}{6}\right]$ = 7. Para se calcular t_0^*, instante de tempo em que ocorre a primeira rejeição, deve-se proceder de forma empírica como se segue:

t_0^*	60	52	36 (6,5s)	33
NS (t_0^*)	11	9	7	6

Obs.: a coluna em negrito representa o 1º t_0^*, pois para t_0^* = 35s, NS (t_0^*) = 6 clientes (6,33s).
Sendo assim, λt_0^* (cheg./seg. x seg. = chegadas) que é a ordem do primeiro usuário rejeitado, é: 36 x 0,333 = 12º. Por isso, o tempo de espera na fila é dado por:

$$\text{TF} = \begin{cases} (6-3)(n-1), & n < 12 \\ (6-1)6, & n \geq 12 \end{cases} \quad \text{TF} = \begin{cases} 3(n-1), & n < 12 \\ 30, & n \geq 12 \end{cases}$$

Conforme apresenta o gráfico a seguir, verifica-se que o TF quando $n < 12$, considerando-se, por exemplo, o 10º usuário, tem-se que TF = $3(10-1) = 27$ segundos, ou seja, ele aguarda até a saída do 9º usuário para que ele ingresse no sistema.



Exemplo 2

Analisar o sistema representado por D/D/1/6/FIFO com $1/\lambda$=3 segundos e $1/\mu$ = 5 segundos.

m = 1/μ ÷ 1/λ = 5/3 = 1,67

$$NS = \begin{cases} 0, & 0 \le t < 3 \\ [t/3] - \left[(t-3)\frac{1}{5}\right], & 3 \le t < t_0^* \\ 5\ ou\ 6, & t \ge t_0^* \end{cases}$$

$$TF = \begin{cases} (5-3)(n-1) & n < 3 \\ \text{não existe expressão geral}, n \ge 14 \end{cases}$$

Como NS (t_0^*) = *k* + 1 = 6 + 1 = 7, então: NS (t_0^*) = $[t_0^*/3] - \left[(t_0^* - 3)\frac{1}{5}\right]$ = 7. Para se calcular t_0^*, instante de tempo em que ocorre a primeira rejeição, deve-se proceder de forma empírica como se segue:

t_0^*	54	48	45	42	39	30
NS (t_0^*)	8	7	7	7	6	5

Sendo assim, λt_0^* que é a ordem do primeiro usuário rejeitado, é: 42 x 0,333 = 14º.

Para se verificar a quantidade de usuários no sistema (NS) para $t \ge t_0^*$, determinado como 5 ou 6, de forma mais facilitada, faz-se necessária a elaboração de um gráfico similar ao do exemplo 1.

5. MODELO M / M / 1

Considerações do Sistema

As equações para este modelo baseiam-se nas seguintes características:

- Formas da chegada à fila e de atendimento seguem o modelo Marcoviano (distribuição de Poisson ou a distribuição exponencial negativa) e;
- Quantidade de canais de atendimento igual a 1.

Expressões

- ✓ Quantidade Média de Clientes no Sistema: NS = λ / (μ - λ)
- ✓ Quantidade Média de Clientes na Fila: NF = λ^2 / [μ (μ - λ)]
- ✓ Fator de Utilização do Servidor – Fração Média de Tempo que o Servidor está Ocupado: ρ = λ / μ
- ✓ Probabilidade de Existirem *n* Clientes no Sistema: P(n) = (1 - λ / μ) $(\lambda / \mu)^n$

Teorema de Little: Para qualquer sistema de filas, no qual exista uma distribuição em regime constante, são válidas as seguintes relações:

NS = λ TS e NF = λ TF

Vale relembrar que sistemas estáveis são caracterizados por $\lambda < \mu$, ou seja, $\rho < 1$. Quanto mais o valor de "ρ" se aproxima de 1 a fila tende a aumentar infinitamente. Observa-se pela expressão de NF que se $\lambda = \mu$, ou seja, $\rho = 1$, o tamanho da fila é infinito.

Exemplo 1:

O número médio de carros que chegam a um posto de informações é igual a 10 carros/hora. Assumir que o tempo médio de atendimento por carro seja de 4 minutos, e ambas as distribuições de intervalos entre chegadas e tempo de serviço sejam exponenciais.

Responder as seguintes questões:
a - Qual a probabilidade do posto de informações estar livre?
b - Qual a quantidade média de carros esperando na fila?
c - Qual o tempo médio que um carro gasta no sistema (tempo na fila mais o tempo de atendimento) ?
d - Quantos carros serão atendidos em média por hora?

Dados do Problema:

Chegada: $\lambda = 10$ carros/hora.
Atendimento: em média, 1 carro a cada 4 minutos, ou seja 15 carros/hora (60/4). Sendo assim, $\mu = 15$ carros/hora.

Solução:

a - $P(0) = (1 - \lambda / \mu)(\lambda / \mu)^0 = (1 - 10 / 15) \times 1 = 1 / 3 = 33{,}33\%$
b - $NF = \lambda^2 / [\mu (\mu - \lambda)] = 10^2 / 15 (15 - 10) = 1{,}33$ carros
c - Dado que $NS = \lambda\, TS$, então:
$TS = NS / \lambda = 1 / (\mu - \lambda)$ $= 1 / 5 = 0{,}2$ horas ou 12 minutos **(Atenção)**
d - Se a ocupação média do posto fosse de 100%, então, o número médio de carros atendidos por hora seria de 15 carros. Sendo a ocupação média, a 100%, igual a 1 - P(0), ou seja, igual a 2/3, então o número de carros atendidos por hora seria de:

$$15 * 2 /3 = 10 \text{ carros por hora.}$$

Exemplo 2:

Supondo-se que a chegada de um navio ao berço portuário siga a distribuição de Poisson, com uma taxa de 6 navios por dia. A duração média de atendimento dos navios é de 3 horas, seguindo-se a distribuição exponencial. Calcule os seguintes valores:

a – Qual a probabilidade de um navio chegar ao porto e não esperar para atracar?
b – Qual é a quantidade média de navios na fila do porto?
c – Qual é a quantidade média de navios no sistema portuário?

d – Qual é a quantidade média de navios utilizando o porto?
e - Qual é o tempo médio de um navio na fila?
f – Qual deve ser a taxa de chegada de um navio para que o tempo médio na fila seja de 3 horas?
g – Qual é a probabilidade do berço portuário estar em uso?

Dados do Problema:

Chegada: $\lambda = 6$ navios/dia.
Atendimento: em média, 1 navio a cada 3 horas, ou seja 8 navios/dia (24/3). Sendo assim, $\mu = 8$ navios/dia.

Solução:

a - $P(0) = (1 - \lambda / \mu)(\lambda / \mu)^0 = (1 - 6 / 8) \times 1 = 1 / 4 = 25\%$
b - $NF = \lambda^2 / [\mu (\mu - \lambda)] = 6^2 / [8 (8 – 6)] = 2,25$ navios
c – $NS = \lambda / (\mu - \lambda) = 6 / (8 – 6) = 3$ navios
d – Navios no Porto = NS – NF = 3 – 2,25 = 0,75 navio
e – TF = ?, como $NF = \lambda$ TF então $TF = NF/ \lambda$, ou seja TF = 2,25 / 6 = 0,375 dia = 9 horas
f – Se TF = 3 horas = 0,125 dia (3/24), mantendo-se a mesma taxa de atendimento (μ), deve-se calcular a nova taxa de chegada (λ). Sendo assim:

Sendo $NF = \lambda^2 / [\mu (\mu - \lambda)]$ e $NF = \lambda\, TF$, então $\lambda^2 / [\mu (\mu - \lambda)] = \lambda\, TF \therefore$
$TF = \{\lambda^2 / [\mu (\mu - \lambda)] \} \times 1 / \lambda = \lambda / [\mu (\mu - \lambda)] = \lambda / (\mu^2 - \mu\lambda) \therefore$
$TF (\mu^2 - \mu\lambda) = \lambda \therefore TF\, \mu^2 – TF\, \mu\, \lambda = \lambda \therefore TF\, \mu^2 = TF\, \mu\, \lambda + \lambda \therefore$
$\lambda (TF\, \mu + 1) = TF\, \mu^2 \therefore \boxed{\lambda = TF\, \mu^2 / (TF\, \mu + 1)}$

$\lambda = 0,125 \times 8^2 / (0,125 \times 8 + 1) = 4$ navios / dia
g – Se a probabilidade de não ter nenhum navio no berço portuário é de 25%, então a probabilidade de ter-se um navio atracado é de 1 – P(0) = 1 – 1 / 4 = 3 / 4 = 0,75 = 75%

Exemplo 3:

Uma distribuidora de combustíveis utiliza caminhões para transportar o seu produto. Sabendo-se que esta empresa só tem um ponto de abastecimento dos caminhões, que os ritmos de chegada e de atendimento seguem as distribuições do modelo Marcoviano, que a taxa de chegada dos caminhões é de 4 unidades por hora, que a taxa de atendimento é de 5 unidades por hora, que os custos horários do funcionário que abastece o veículo é de 5,00 unidades monetárias e do motorista é de 12,00 unidades monetárias, calcule o custo horário do sistema e a probabilidade do funcionário que abastece ficar sem nenhum caminhão para abastecer.

Dados do Problema:

Chegada: $\lambda = 4$ caminhões/hora.
Atendimento: $\mu = 5$ caminhões/hora.
Custo do funcionário que abastece o caminhão: 5,00 unidades monetárias.

Custo do motorista: 12,00 unidades monetárias.
Solução:

Quantidade de Caminhões no Sistema:

NS = λ / (μ - λ) = 4 / (5 – 4) = 4 caminhões

Por hora, 1 funcionário abastece 4 caminhões, então calcula-se o custo do sistema por:

Custo Horário do Sistema = 5,00 + (12,00 x 4) = 53,00 unidades monetárias.

Probabilidade de não ter nenhum caminhão para ser abastecido:

$P(0) = (1 - \lambda / \mu)(\lambda / \mu)^0 = (1 - 4 / 5) \times 1 = 20\%$.

Exemplo 4:

Um técnico de laboratório gasta 30 minutos em média para reparar relés. Considerar que o tempo distribui-se conforme uma distribuição exponencial negativa. Os relés chegam à recepção do laboratório segundo a distribuição de Poisson a uma taxa média de 10 relés por dia. Considerar um turno de 8 horas de trabalho. Eles são reparados de acordo com a ordem de chegada.
a) Qual a folga média do técnico por dia de trabalho?
b) Em média, quantos relés se encontram na oficina aguardando reparação?

Solução:

Tempo para reparação: 30 minutos, ou seja, $\mu = 2$ relés/h
Taxa de chegada: 10 relés/8 horas de trabalho, ou seja, $\lambda = 1{,}25$ relés/h

a) $P(0) = (1 - \lambda / \mu)(\lambda / \mu)^0 = (1 – 1{,}25/2)(1{,}25/2)^0 = 0{,}375 = 37{,}5\%$
   Como são 8 horas diárias de trabalho: 8 x 0,375 = 3 h
b) $NF = \lambda^2 / [\mu (\mu - \lambda)] = 1{,}25^2 / [2 (2 – 1{,}25)] = 1{,}04$ aparelhos.

Exercício 1:

Uma loja de departamentos 24h tem um técnico para fazer manutenção nas máquinas registradoras. Sabe-se que na loja de Recife, quatro máquinas falham, em média, por dia, seguindo distribuição de Poisson. O técnico repara, em média, seis máquinas por dia, seguindo-se uma distribuição exponencial. Como a loja depende da média dos registros de venda, pela falta de máquina, ela deixa de registrar R$ 3.500,00 por hora. O técnico usa materiais para manutenção que custam, em média por máquina, R$ 310,00. Por isso, determine:
a) Qual é o custo total de máquina parada? R$ 168.620,00
b) Qual a probabilidade do técnico não estar executando a manutenção de máquinas? 33%
c) Qual é a quantidade média de máquinas na fila? 1,33 máquina
d) Qual é a quantidade média de máquinas no sistema? 2 máquinas
e) Qual é o tempo médio das máquinas na fila? 7,92h
f) Qual é o tempo médio das máquinas no sistema? 12h

Exercício 2:

Considere um sistema de fila tipo M/M/1 em que navios chegam a um porto para carregar algum produto. A seguir estão anotados os valores, para 20 navios, dos intervalos entre chegadas (em horas) e da duração de carga (em horas) de cada navio. Baseando-se neles, determine:

Abertura do Porto (t=0)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
A	10	2	13	7	2	8	8	8	10	9	1	14	14	1	10	9	9	9	8	14
B	5	5	3	3	6	7	6	8	2	5	8	8	8	3	4	3	3	4	5	5

A – Intervalo entre chegadas
B – Duração dos atendimentos

a) A taxa de chegada; 2,86 navios/dia
b) A taxa de atendimento; 4,76 navios/dia
c) O tamanho médio da fila; 0.9 navio
d) A quantidade de navios no sistema; 1,51 navio
e) O tempo médio dos navios na fila; 0,31 dia = 7,44h
f) A probabilidade de existirem três navios no sistema. 9%

Exercício 3:

Em uma mineradora verificou-se que o tempo médio dos caminhões junto a um sistema de carregadeiras tipo M/M/1 é de 3 minutos e que, em média, existem 6 caminhões neste sistema. Qual é a taxa de chegada dos caminhões? Resp. 2 caminhões/minuto

Exercício 4:

Um operador logístico recebe, em média, 4.000 itens de produtos em uma hora. O ponto de entrada desses itens é único para recebimento e avaliação crítica. Existe um único atendente para este ponto de entrada. Ele tem capacidade para atender 4.200 itens por hora, em média. Sabendo-se que a chegada e o atendimento podem ser representados por distribuição exponencial, determine:

a) Qual a fração média do tempo que o atendente está ocupado? 95%
b) Qual a quantidade média dos itens na fila? NF=19 itens
c) Qual a quantidade média dos itens no sistema? NS=20 itens
d) Qual o tempo médio dos itens na fila? TF=17,14s
e) Qual o tempo médio dos itens no sistema? TS=18s
f) Qual a probabilidade de não existirem itens no sistema? 5%
g) Qual deverá ser a taxa de chegada dos itens para uma redução do tempo no sistema de 30%? 3914,3 itens/h
h) Quantos itens serão efetivamente atendidos? 3990 itens/h

Exercício 5:

Numa sala de espera, com 1 médico atendendo, há 15 clientes em média. A taxa de chegada é de 1 cliente a cada 30 segundos. Qual é o tempo médio de espera dos clientes na sala? Os clientes são atendidos por ordem de chegada (FIFO).
Resp.: 7,5 minutos

Exercício 6:

Um sistema para atendimentos está associado a 100 computadores. O tempo médio para resposta à requisição do computador ao sistema de atendimento é de 0,6 segundos. No horário de pico são efetuadas 20 consultas/minuto. Qual é a probabilidade do sistema de atendimento estar livre? Resp. 80%

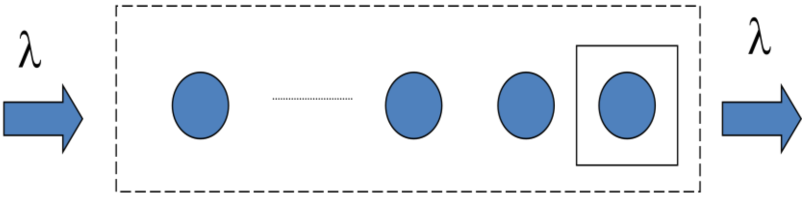
Exercício 7:

Uma indústria deseja contratar um especialista para manutenção de máquinas que apresentam 3 falhas/h. Para isso, a indústria possui 2 opções:
- 1 Especialista lento: ritmo de manutenção de 4 falhas/h a R$ 3,00/hora
- 1 Especialista rápido: ritmo de manutenção de 6 falhas/h a R$ 5,00/hora

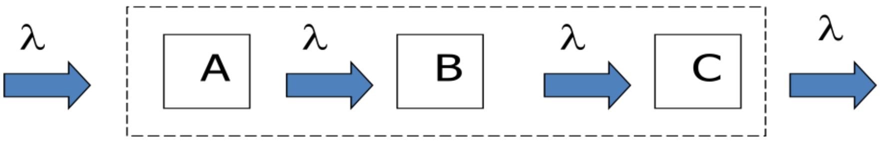
O custo de máquina parada é de R$ 5,00/h. Qual é a melhor contratação que deve ser efetuada pela indústria de forma a minimizar o custo total? Resp. Especialista rápido R$ 10,00/h

<u>5.1. SISTEMA DE FILAS SEQUENCIAIS – M/M/1</u>

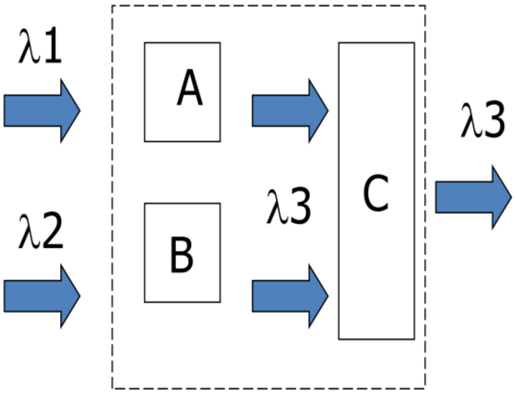
Em qualquer sistema estável, o fluxo que entra é igual ao fluxo que sai.



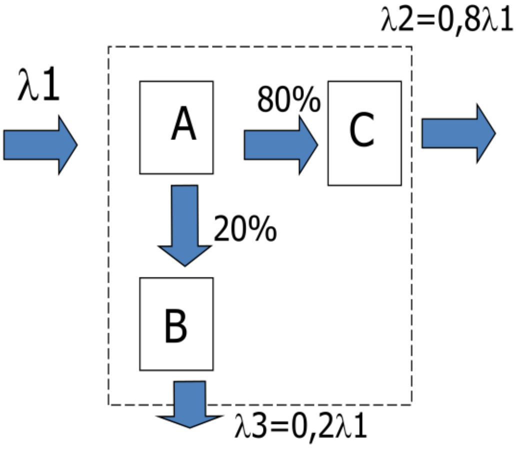
Em um sistema estável, o fluxo de entrada se mantém nas diversas seções do sistema.



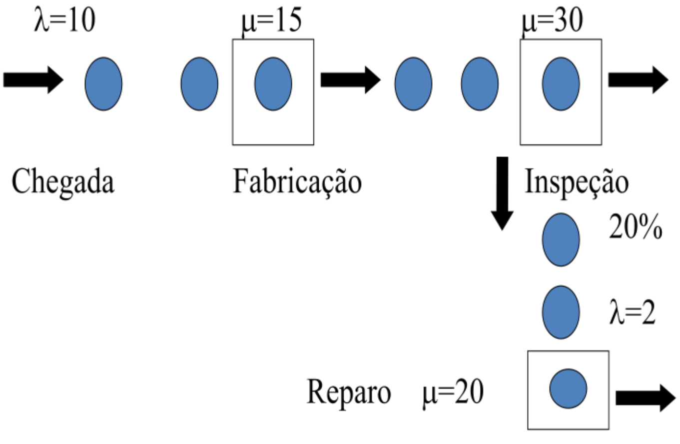
Em um sistema estável, a junção de fluxos equivale às suas somas.



Em um sistema estável, o fluxo se desdobra aritmeticamente.



Exemplo 1: em um sistema de filas sequenciais, conforme figura, calcule o tamanho das filas que se formam em cada servidor.



6. MODELO M / M / 1 / k

k: número máximo (capacidade máxima) de usuários no sistema.

Expressões

- ✓ Probabilidade de Existirem *n* Clientes no Sistema: P(n) = $\rho^n \left[\frac{(1-\rho)}{(1-\rho^{k+1})}\right] para\, n \leq k$
  sendo P(0) = $\frac{(1-\rho)}{(1-\rho^{k+1})}$
- ✓ P(n) = $0\ para\, n > k$
- ✓ Quantidade Média de Clientes no Sistema: NS = $\frac{\rho[1-(k+1)\rho^k+k\rho^{k+1}]}{(1-\rho^{k+1})(1-\rho)}$
- ✓ Quantidade Média de Clientes em Atendimento: NA = 1 - P(0)
- ✓ Quantidade Média de Clientes na Fila: NF = NS – NA
- ✓ Tempo Médio de Clientes no Sistema: TS = $\frac{NS}{\{\lambda[1-P(k)]\}}$
- ✓ Tempo Médio de Clientes na Fila: TF = $\frac{NF}{\{\lambda[1-P(k)]\}}$

Observação: embora a taxa de chegada de clientes ao sistema seja λ, a taxa de clientes que permanecem no sistema é λ [1 − P(k)]

Exemplo 1:

Uma barbearia tem 1 barbeiro e um total de 10 cadeiras. O intervalo de tempo entre chegada de clientes à barbearia é em média de 20 clientes por hora. Aqueles clientes que chegam e encontram a barbearia cheia, não entram. Os barbeiros levam em média 12 minutos para cortar o cabelo de cada cliente. Os tempos gastos nos cortes de cabelo são distribuídos exponencialmente.
1 - Na média, qual é a taxa de clientes que **não** serão **efetivamente** atendidos (ficarão à espera)?
2 - Na média, quanto tempo cada cliente gasta na barbearia?

Solução:

Taxa de ocupação do sistema: ρ = λ / μ = 20 / 5 = 4
1 - Número médio de clientes que entram por hora na barbearia:
a) probabilidade de ter menos de 10 clientes na barbearia:
$1 - P(10) = 1 - 4^{10} [(1 - 4) / (1 - 4^{11})] = 0{,}25$
b) número médio de clientes que **permanecem** na barbearia:
λ [1 - P(10)] = 20 x 0,25 = 5 clientes / hora
Obs.: Número de clientes que **não** serão atendidos (ficarão à espera) é igual a:
20 - 5 = 15 clientes / hora
2 - Tempo gasto por cliente na barbearia:
a) Número médio de clientes na barbearia:
$NS = \{ 4 [1 - 11(4^{10}) + 10(4^{11})] \} / \{ (1- 4^{11})(1 - 4) \} = 9{,}67$ clientes
b) Tempo médio gasto por cliente na barbearia:
TS = 9,67 / (20 x 0,25) = 1, 93 horas

Exercício 1:

Em um Centro de Distribuição (CD) existe a capacidade de receber, no máximo, 40 caminhões. Sabe-se que se pode atender um caminhão de cada vez. As taxas de chegada e de atendimento do CD seguem uma distribuição Markoviana. Dez caminhões são atendidos, em média, por hora e chegam ao CD a cada 10 minutos. Determine:

a) Qual a probabilidade do CD estar em atendimento? 1 – P(0) = 60%
b) Qual é o tempo de espera no CD? TS = 0,25h (15 min.)
c) Qual é a quantidade média de caminhões na fila? NF = 0,89 cam.
d) Qual é a quantidade média de caminhões no sistema? NS = 1,49 cam.
e) Qual é o tempo médio dos caminhões na fila? TF = 0,15h (9 min.)

Exercício 2:

Um pátio ferroviário tem capacidade para receber no máximo 30 vagões de minério, sem influenciar no tráfego. Sabe-se que só existe um virador de vagão para descarregar a carga. As taxas de chegada dos vagões e de atendimento seguem as distribuições de Poisson e exponencial, respectivamente. Um vagão é virado a cada 10 minutos. Chega ao pátio 1 vagão a cada 15 minutos. Determine os seguintes parâmetros do sistema:

a) Qual é a probabilidade de não existir vagão no pátio? P(0) = 34%
b) Qual é o tempo médio gasto, por cada vagão, no pátio? TS = 30 minutos
c) Qual é o tempo médio do vagão na fila? TF = 19,8 minutos
d) Qual é a probabilidade do pátio ter algum trem em atendimento? 1 – P(0) = 0,66

Exercício 3:

Um porto, com um ponto para descarga, recebe navios RO-RO. Nesse ponto pode-se descarregar, em média, 5 navios/dia, acostando, no máximo, 2 navios de cada vez. Quando o porto está ocupado, os navios que chegam são desviados, acarretando R$ 20.000,00 por navio desviado. O navio parado no porto custa R$ 12.000,00/dia e por navio. Sabe-se que a chegada segue uma distribuição de Poisson com taxa de 3 navios/dia. Os tempos de chegada seguem uma distribuição exponencial negativa. Avalie a viabilidade econômica para ampliar o cais para receber mais um navio (3 ao todo), o que levaria o custo diário em R$ 1.000,00, ou manter a configuração atual.
Resp.: Ampliando para três berços, a economia seria de R$ 1.220,00.

<u>7. MODELO M / M / c</u>

c: especifica o número de canais de atendimento ou número de servidores.

Expressões

- ✓ Taxa de Ocupação do Sistema: $\rho = \lambda / c\mu$ onde μ é a taxa por canal de atendimento
- ✓ Probabilidade de Existirem *n* Clientes no Sistema: P(n) = $\frac{(c\rho)^n P(0)}{n!}$ $para\ n = 1, 2, \dots, c$ sendo P(0) = $\frac{1}{\sum_{i=0}^{i=c-1} [\frac{(c\rho)^i}{i!} + \frac{(c\rho)^c}{c!(1-\rho)}]}$

- ✓ P(n ≥ c) = $\frac{(c\rho)^c P(0)}{c!(1-\rho)}$
- ✓ Quantidade Média de Clientes na Fila: NF= $\frac{\mathrm{P(n \geq c)}\rho}{(1-\rho)}$
- ✓ Quantidade Média de Clientes no Sistema: NS = NF + NA = NF + (λ / μ)
- ✓ Tempo Médio de Clientes no Sistema: TS = $\frac{NS}{\lambda}$
- ✓ Tempo Médio de Clientes na Fila: TF = $\frac{NF}{\lambda}$
- ✓ Tempo Médio de Clientes em Atendimento: TA = 1 / μ

Exemplo 1:

Considere um banco com 2 caixas para atendimento de clientes. Uma média de 80 clientes por hora chegam ao banco e esperam em 1 única fila para serem atendidos. O tempo médio de atendimento por cliente é de 1,2 minutos. Assumir que o intervalo de tempo entre chegadas de clientes e o tempo de atendimento são exponenciais.

Determinar:
1 - O número esperado de clientes no banco.
2 - O tempo que cada cliente gasta no banco.

Dados:
λ = 80 clientes/h
μ = 60/1,2 = 50 clientes/h
c = 2

Solução:

Taxa de ocupação do banco: ρ = λ / cμ = 80 / (2 x 50) = 0,80
1 –

a) Probabilidade de se ter mais de 2 clientes no banco:

   P(0) = $\frac{1}{\left[\frac{(2\times0{,}8)^0}{0!}+\frac{(2\times0{,}8)^2}{2!(1-0{,}8)}\right]+\left[\frac{(2\times0{,}8)^1}{1!}+\frac{(2\times0{,}8)^2}{2!(1-0{,}8)}\right]} = 0{,}065$

   P(n ≥ c) = $\frac{(2\times0{,}8)^2\times0{,}065}{2!(1-0{,}8)} = \frac{0{,}1664}{0{,}4} = 0{,}416$

b) Número de clientes esperando na fila - NF= $\frac{\mathrm{P(n \geq c)}\rho}{(1-\rho)}$ = (0,416 x 0,80) / (1 – 0,8) = 1,7 cliente
c) Número de clientes esperando no Sistema: NS = NF + (λ / μ) = 1,7 + (80/50) = 3,3 clientes

2 – TS = NS / λ = 3,3 / 80 = 0,041 h = 2,46 minutos

Exercício 1:

Uma agência de reciclagem de alumínio tem duas máquinas para extrusão e funciona em períodos de 8 h diárias. A carga chega à agência a cada 10 minutos e, descarregada e extrusada a cada 15 minutos. A empresa pretende colocar mais uma extrusadora, mas não sabe se será interessante quanto ao tempo da carga na fila. Verifique isso. Resp. melhor opção C = 3.

8. MODELO M / Er / 1

Segundo Pereira (2009), o modelo de Erlang, com um servidor, as chegadas seguindo a distribuição de Poisson com parâmetro *λ* e os tempos de atendimento seguindo a distribuição de Erlang de ordem *r* com parâmetros *r* e *μ,* aplica-se, por exemplo, no caso em que se tem uma atividade para passar, etapa para etapa, por uma série de *r* fases de produção independentes. Cada etapa tem um tempo com distribuição exponencial com um parâmetro comum μ. A análise do modelo M/E*r*/1 é similar a do modelo M/M/1.

A taxa de ocupação é dada por $\rho = r\lambda / \mu$. Para que o sistema atinja o estado de equilíbrio é necessário e suficiente que $\rho < 1$.

As expressões que fundamentam a resolução do modelo de Erlang estão expostas na tabela a seguir.

Expressões para o Modelo de Filas de Erlang de Ordem r

Probabilidade de ocorrência do estado 0	$P_0 = \left[\sum_{k=1}^{r} \frac{-\mu/\lambda}{z_k^r \prod_{i=1,\ i\neq k}^{r} (1-z_i/z_k)} \frac{z_k}{z_k-1}\right]^{-1}$
Probabilidade de ocorrência do estado n	$P_n = P_0 \sum_{k=1}^{r} \frac{-\mu/\lambda}{z_k^r \prod_{i=1,\ i\neq k}^{r} (1-z_i/z_k)} \left(\frac{1}{z_k}\right)^n$
Número médio de clientes no sistema	$L_s = \rho \frac{\rho+r(2-\rho)}{2r(1-\rho)}$
Comprimento médio da fila	$L_q = \frac{\rho^2(r+1)}{2r(1-\rho)}$
Tempo médio de espera no sistema	$W_s = \frac{\rho+r(2-\rho)}{2\mu(1-\rho)}$
Tempo médio de espera na fila	$W_q = \frac{\rho(r+1)}{2\mu(1-\rho)}$
Taxa de ocupação	$\rho = \frac{\lambda r}{\mu}$

Fonte: Pereira (2009)

Obs.: cuidado com as variáveis que são diferentes das usadas nesta apostila

9. CADEIAS DE MARKOV

9.1. Sinopse de Conceitos Associados à Teoria da Probabilidade

Considerando-se experimentos em que os resultados não sejam previsíveis antecipadamente, tais como lançamento de uma moeda, jogar um dado, vida útil de um equipamento mecânico etc., pode-se considerar como espaço amostral os resultados possíveis destes.

Para os experimentos anteriores, tem-se: uma moeda lançada, Espaço Amostral = {cara, coroa}; para o dado jogado, Espaço Amostral = {1, 2, 3, 4, 5, 6}; para vida útil de um equipamento mecânico, Espaço Amostral = [0, ∞).

Para tais experimentos, tem-se que o subconjunto de cada espaço amostral é denominado evento. Para a moeda, Evento 1 = {cara} e Evento 2 = {coroa}; para o dado, considerar os resultados que são números pares, ou seja, Evento 1 = {2, 4, 6}; para um equipamento que dure ao menos 1 anos, mas não complete o segundo, tem-se Evento 1 = [1,2).

Tomando-se, então, os resultados de um experimento que pode ser listado pelo espaço amostral (S) com os seus eventos (E), observa-se que a probabilidade P de um certo evento - P(E) – compreende-se entre 0 e 1, isto é: $0 \leq P(E) \leq 1$; Além disso, a probabilidade deste espaço amostral é igual a 1 - P(S)=1.

A probabilidade condicional é a probabilidade de certo evento ocorra sabendo-se da ocorrência de um anterior. Por exemplo, qual é a probabilidade da soma de dois dados lançados ter resultado 10, sabendo-se que o primeiro saiu 4? Os eventos possíveis são (4,1), (4,2), (4,3), (4,4), (4,5), (4,6) e, por isso, a probabilidade do resultado ser 10 é de 1/6.

Pode-se expressar a probabilidade condicional de um evento E ocorra sabendo-se que outro evento F ocorreu, por $P(E|F) = P(E \cap F)/P(F)$. Por exemplo, considerar que uma moeda será lançada 2 vezes, qual será a probabilidade condicional de que resulte duas vezes cara (E), tomando-se que pelo menos uma cara foi observada (F)?
S = {(cara, coroa), (coroa, cara), (cara, cara), (coroa, coroa)}
E = {cara, cara};
F= {(cara, coroa), (coroa, cara), (cara, cara)}
$P(E|F) = P(E \cap F)/P(F) =$
P({(cara,cara)}) / P({(cara, coroa),(coroa, cara),(cara, cara)}) = ¼ / ¾ = 1/3

Nestes termos, uma variável aleatória pode ser entendida como o resultado de uma medição de algum parâmetro que pode gerar um valor diferente a cada medida, ou seja, diz respeito à característica do experimento que se quer estudar. Matematicamente, ela é a função que associa cada elemento de um espaço amostral a um número real. Por exemplo, se ao lançar uma moeda três vezes, obtém o seguinte espaço amostral: S = {(ccc), (kcc), (ckc), (cck), (kkk), (kkc), (kck), (ckk)}, sendo *c* representando “cara” e *k*, “coroa”. Necessita-se avaliar a quantidade de caras possíveis. Assim, a variável aleatória *X*, que representa a quantidade de “caras”, pode ser expressa da seguinte forma:
x = 0 {(kkk)}
x = 1 {(kkc)(kck)(ckk)}
x = 2 {(kcc)(ckc)(cck)}
x = 3 {(ccc)}

Uma variável aleatória discreta assume cada um dos seus valores com certa probabilidade, conforme a seguir:

x	**0**	**1**	**2**	**3**
P(X = x)	1/8	3/8	3/8	1/8

As variáveis aleatórias podem ser classificadas em:

- Discretas (VAD) – a quantidade de valores possíveis, assumidos por X, for contável e finita (ou infinita).
- Contínua (VAC) – a quantidade de valores possíveis, assumidos por X, for formada por intervalos, ou seja, por valores não-contáveis.

Exemplos:

1) Para VAD:

a) Jogar um dado não viciado:
   S = {1, 2, 3, 4, 5, 6}
   X = 1 se ponto for igual a 6
   X = 0 caso contrário
   X = {0, 1}

b) Jogar uma moeda até tirar uma cara:
   X assume a quantidade de jogadas até tirar uma cara (incluindo-se a cara)
   X = {1, 2, 3, ...}
   X assume a quantidade de coroas até tirar uma cara
   X = {0, 1, 2, ...}

2) Para VAC:

a) X distância entre dois pontos positivos:
   $X = [0,+\infty[$

b) X distância entre dois pontos quaisquer:
   $X =]-\infty,+\infty[$

Quando nos depararmos com situações em que as variáveis aleatórias são dependentes umas das outras, ou suas distribuições de probabilidade mudam com o tempo, ou ambas as coisas acontecem; estudam-se tais situações baseando-se na teoria de funções aleatórias, ou seja, na teoria de processos estocásticos.

Os termos processo estocástico e processo aleatório são sinônimos e abrangem toda a teoria de probabilidades. Na prática, entretanto, o termo processo estocástico é reservado para quando o parâmetro temporal é introduzido.

9.2. Processo Markoviano

Tomando-se $n = 1,2,..., \infty$ e qualquer sequência de estados possíveis $s_1, s_2,..., s_{n+1}$, com $X_n, X_{n-1},..., X_1$ conhecidos, tem-se para X_{n+1} como uma variável aleatória discreta:

$$P(X_{n+1} = s_{n+1} \mid X_1=s_1, X_2=s_2,..., X_n=s_n) = P(X_{n+1} = s_{n+1} \mid X_n=s_n)$$

Traduzindo: as probabilidades de todos os estados futuros X_j $(j > n)$ dependem somente do estado atual X_n, mas não dependem dos estados anteriores $X_1,..., X_{n-1}$.

Um processo Markoviano é um processo estocástico cuja dinâmica do comportamento é tal que a distribuição de probabilidade do futuro depende somente do estado presente e não levando em consideração como o processo chegou a tal estado (passado).

Os processos markovianos são modelados formalmente pelos modelos de Markov, que são sistemas de transições de estados, onde os estados são representados em termos de seus vetores probabilísticos, que podem variar no espaço temporal (discreto ou contínuo), e as transições entre estados são probabilísticas e dependem apenas do estado corrente.

Se o espaço de estados é discreto, o processo Markoviano é conhecido como cadeia de Markov.

Em uma cadeia de Markov, o símbolo p_{ij} (probabilidade de passar do estado *i* para o estado *j* em uma fase) é usado para representar a probabilidade (condicional) de que, dado que o sistema esteja no estado *i* em um certo momento, venha a estar no estado *j* no intervalo de tempo seguinte. O símbolo p_{ij} também pode ser denominado como probabilidade de transição de cadeias de Markov.

Se A é o evento do sistema no estado *i* no tempo *n* e B é evento do sistema no estado *j* no tempo *n+1*, então p_{ij} = P(A|B), ou seja, é a probabilidade de ocorrer o evento A sabendo-se que B ocorreu.

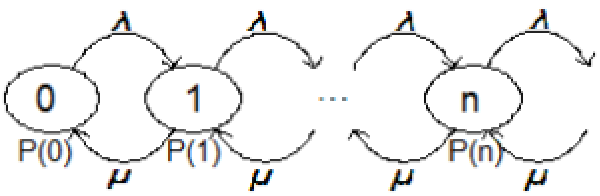
9.3. Matriz de Transição

Considere uma cadeia de Markov com estados 1,2,...,N. Seja p_{ij} a probabilidade de transição do estado *i* para o estado *j*. Então a matriz NxN P=[p_{ij}] denomina-se matriz de transição de cadeia de Markov. Por exemplo, se a cadeia de Markov tem quatro estados 1, 2, 3, 4, a matriz de transição pode ser representada assim:

$$P = \begin{bmatrix} p_{11} & \cdots & p_{14} \\ \vdots & \ddots & \vdots \\ p_{41} & \cdots & p_{44} \end{bmatrix}$$

9.4. Modelo M / M / 1

Uma fila que segue o modelo M/M/1 pode ser representada pela figura a seguir, onde os números representam os estados possíveis e as setas as possíveis transições.



Fonte: Pereira (2009)

No estado 0, o número de entradas por unidade de tempo é μP_1, que ocorre na passagem do estado 1 para o estado 0 por saída do sistema de um cliente atendido, e o número de saídas por unidade de tempo é λP_0, que ocorre na passagem do estado 0 para o estado 1 por chegada de um cliente ao sistema.

Para que o estado de equilíbrio ou estacionário ocorra é necessário que o número de entradas por unidade de tempo e o número de saídas por unidade de tempo sejam iguais, isto é: $\mu P_1 = \lambda P_0 \rightarrow P_1 = \lambda P_0/\mu$

No estado 1 o número de entradas por unidade de tempo é $\lambda P_0 + \mu P_2$, que acontece na passagem do estado 0, com a chegada de um cliente ao sistema, ou na passagem do estado 2 por saída do sistema de um cliente atendido, e o número de saídas por unidade de tempo é $\lambda P_1 + \mu P_1$, que acontece na passagem para o estado 2 por chegada de um cliente ou para o estado 0 por saída do sistema de um cliente atendido.

A expressão que representa o estado de equilíbrio da fila é: $P_n = (\lambda/\mu)^n P_0$. Para reavaliar o desenvolvimento desta expressão, deve-se se reportar a Pereira (2009).

Sabendo-se que a taxa de ocupação (ρ) é igual a λ/μ, então $P_n = (\rho)^n P_0$, para *n* sendo a quantidade de atendimentos aos clientes.

Para condição de equilíbrio da fila, tem-se:

- Taxa de ocupação: $\rho = \lambda/\mu$
- O número médio de clientes no sistema: $NS = \rho / (1 - \rho)$.
- O número médio de clientes na fila $NF = \rho^2 / (1 - \rho)$.
- O número médio de clientes esperando quando existe pelo menos uma pessoa esperando (caso especial de NF) $NF' = 1/ (1- \rho) = \mu/(\mu - \lambda)$.
- O tempo médio de clientes no sistema $TS = 1 / \mu (1 - \rho)$.
- O tempo médio de clientes na fila $TF = \rho / \mu (1 - \rho)$.
- Probabilidade de ocorrência do estado 0: $P_0 = 1 – \rho$
- Probabilidade de ocorrência do estado *n*: $P_n = \rho^n P_0$
- Probabilidade de existirem no sistema *k* ou mais clientes: $P(N \geq k) = \rho^k$
- Probabilidade de o tempo de espera na fila ser zero: $P(TF = 0) = P_0$
- Probabilidade de o tempo de espera na fila exceder *t*: $P(TF > t) = \lambda/\mu\ [e^{(\lambda - \mu)t}]$
- Probabilidade de o tempo gasto no sistema exceder *t*: $P(TS > t) = e^{(\lambda - \mu)t}$

Exemplo 1:

A Sra. Cutt possui um salão unissex. Ela não marca hora, de forma que, o atendimento ocorre numa disciplina “primeiro a chegar, primeiro a ser atendido” (FCFS). Ela acredita que está muito ocupada aos sábados pela manhã, de modo que, ela está estudando a possibilidade de contratar uma assistente e possivelmente mudar-se para um estabelecimento maior. Entretanto, antes de tomar essa decisão ela deseja saber qual o número médio de clientes no salão e qual o número médio de clientes esperando para serem atendidos.

Para tanto, a Sra. Cutt observou que os clientes chegavam ao salão de acordo com um processo de Poisson com uma taxa média de chegada de 5 clientes/h. Além disso, foi observado que o tempo de atendimento era exponencialmente distribuído e em média de 10 minutos. Sabe-se que a sala de espera do salão tem 4 assentos, e por isso ela quer saber qual é a probabilidade dos seus cliente terem que ficar de pé esperando pela sua vez.

Dados do problema:

λ = 5 clientes/h e μ = 60/10 = 6 clientes/h
Taxa de ocupação (ρ) = 5/6 = 0,83
Número médio de clientes no sistema: NS = 0,83 / (1 – 0,83) = 4,88 ≈ 5 clientes
Número médio de clientes na fila: NF = $0{,}83^2$ / (1 – 0,83) = 0,69 / 0,17 = 4,06 ≈ 4 clientes
O número médio de clientes esperando quando existe pelo menos uma pessoa esperando (caso especial de NF) NF' = 1/ (1- 0,83) = 5,88 ≈ 6 clientes

Ela também está interessada em saber a porcentagem de tempo que os clientes são atendidos sem terem que esperar em fila, ou seja, a probabilidade que uma chegada ocorra e o sistema esteja vazio é: $P_0 = 1 - \rho = 1 - 0{,}83 = 0{,}17 = 17\%$

Desse modo, 17% do tempo da Sra. Cutt está ociosa, fazendo com que o próximo cliente a chegar possa ser atendido sem ter que esperar. Devido ao fato do processo de Poisson governar sua chegada e por causa de sua propriedade completamente aleatória, a porcentagem de clientes que entrarão em serviço diretamente sem ter que esperar é 17% também. Por outro lado, 83% dos clientes devem esperar antes de serem atendidos.

A sala de espera da Sra. Cutt tem quatro assentos, de forma que, ela está interessada em saber a probabilidade de um cliente, após chegar, não ter lugar para sentar.

Pr{cliente não encontrar nenhum assento disponível} =
Probabilidade de existirem no sistema *k* ou mais clientes = $P(N \geq k) = \rho^k$
$Pr\{N \geq 5\} = \rho^5 = 0{,}83^5 = 0{,}39$

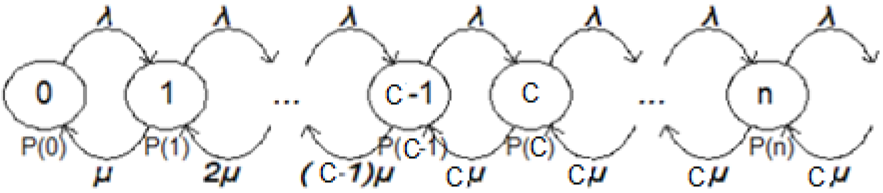
Isto implica que em quase 40% do tempo, os clientes não encontrarão lugar para sentar.

Tempo médio de atendimento no salão e tempo médio na fila são:
O tempo médio de clientes no sistema TS = 1 / μ (1 - ρ) = 1 / 6(1 – 0,83) = 0,98 ≈ 1 hora
O tempo médio de clientes na fila TF = ρ / μ (1 - ρ) = 0,83 / 6(1 – 0,83) = 0,81 horas x 60 minutos ≈ 49 minutos.

9.5. Modelo M / M / C

c: especifica o número de canais de atendimento ou número de servidores.

A próxima figura expressa o modelo M/M/C considerando uma cadeia de Markov.



Fonte: Adaptado de Pereira (2009)

Expressões (Fonte: Costa, s/d)

- Taxa de ocupação: $\rho = \lambda/C\mu$
- Probabilidade de ocorrência do estado 0: $p_0 = \left(\sum_{n=0}^{c-1} r^n/n! + \frac{r^c}{c!(1-\rho)}\right)^{-1}$ $(r/c = \rho < 1)$

  Sendo $r = \lambda/\mu$

- Probabilidade de existirem no sistema *k(n)* ou mais clientes: $p_n = \frac{\lambda^n}{c^{n-c}c!\mu^n} p_0$ $(n>=c)$.
- O número médio de clientes na fila (L_q=NF): $L_q = \left(\frac{r^c\rho}{c!(1-\rho)^2}\right) p_0$ sendo $r = \lambda/\mu$
- O número médio de clientes no sistema (L=NS): $L = r + \left(\frac{r^c\rho}{c!(1-\rho)^2}\right) p_0$
- O tempo médio de clientes na fila (W_q=TF): $W_q = L_q/\lambda = \left(\frac{r^c}{c!(c\mu)(1-\rho)^2}\right) p_0$
- O tempo médio de clientes no sistema (W=TS): $W = 1/\mu + \left(\frac{r^c}{c!(c\mu)(1-\rho)^2}\right) p_0$

Exemplo 1:

Uma clínica de olhos da cidade oferece consultas grátis todas as quartas à tarde. Existem três oftalmologistas na clínica. Um teste oftalmológico leva, em média, 20min, sendo considerado como exponencialmente distribuído ao redor desta média. Os pacientes chegam, por distribuição de Poisson, com média de 6 pacientes/h. Eles são atendidos considerando a forma FCFS. Os planejadores do hospital estão interessados em saber:
(1) Qual é o número médio de pessoas esperando (NF);
(2) A quantidade média de tempo que um paciente gasta na clínica (TS);

Dados do Problema:

C = 3, λ = 6 pacientes/h e μ = 60/20min = 3 pacientes/h

Taxa de Ocupação (ρ) = $\lambda/C\mu = 6/(3 \times 3) = 2/3 = 0,67$
$r = \lambda/\mu = 6/3 = 2$

Da expressão de P_0 $p_0 = \left(\sum_{n=0}^{c-1} r^n/n! + \frac{r^c}{c!(1-\rho)}\right)^{-1}$ $(r/c = \rho < 1)$ chega-se a:

n = c-1 = 2
$P_0 = \left(\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!(1-0,67)}\right)^{-1} = 0,11$

Da equação NF = tem-se que: $L_q = \left(\dfrac{r^c \rho}{c!(1-\rho)^2}\right) p_0$

NF = [(2^3 2/3)/3!(1-2/3)^2]1/9 = 8/9 = 0,89

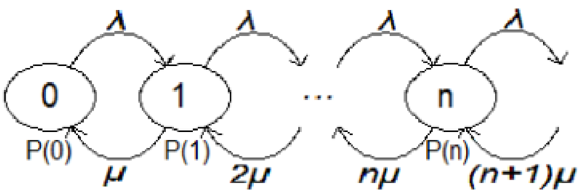
TS = $W = 1/\mu + \left(\dfrac{r^c}{c!(c\mu)(1-\rho)^2}\right) p_0$

$$TS = \frac{1}{3} + \frac{2^3}{3!\,(3.3)(1-0{,}67)^2} \times 0{,}11 = \frac{0{,}88}{5{,}94} + 0{,}33 = 0{,}48h$$

TS = 1/μ + NF/λ = 1/3 + (8/9)/6 = 13/27h = 28,9min.

9.6. Modelo de 1 Canal e 1 Fila com População Infinita (M / M / ∞) → c = ∞

A próxima figura expressa o modelo M/M/C considerando uma cadeia de Markov.



Fonte: Pereira (2009)

Expressões

Probabilidade de ocorrência do estado 0	$P_0 = e^{-\rho}$
Probabilidade de ocorrência do estado n	$P_n = \frac{\rho^n}{n!}e^{-\rho}$
Número médio de clientes no sistema	$L_s = \rho$
Comprimento médio da fila	$L_q = 0$
Tempo médio de espera no sistema	$W_s = \frac{1}{\mu}$
Tempo médio de espera na fila	$W_q = 0$
Taxa de ocupação	$\rho = \frac{\lambda}{\mu}$

Fonte: Pereira (2009)

Obs.: cuidado com as variáveis que são diferentes das usadas nesta apostila

10. BIBLIOGRAFIA

Andrade, Eduardo Leopoldino de, Introdução à pesquisa operacional, LTC - Livros Técnicos e Científicos, Rio de Janeiro, 2000.

Albernaz, Marco Aurélio, Teoria das Filas – Apontamentos da Disciplina Pesquisa Operacional II – Pontifícia Universidade Católica – RJ, 2004.

Costa, Luciano Cajado **Teoria das Filas** Universidade Federal do Maranhão, Centro Tecnológio, Disponível em http://www.deinf.ufma.br/~mario/grad/filas/TeoriaFilas_Cajado.pdf, Capturado em27/07/2010.

Costa, Renato Aurélio Castro, Determinação de Estoques, Dissertação de Mestrado, Universidade Federal do Paraná, Curitiba, 2003

Fogliatti, Maria Cristina; Mattos, Neli Maria Costa. Teoria de filas. Rio de Janeiro, Interciência, 2007.

Grigoletti, Pablo Souza **Cadeias de Markov** Escola de Informática – Universidade Católica de Pelotas, capturado de http://www.assembla.com/spaces/ptfe/documents/c5dGKM7har3jj8abIlDkbG/download/02-markov.pdf, disponível em 25/07/2010, Pelotas.

Pereira, Cláudia Rossana Velosa **Uma Introdução às Filas de Espera** Mestrado em Matemática, Universidade da Madeira - Departamento de Matemática e Engenharias, Portugal, 2009.

Prado, Darci Santos do. **Teoria das Filas e da Simulação**, ISBN 85-86948-12-8, INDG Tecnologia e Serviços LTDA, Belo Horizonte, 2004.

ANEXO

Letras Gregas

Maiúsculo/Minúsculo/Nome Port./Nome Inglês							
Αα	Alfa/Alpha	**Ηη**	Eta, Éta Usado como nano 10^{-9}	**Νν**	Ni, Nu	**Ττ**	Tau/Tau
Ββ	Beta/Beta	**Θθ**	Teta/Theta	**Ξξ**	Xi “Ksi”/ Xi”csai”	**Υυ**	Úpsilon/ Upsilon
Γγ	Gama/Gamma	**Ιι**	Iota/Iota	**Οο**	Ômicron/ Ómicron	**Φφ**	Fi/Phi
Δδ	Delta/Delta	**Κκ**	Kapa/Kappa	**Ππ**	Pi/Pi	**Χχ**	Qui/Chi
Εε	Épsilon/Epsilon	**Λλ**	Lâmbda/Lambda	**Ρρ**	Rô/Rho	**Ψψ**	Psi/Psi
Ζζ	Zeta/Zeta	**Μμ**	Mi, Um Usado como micro 10^{-6}	**Σσς**	Sigma/Sigma	**Ωω**	Ômega, Ômega