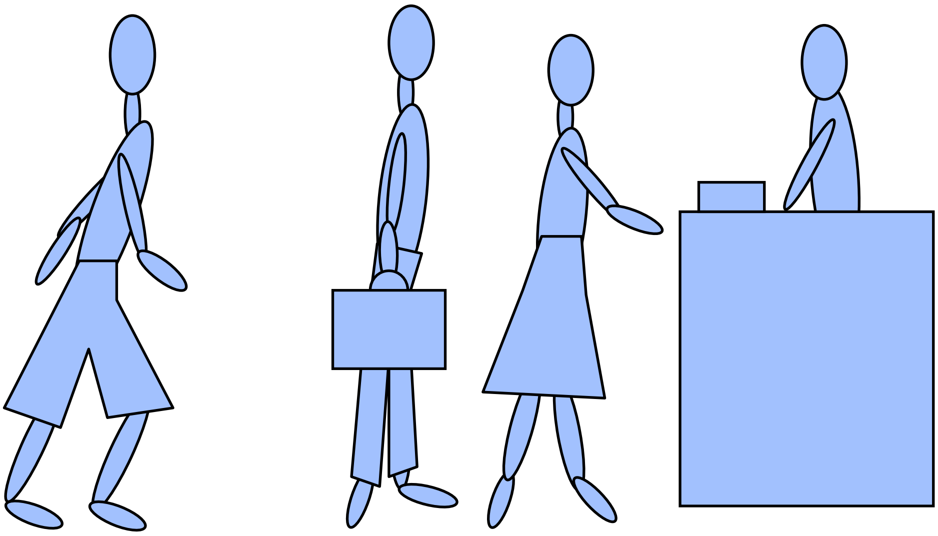
Sistemas de Filas Simples

1 Teoria de Filas



Processo de chegada: se os usuários de uma fila chegam nos instantes t_1, t_2, t_3,..., t_j, as variáveis aleatórias $\tau_j = t_j - t_{j-1}$ são chamadas de intervalos entre chegadas.

As variáveis aleatórias formam uma seqüência de valores aleatórios Independentes e Identicamente Distribuídos (IID).

O processo mais comum de chegada é o **processo de Poisson**. Neste caso os intervalos entre chegadas são distribuídos exponencialmente IID. Outras distribuições podem também ser utilizadas tais como: Determinística, Hiper-exponencial e Genérica.

Processo de serviço: O tempo que cada usuário do sistema de fila passa em serviço define o seu tempo de serviço. É usual considerar que estes tempos de serviços são variáveis aleatórias IID. A distribuição mais comum é a exponencial, porém outras distribuições também são utilizadas, tais como: determinística, Hyper-exponencial e genérica.

Número de servidores: define o número de servidores disponíveis no sistema de fila. Normalmente, estes servidores são idênticos.

Capacidade do sistema: é o número máximo de usuários que o sistema de fila pode apresentar. Este número considera tanto os usuários na fila como aqueles em serviço. Quando este parâmetro não representa nenhuma grande limitação é comum utilizar-se um valor infinito para a sua capacidade.

Tamanho da população: representa o número total de usuários em potencial, que podem chegar no sistema de fila. Na maioria dos sistemas reais a população é finita, porém se esse número é suficientemente grande, pode-se utilizar o valor infinito como tamanho da população.

Disciplina de Serviço: a ordem com que os usuários do sistema são atendidos define a disciplina de serviço ou atendimento. A disciplina mais comuns são:

- **FCFS** (First Come First Served) – O primeiro a chegar é o primeiro a ser atendido.
- **RR** (Round Robin) – Um usuário é atendido por um tempo máximo pré-definido e se não for possível realizar todo o serviço, o usuário deve ir para o final da fila.
- **LCFS** (Last Come First Served) – O último a chegar é o próximo a ser atendido.
- **LCFS-preemptivo** – O usuário que chegou por último pode interromper o usuário que está sendo atendido.
- **PS** (Processor Sharing) – Corresponde à disciplina RR com um quantum suficientemente pequeno comparado com o tempo médio de serviço. Equivale a repartir o processador em n partes iguais.

Sistemas de filas que não possuem tempo de espera são chamados de **Centros de Atraso** (Delay Center). Em geral, sistemas com **infinitos servidores** (**IS**) possuem esta característica.

Notação

Para a definição de um sistema de filas é necessário se especificar seis parâmetros:

A/S/m/B/K/SD onde

A: Distribuição dos intervalos entre chegadas;
S: Distribuição dos tempos de serviço;
m: Número de servidores;
B: Número máximo de usuários no sistema;
K: Tamanho máximo da população;
SD: Disciplina de atendimento ou serviço.

As distribuições dos intervalos entre chegadas e dos tempos de serviço são representadas em geral por uma letra:

M Exponencial;
D Determinística
G Genérica

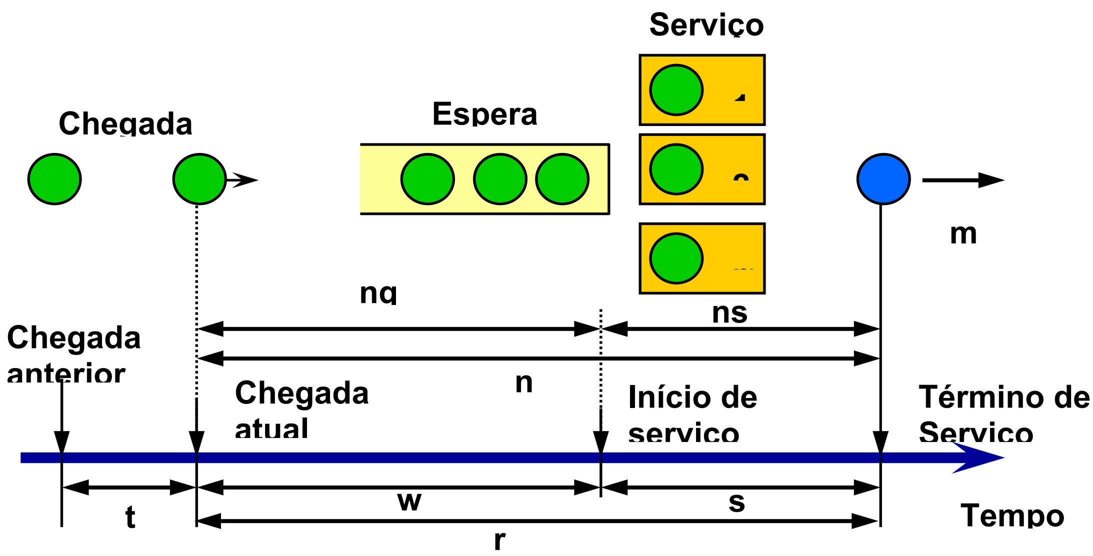
Uma **distribuição determinística** é aquela que define tempos constantes, portanto não existe nenhuma variabilidade.

Uma **distribuição genérica** significa uma distribuição não especificada. Os resultados assim obtidos são válidos para qualquer distribuição.

A **distribuição exponencial** possui a propriedade de não apresentar memória ("memoryless"), isto é, o próximo estado só depende do estado atual e não dos estados anteriores.

Se os intervalos entre chegadas são distribuídos exponencialmente com média $1/\lambda$, o tempo esperado para a próxima chegada é sempre $1/\lambda$, independente do tempo que já transcorreu desde a última chegada.

Variáveis associadas a um sistema de filas:



τ	Intervalo entre chegadas sucessivas
λ	Taxa média de chegada = $1/E[\tau]$, em alguns sistemas este parâmetro pode ser função do estado do sistema
s	Tempo de serviço de um usuário
μ	Taxa média de serviço por servidor = $1/E[s]$. A taxa total de serviço para **m** servidores é $\mathbf{m}\mu$
n	número de usuários no sistema. É chamado também de tamanho da fila, incluindo os usuários que estão em serviço
nq	Número de usuários esperando para receber serviço. É sempre menor que **n**, pois não inclui os usuários em serviço
ns	Número de usuários em serviço no sistema
r	Tempo de resposta ou simplesmente tempo no sistema. Inclui tanto o tempo de espera como o tempo de serviço
w	Tempo de espera, isto é, intervalo de tempo entre o instante de chegada e o início do serviço

Todas as variáveis acima, com exceção de λ e μ, são variáveis aleatórias.

2 Relações entre as variáveis

As relações discutidas a seguir são válidas para o sistema G/G/m (ou seja, para qualquer sistema de filas):

Condição de estabilidade: se o número de usuários cresce continuamente e se torna infinito, o sistema é denominado instável. Para que um sistema seja estável é necessário que a seguinte relação seja verdadeira:

$$\lambda < m\mu$$

Relação entre o número de usuários no sistema e número na fila:

$$n = nq + ns$$

Observe que estas variáveis são aleatórias. Em particular esta igualdade leva a uma outra igualdade entre os seus valores médios:

$$E[n] = E[nq] + E[ns]$$

Se a taxa de serviço de cada servidor é independente do número de usuários na fila, ou seja,

$$Cov(nq,ns) = 0$$

Então

$$Var[n] = Var[nq] + Var[ns]$$

Número x Tempo: Se usuários não são perdidos, o número médio de usuários no sistema está relacionado ao seu tempo de resposta da seguinte maneira:

$$E[n] = \lambda * E[r]$$

ou de maneira semelhante:

$$E[nq] = \lambda * E[w]$$

Estas equações são conhecidas como **Resultado de Little** e são válidas desde que o sistema esteja em situação de equilíbrio, isto é, em um intervalo grande de observação, o número de saídas é igual ao número de chegadas ao sistema. Este resultado é independente das distribuições dos intervalos de chegada e dos tempos de atendimento,

Tempo no sistema x Tempo na fila: o tempo no sistema é expresso como a soma do tempo de espera com o tempo de serviço:

$$r = w + s$$

Note que **r**,**w** e **s** são variáveis aleatórias. Portanto existe entre seus valores médios a seguinte relação:

$$E[r] = E[w] + E[s]$$

e quando o tempo de serviço é independente do número de usuários na fila então:

$$Cov(w,s) = 0$$

e

Var[r] = Var[w] + Var[s]

Resultado de Little

Suponha que o sistema seja monitorado por um intervalo de tempo **T** e que seja mantido um registro sobre cada instante de chegada ou de partida. Se **T** for suficientemente grande o número de chegadas se aproxima do número de partidas. Seja esse número igual a **N**.
Então

Taxa de chegada = Total de chegadas/Tempo total = N/T

Seja **J** a soma do tempo gasto no sistema por todos os usuários que passaram pelo sistema no intervalo **T**, então se podem escrever as seguintes relações:

E[r] = J/N Tempo médio no sistema
E[n] = J/T Número médio de usuários no sistema

Portanto:

E[n] = (N/T)*(J/N)

Ou seja,

E[n] = λ * E[r] **Resultado de Little**

3 Processos Estocásticos

Uma seqüência de variáveis aleatórias idênticas indexadas por um mesmo parâmetro, por exemplo, tempos, formam um processo estocástico.

Exemplo: O número de usuários na fila de um sistema em função do tempo, w(t) é um processo estocástico.

Os processos mais comuns utilizados nos sistemas de fila são:

Processos estocásticos com estado discreto

Dependendo dos valores que a variável aleatória pode tomar, podem-se ter processos com estados discretos ou contínuos. Por exemplo: O número de usuários na fila é um processo discreto.

Processo de Markov

Se os estados futuros do processo dependem exclusivamente do estado atual, o processo é chamado de Markov, como no caso de variáveis aleatórias com distribuição exponencial que possuem a propriedade "memoryless". Neste caso, o tempo em que o sistema se encontra no estado atual não afeta o próximo intervalo de tempo.

Cadeia de Markov

Um processo de Markov com estados discretos é chamado de cadeia de Markov.

Os sistemas de fila do tipo M/M/m podem ser modelados como uma cadeia de Markov considerando-se como estado o número de usuários na fila.

Processo Nascimento e Morte

Os processos de Markov que possuem um espaço de estado discreto e no qual as transições de estado só podem ocorrer entre estados vizinhos, são denominados de processos de Nascimento e Morte.

Neste caso, pode-se representar os estados por um número inteiro N e as possíveis mudanças de estado são para (N+1) ou (N-1).

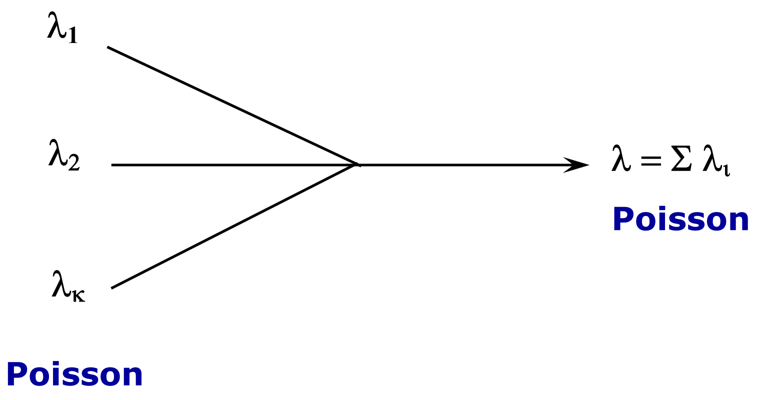
Por exemplo, o número de usuários N em um sistema de filas que apresenta somente chegadas individuais pode ser modelado por um processo de Nascimento e Morte.

Processo de Poisson

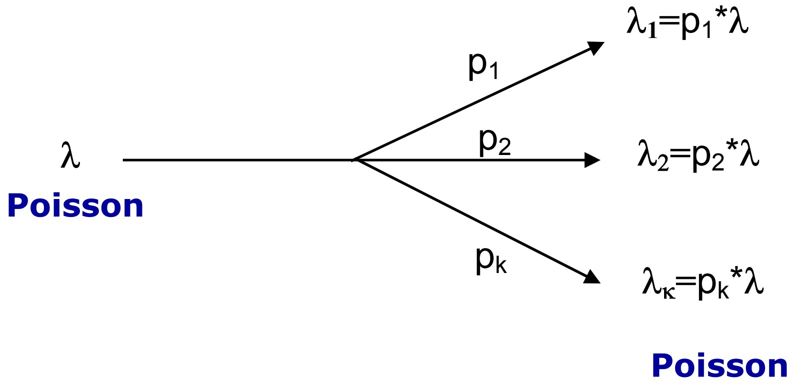
Se os intervalos entre chagadas são IID e exponencialmente distribuídos, o número de chegadas num intervalo de tempo (t,t+x) possui uma distribuição de Poisson.

Os processos de Poisson são muito utilizados nos sistemas de filas e apresentam algumas propriedades interessantes.

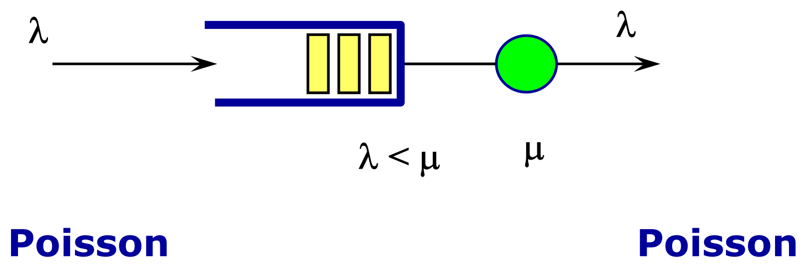
a) Junção de processos de Poisson:



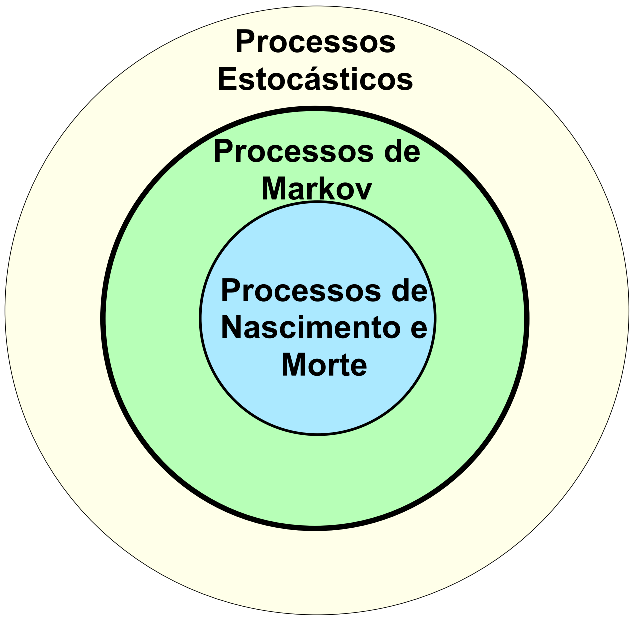
b) Distribuição de processos de Poisson



c) Partida de um sistema M/M/1

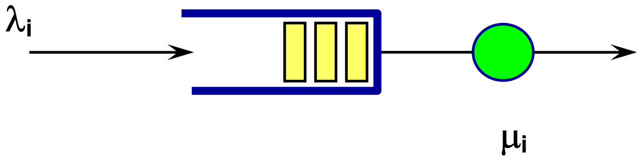


Relação entre os diversos processos estocásticos



4 Análise de Fila Única

Seja um sistema de fila única:



Este sistema pode ser descrito como um Processo de Nascimento e Morte em que o estado é representado pelo número de usuários no sistema (ver capítulo de Noções de Processos Estocásticos e de cadeia de Markov).

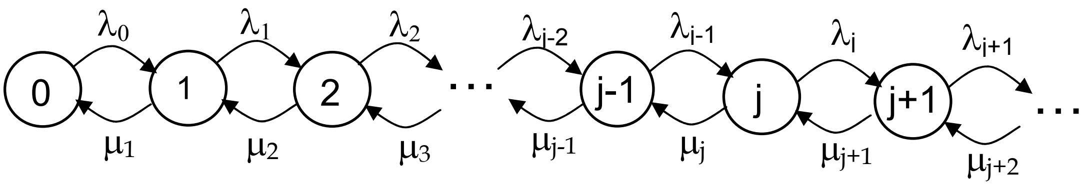


Diagrama de Transição de Estados

Probabilidades em Equilíbrio

Consideremos que em um instante t o sistema está no estado j, isto é, existem j usuários no sistema.

p_j (t) é a probabilidade do sistema estar no estado j no instante t, isto é, de possuir j usuários.

O cálculo de p_j (t) na situação de equilíbrio pode ser feito através de uma equação de balanceamento de fluxo de probabilidades em cada estado da Cadeia de Markov. A última equação necessária é a soma das probabilidades de todos os estados que é igual a um.

Em uma intervalo de tempo Δt, o sistema poderá continuar no estado j ou mudar para o estado j-1 ou j+1 de acordo com as probabilidades:

Pr[ir para o estado j+1 | está no estado j] = $\lambda_j \Delta t$
Pr[ir para o estado j-1 | está no estado j] = $\mu_j \Delta t$
Pr[continuar no estado j] = $(1 - \lambda_j \Delta t - \mu_j \Delta t)$

O intervalo de tempo Δt deve ser pequeno o suficiente para não haver dois eventos simultâneos neste intervalo. Podemos montar um conjunto de equações para as transições de estado:

$$p_0(t + \Delta t) = (1 - \lambda_0 \Delta t)\, p_0(t) - \mu_1 \Delta t\, p_1(t)$$
$$p_1(t + \Delta t) = \lambda_0 \Delta t\, p_0(t) + (1 - \lambda_1 \Delta t - \mu_1 \Delta t)\, p_1(t) + \mu_2 \Delta t\, p_2(t)$$
$$p_2(t + \Delta t) = \lambda_1 \Delta t\, p_1(t) + (1 - \lambda_2 \Delta t - \mu_2 \Delta t)\, p_2(t) + \mu_3 \Delta t\, p_3(t)$$
$$\dots$$
$$p_j(t + \Delta t) = \lambda_{j-1} \Delta t\, p_{j-1}(t) + (1 - \lambda_j \Delta t - \mu_j \Delta t)\, p_j(t) + \mu_{j+1} \Delta t\, p_{j+1}(t)$$

A última equação pode ser escrita como:

$$p_j(t + \Delta t) = \lambda_{j-1} \Delta t\, p_{j-1}(t) + p_j(t) - (\lambda_j + \mu_j)\, \Delta t\, p_j(t) + \mu_{j+1} \Delta t\, p_{j+1}(t)$$

$$p_j(t + \Delta t) - p_j(t) = \lambda_{j-1} \Delta t\, p_{j-1}(t) - (\lambda_j + \mu_j)\, \Delta t\, p_j(t) + \mu_{j+1} \Delta t\, p_{j+1}(t)$$

$$(p_j(t + \Delta t) - p_j(t))/\Delta t = \lambda_{j-1}\, p_{j-1}(t) - (\lambda_j + \mu_j)\, p_j(t) + \mu_{j+1}\, p_{j+1}(t)$$

No limite quando $\mathbf{\Delta t \to 0}$ tem-se:

$$\lim_{\Delta t \to 0} \frac{p_j(t + \Delta t) - p_j(t)}{\Delta t} = \lambda_{j-1} p_{j-1}(t) - (\lambda_j + \mu_j) p_j(t) + \mu_{j+1} p_{j+1}(t)$$

Então

$$\frac{dp_j(t)}{dt} = \lambda_{j-1} p_{j-1}(t) - (\lambda_j + \mu_j) p_j(t) + \mu_{j+1} p_{j+1}(t)$$

Quando o sistema está na, situação de equilíbrio, isto é, estável, a probabilidade de estar em um estado **j** não varia com o tempo o que significa que a derivada de $p_j(t)$ é nula.

$$0 = \lambda_{j-1} p_{j-1}(t) - (\lambda_j + \mu_j) p_j(t) + \mu_{j+1} p_{j+1}(t)$$

Neste caso, $p_j(t)$ será indicado como p_j .

$$0 = \lambda_{j-1} p_{j-1} - (\lambda_j + \mu_j) p_j + \mu_{j+1} p_{j+1} \qquad \text{para } j=1,2,\dots,\infty$$

$$p_1 = \frac{\lambda_0 p_0}{\mu_1}$$

$$p_{j+1} = \frac{(\lambda_j + \mu_j) p_j}{\mu_{j+1}} - \frac{\lambda_{j-1} p_{j-1}}{\mu_{j+1}}$$

Calculando p_j em função de p_0 temos:

$$p_1 = \frac{\lambda_0 p_0}{\mu_1}$$

$$p_2 = \frac{(\lambda_1 + \mu_1)p_1}{\mu_2} - \frac{\lambda_0 p_0}{\mu_2} = \frac{(\lambda_1 + \mu_1)}{\mu_2}\frac{\lambda_0 p_0}{\mu_1} - \frac{\lambda_0 p_0}{\mu_2} = \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} p_0$$

Para n=1,2,..., ∞

$$p_n = \frac{\lambda_0 \lambda_1 \lambda_2 \ldots \lambda_{n-1}}{\mu_1 \mu_2 \mu_3 \ldots \mu_n} p_0$$

A probabilidade de equilíbrio de um processo nascimento e morte se encontrar num determinado estado é:

$$p_n = p_0 \prod_{j=0}^{n-1} \frac{\lambda_j}{\mu_{j+1}}$$

Para n=1,2,..., ∞

Podemos determinar p_0, que é a probabilidade do sistema possuir zero usuários, considerando que

$$\sum_{n=0}^{\infty} p_n = 1$$

Cálculo de p_0:

$$1 = p_0 + \sum_{n=1}^{\infty} p_n = p_0 + \sum_{n=1}^{\infty} p_0 \prod_{j=0}^{n-1} \frac{\lambda_j}{\mu_{j+1}}$$

$$1 = p_0 (1 + \sum_{n=1}^{\infty} \prod_{j=0}^{n-1} \frac{\lambda_j}{\mu_{j+1}})$$

Desta forma, a solução geral de um Processo Nascimento e Morte é dada por:

$$p_0 = \frac{1}{1 + \sum_{n=1}^{\infty} \prod_{j=0}^{n-1} \frac{\lambda_j}{\mu_{j+1}}}$$

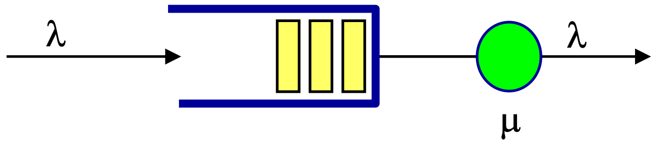
$$p_n = p_0 * \prod_{j=0}^{n-1} \frac{\lambda_j}{\mu_{j+1}}$$

5 Filas Markovianas

São as filas únicas com chegada e atendimento pelo processo de Poisson, indicadas de forma genérica como M/M/m/B sendo M o número de servidores, e B o número máximo de usuários no sistema. Neste caso considera-se que o sistema se encontra na situação de equilíbrio em que a taxa de chegada não se altera, sendo igual a λ, e as taxas de serviço de dos servidores não se alteram, sendo que todos servidores tem a mesma taxa de atendimento μ.

5.1 Fila Única M/M/1

O sistema M/M/1 se encontra na situação de equilíbrio em que os λ e μ não se alteram conforme o número de usuários no sistema.



$\lambda < \mu$

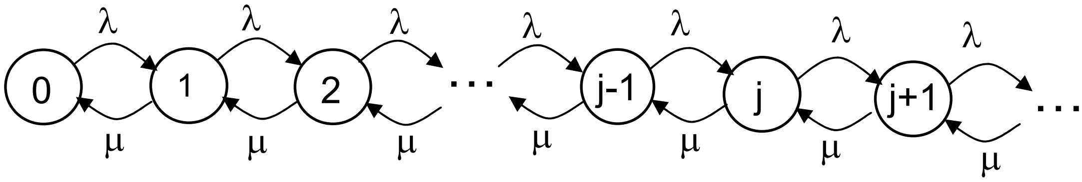


Diagrama de transição de estados

Para este sistema tem-se:

$$\lambda_n = \lambda$$

$$\mu_n = \mu$$

Aplicando a forma geral de solução do processo nascimento e morte, chega-se:

$$p_n = \left(\frac{\lambda}{\mu}\right)^n * p_0$$

onde λ/μ é chamada de intensidade de tráfego e é usualmente representada por ρ, portanto

$$p_n = p_0 * \rho^n \text{, onde } p_0 = 1 - \rho$$

Baseado nesta solução pode-se derivar os principais parâmetros do sistema M/M/1:

a. Fator de utilização do servidor: U

$$U = 1 - P_0 = \rho$$

b. Número médio de usuários no sistema:

$$E(n) = \sum_{n=1}^{\infty} n.P_n = \sum_{n=1}^{\infty} n.(1-\rho).\rho^n = \frac{\rho}{(1-\rho)}$$

c. Variância do número de usuários no sistema:

$$Var(n) = E(n^2) - E(n)^2 = \left[\sum_{n=1}^{\infty} n^2.(1-\rho).\rho^n\right] - E(n)^2 = \frac{\rho}{(1-\rho)^2}$$

d. Probabilidade de se ter n ou mais usuários no sistema:

$$p_{\geq n} = \sum_{j=n}^{\infty} p_j = \sum_{j=n}^{\infty} \rho^j.(1-\rho) = \rho^n$$

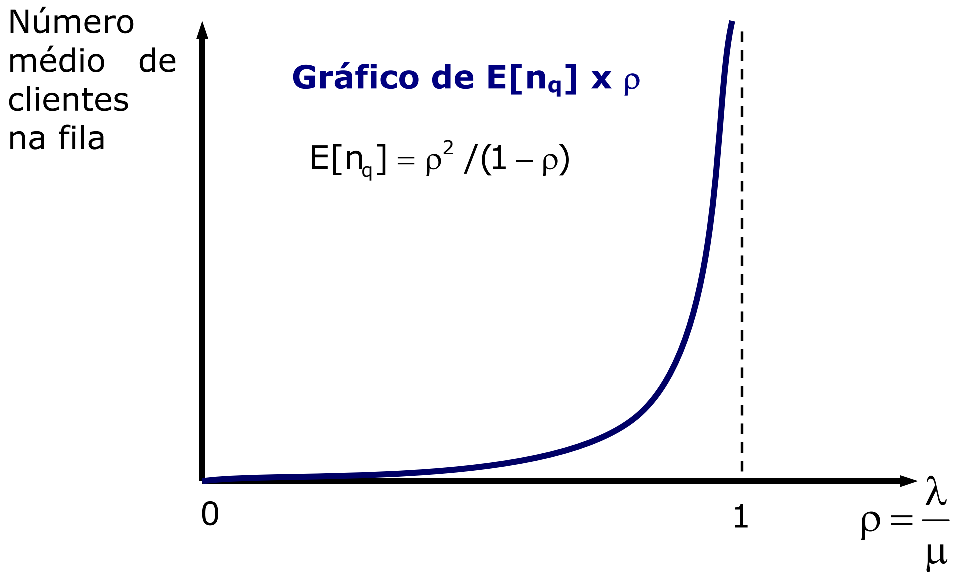
e. Tempo médio de resposta: E[r]

Considerando que E[n] = λ * E[r] pelo Resultado de Little tem-se que:

$$E(r) = \frac{E(n)}{\lambda} = \frac{\rho}{(1-\rho)}.\frac{1}{\lambda} = \frac{\frac{1}{\mu}}{1-\rho}$$

f. A distribuição de probabilidades do tempo de resposta no sistema **F(r)** é uma distribuição exponencial:

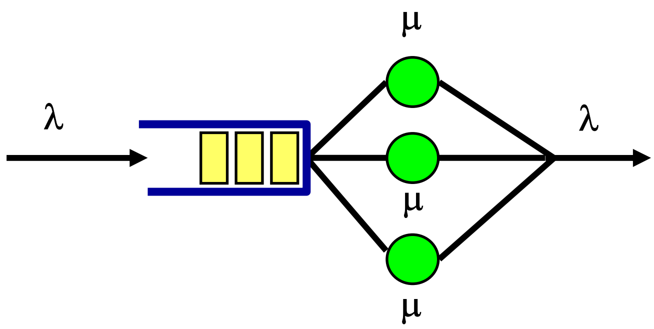
$$F(r) = 1 - e^{-r\mu(1-\rho)}$$



Observar: quando ρ se aproxima de 1 o tempo na fila e o número de clientes na fila tendem a infinito.

5.2 Fila Única M/M/m

Este sistema possui uma fila e m servidores cada um com taxa de atendimento μ.



Conforme os usuários entram no sistema, pelo fato de servidores entrarem em operação, a taxa de serviço vai aumentando em múltiplos de μ, até que todos os servidores estejam ocupados. A partir deste ponto a taxa de atendimento se mantém igual a mμ.

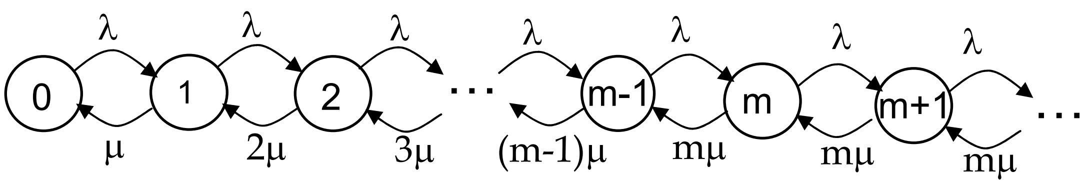
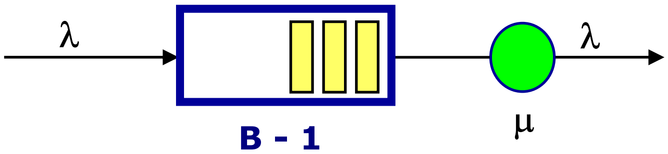


Diagrama de transição de estados

5.3 Fila Única M/M/1/B

Este sistema possui uma fila e um servidor com taxa de atendimento μ, sendo que a fila possui tamanho B-1, isto é, o limite no número de usuários no sistema é B.



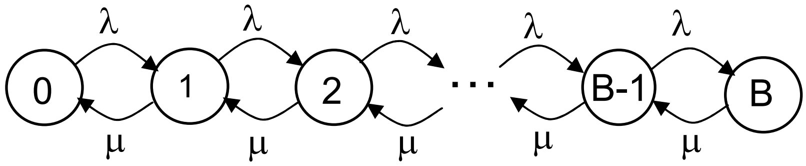
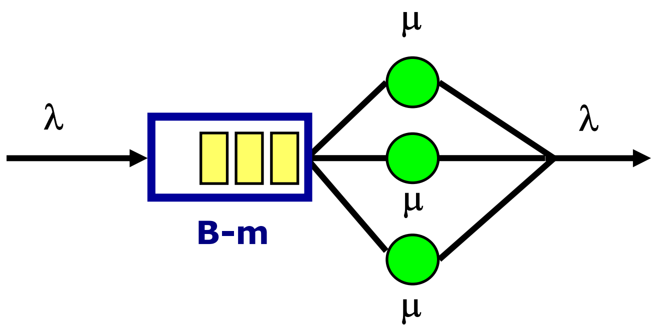


Diagrama de transição de estados

5.4 Fila Única M/M/m/B

Este sistema possui uma fila e m servidores cada um com taxa de atendimento μ sendo que a fila possui tamanho B-m, isto é, o limite no número de usuários no sistema é B.



Conforme os usuários entram no sistema, pelo fato de servidores entrarem em operação, a taxa de serviço vai aumentando em múltiplos de μ. Tem-se duas situações possíveis: B≤m e B>m.

a) Se B≤m

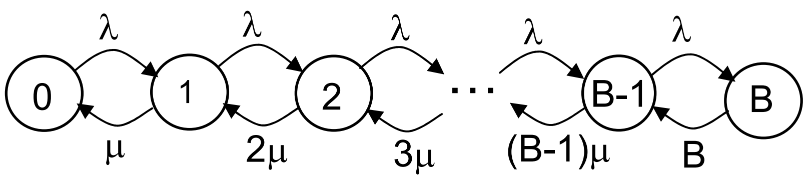


Diagrama de transição de estados

b) Se B>m

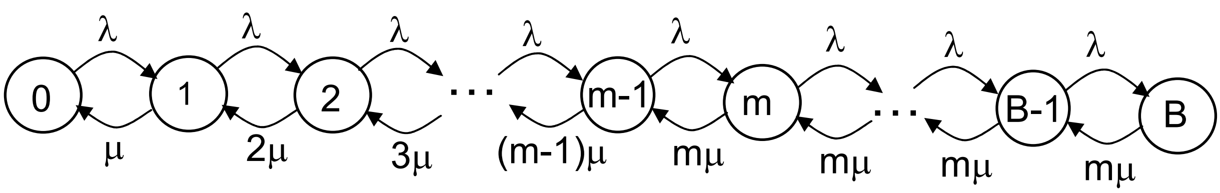


Diagrama de transição de estados

6 Bibliografia

[1] Jain, R., "The Art of Computer Systems Performance Analysis", John Wiley & Sons Inc, ISBN: 0-471-50336-3, 1991, 685 p.
[2] Cassandras, C. G., "Discrete Event Systems: Modeling and Performance Analysis", Aksen Associates Incorporated Publishers, 1993 , ISBN: 0-256-11212-6, 790p.
[3] Menascé, D. A., Almeida, V. A. F., "Scaling E-Business: Technologies, Models, Performance and Capacity Planning", Prentice-Hall, ISBN: 0-13-086328-9, 2000, 449p.
[4] Marsan, M. A., Balbo, G., Conte, G., Donatelli, S., Franceschinis, G., "Modeling with Generalized Stochastic Petri Nets", John Wiley & Sons, ISBN: 0-471-93059-8, 1995, 301p.

7 Exercícios

1) Durante uma hora de observação, um servidor de nomes de um sistema recebeu 10.800 requisições. O tempo médio de resposta observado a essas requisições foi de 1/3 de segundo. Qual o número médio de consultas no servidor? O que deve ser assumido a respeito do sistema? O número médio de consultas seria diferente se o tempo de serviço não fosse distribuído exponencialmente?
   Resp.: 1, fluxo balanceado, não pois foi utilizada a fórmula de Little que não exige distribuição exponencial.

2) O tempo médio de resposta em um sistema de bases de dados do campus de uma universidade é 3 segundos. Durante um período de observação de 1 minuto, o tempo ocioso no sistema foi medido como 10 segundos. Usando o modelo M/M/1 para o sistema determine o seguinte:
   a) Utilização do sistema
   b) Tempo médio de serviço por consulta
   c) Número de consultas completadas durante o intervalo de observação
   d) Número médio de consultas no sistema
   e) Probabilidade do número de consultas no sistema ser maior que 10
   f) Tempo de resposta em 90%
   g) Tempo de espera em 90%.

   Resp.: a) 5/6; b) 0,5 s; c)100; d) 5; e) 0,135; f) 6,9 s; g) 6,36 s.

3) Considerando que o tempo médio de resposta no exercício anterior não está aceitável, a universidade está analisando uma das seguintes alternativas. Qual dessas alternativas garante um tempo de resposta menor?
   a) Substituir o computador por um que seja duas vezes mais rápido.
   b) Colocar outro computador idêntico ao primeiro em outro lugar do campus.

4) Um servidor de rede possui uma unidade de disco com tempo médio para processamento de uma requisição de Entrada/Saída de 50 ms. A taxa de chegada de requisições é de 16 requisições por segundo.
   Determine:
   a) Fator de utilização do disco.

b) Probabilidade de o servidor estar ocioso.
c) Probabilidade de ter que esperar na fila.
d) Número médio de requisições no servidor.
e) Número médio de requisições esperando na fila.
f) Tempo médio de resposta a uma requisição.
g) Variância do tempo de resposta.
h) Tempo de espera em 90%.

5) Para melhorar o tempo de resposta do servidor da questão anterior devem ser analisadas as seguintes alternativas:
   I. Colocar mais um disco com mesmo tempo de atendimento (50 ms por requisição) e com uma fila atendendo aos dois discos.
   II. Colocar mais um disco com mesmo tempo de atendimento (50 ms por requisição) e com uma fila para cada disco.
   III. Trocar o disco por outro com o dobro da velocidade.

   Em relação às três alternativas responda:
   a) Qual o tempo de resposta e o tempo de espera na fila para cada uma das alternativas,
   b) Qual a melhor solução e qual critério utilizou nesta decisão.

6) Um sistema de armazenamento de um computador consiste de 3 unidades de disco compartilhando uma fila comum. O tempo médio de atendimento de uma requisição de E/S é de 50 ms. As requisições de E/S chegam ao sistema a uma taxa de 30 requisições por segundo. Utilizando o modelo M/M/3 para este sistema, determine:
   a) Taxa de utilização média dos discos.
   b) Probabilidade de o sistema estar ocioso, p_0.
   c) Probabilidade de ter que esperar na fila.
   d) Número médio de requisições no sistema, E[n].
   e) Número médio de requisições esperando na fila, $E[n_q]$.
   f) Tempo médio de resposta, E[r].
   g) Variância do tempo de resposta.
   h) Tempo de espera em 90%.

   Resp.: a) 0,5; b) 0,21; c) 0,24; d) 1,7; e) 0,25; f) 0,0579s; g) 0,00296s^2; h) 0,0287s.

7) Repita o exercício anterior assumindo que uma fila separada é mantida para cada unidade de disco do sistema. Assuma também a mesma taxa de chegada de requisições.
   Resp.: a) 0,5; b) 0,5; c) 0,5; d) 1; e) 0,5; f) 0,1s; g) 0,01s2; h) 0,16s.

8) Assumindo que existem apenas 4 buffers no exercício 6, determine:
   a) Probabilidade p_n de n requisições no sistema, n=0, 1, 2, 3 e 4.
   b) Número médio de requisições no sistema, E[n].
   c) Número médio de requisições na fila, $E[n_q]$.
   d) Variância do número de requisições no sistema Var[n].
   e) Taxa de chegada efetiva.
   f) Taxa de perda de requisições.
   g) Utilização das unidades de disco.
   h) Tempo médio de resposta.

Resp.: a) 0,22; 0,34; 0,25; 0,13; 0,0629 b) 1,5 req; c) 0,0629; d) 1,3; e) 28; f) 1,9; g) 0,47; h) 0,0522.

9) 50% dos alunos que vão à sala de um professor para esclarecer dúvidas desistem se o professor já está ocupado e 100% desistem se o professor está ocupado e já existe um colega à espera. Desenhe a cadeia de Nascimento e Morte deste sistema considerando que a taxa inicial de chegada de alunos é λ e o tempo médio de atendimento de cada aluno é $1/\mu$. Determine as probabilidades de equilíbrio deste sistema. Qual o número médio de alunos sendo atendidos se λ=2 alunos/hora e $1/\mu$=15 minutos.

10) Um supermercado oferece R$10,00 aos seus próximos 1000 clientes que encontrarem todas as caixas ocupadas. Considere que existe uma fila única para todos os caixas. Os clientes chegam à fila a uma taxa de 3 clientes por minuto e a taxa de serviço de cada caixa é 2 clientes por minuto. O gerente do supermercado quer saber quantos caixas deve manter em funcionamento para que seu orçamento de R$2.500,00 não seja excedido, isto é, a promoção não deve gastar mais que R$2.500,00 para esses primeiros 1000 clientes.

11) Um banco quer determinar quantas ATMs (máquinas de auto-atendimento) devem ser instaladas em uma localidade onde chegam 40 clientes por hora, sendo que uma transação leva em média 1 minuto para ser processada. O banco cobra R$ 1,00 por transação e assume que quando existem mais de 2 pessoas na fila o cliente desiste. Supondo que a chegada é Poisson e o tempo de processamento da transação é exponencial determine quantas ATM devem ser colocadas de forma que a perda esperada por hora seja inferior a R$5,00?

8 Fórmulas

	M/M/1	M/M/m
1. Taxa de chegada Taxa de serviço Número de Servidores	λ μ 1	λ μ m
2. Fator de utilização	$U = \rho = \frac{\lambda}{\mu}$	$U = \rho = \frac{\lambda}{m \cdot \mu}$
3. Probabilidade de zero usuários no sistema	$p_0 = 1 - \rho$	$p_0 = \left[1 + \frac{(m \cdot \rho)^m}{m!(1-\rho)} + \sum_{n=1}^{m-1} \frac{(m \cdot \rho)^n}{n!} \right]^{-1}$
4. Probabilidade de n usuários no sistema	$p_n = (1-\rho) \cdot \rho^n$ n= 0,1,2,...,∞	$p_n = p_0 \frac{(m \cdot \rho)^n}{n!}$ n<m $p_n = p_0 \frac{m^m \cdot \rho^n}{m!}$ n≥m
5. Probabilidade de n ou mais usuários no sistema	$p_{\geq n} = \rho^n$	
6. Probabilidade de esperar na fila	$p_{\geq 1} = \rho$	$\varsigma = P(\geq m \text{ usuários}) = \frac{(m\rho)^m}{m!(1-\rho)} P_0$
7. Número médio de usuários no sistema	$E[n] = \rho / (1-\rho)$	$E[n] = m\rho + \rho\varsigma / (1-\rho)$
8. Variância do número de usuários no sistema	$Var[n] = \rho / (1-\rho)^2$	$Var[n] = m\rho + \rho\varsigma \left[\frac{1 + \rho - \rho\varsigma}{(1-\rho)^2} + m \right]$

	M/M/1	**M/M/m**
9. Tempo médio de resposta	$E[r]=(1/\mu)/(1-\rho)$	$E[r]=\frac{1}{\mu}\left[1+\frac{\varsigma}{m(1-\rho)}\right]$
10.Variância do tempo de resposta	$Var[r]=(1/\mu^2)/(1-\rho)^2$	$Var[r]=\frac{1}{\mu^2}\left[1+\frac{\varsigma(2-\varsigma)}{m^2(1-\rho)^2}\right]$
11.Probabilidade de k usuários na fila	$p(n_q=k)=1-\rho^2$ k=0 $p(n_q=k)=(1-\rho)\cdot\rho^{k+1}$ k>0	
12.Número médio de usuários na fila	$E[n_q]=\rho^2/(1-\rho)$	$E[n_q]=\rho\varsigma/(1-\rho)$
13.Variância do número de usuários na fila	$Var[n_q]=\rho^2(1+\rho-\rho^2)/(1-\rho)^2$	$Var[n_q]=\rho\varsigma(1+\rho-\rho\varsigma)/(1-\rho)^2$
14.Tempo médio de espera	$E[w]=\rho/[\mu(1-\rho)]$	$E[w]=\varsigma/[m\mu(1-\rho)]$
15.Variância do tempo de espera	$Var[w]=(2-\rho)\rho/[\mu^2(1-\rho)^2]$	$Var[w]=\varsigma(2-\varsigma)/[m^2\mu^2(1-\rho)^2]$
16.Tempo de resposta em q%	$r_{q\%}=E[r]\ln\left(\frac{100}{100-q}\right)$	
17.Tempo de espera em q%	$w_{q\%}=\max\left[0,\frac{E[w]}{\rho}\ln\left(\frac{100\rho}{100-q}\right)\right]$	$w_{q\%}=\max\left(0,\frac{E[w]}{\varsigma}\ln\frac{100\varsigma}{100-q}\right)$

	M/M/1/B	**M/M/m/B**
1. Taxa de chegada Taxa de serviço Número de servidores Número de usuários	λ μ 1 B ($B \geq 1$)	λ μ m B ($B \geq m$)
2. Fator de utilização	$U = \rho(1-P_B)$ sendo $\rho = \lambda/\mu$ O sistema é sempre estável ($\rho < \infty$)	$U = \rho(1-P_B)$ sendo $\rho = \lambda/(m\mu)$ O sistema é sempre estável ($\rho < \infty$)
3. Probabilidade de zero usuários no sistema	$p_0 = \frac{(1-\rho)}{(1-\rho^{B+1})}$ $\rho \neq 1$ $p_0 = \frac{1}{(B+1)}$ $\rho = 1$	$p_0 = \left[1 + \frac{(1-\rho^{B-m+1})(m \cdot \rho)^m}{m!(1-\rho)} + \sum_{n=1}^{m-1} \frac{(m \cdot \rho)^n}{n!}\right]^{-1}$
4. Probabilidade de n usuários no sistema	$p_n = \frac{(1-\rho)}{(1-\rho^{B+1})}\rho^n$ $\rho \neq 1$ e $0 \leq n \leq B$ $p_n = \frac{1}{(B+1)}$ $\rho = 1$ e $0 \leq n \leq B$ $p_n = 0$ $n > B$	$p_n = p_0 \frac{(m \cdot \rho)^n}{n!}$ $0 \leq n \leq m$ $p_n = p_0 \frac{m^m \cdot \rho^n}{m!}$ $m \leq n \leq B$
5. Taxa de chegada efetiva	$\lambda' = \lambda(1-P_B)$	$\lambda' = \lambda(1-P_B)$
6. Taxa de perda	λP_B	λP_B
7. Número médio de usuários no sistema	$E[n] = \frac{\rho}{1-\rho} - \frac{(B+1)\rho^{B+1}}{1-\rho^{B+1}}$	
8. Número médio de usuários na fila	$E[n_q] = \frac{\rho}{1-\rho} - \rho\frac{1+B\rho^B}{1-\rho^{B+1}}$	
9. Tempo médio de resposta	$E[r] = E[n]/[\lambda(1-P_B)]$	$E[r] = E[n]/[\lambda(1-P_B)]$
10. Tempo médio de espera	$E[w] = E[n_q]/[\lambda(1-P_B)]$	$E[w] = E[n_q]/[\lambda(1-P_B)]$

Fórmulas de Progressões Geométricas

Sendo

a_0 - primeiro termo
a_n - último termo
q - razão

1. Soma dos n primeiros termos de uma PG finita

$$S_n = a_0 + a_1 + ... + a_n = \sum_{i=0}^{n} a_i q^n$$

$$S_n = \frac{a_0(1 - q^{n+1})}{(1 - q)}$$

2. Soma de uma PG infinita (série) com razão q < 1

$$S_n = a_0 + a_1 + ... + a_n + ... = \sum_{i=0}^{\infty} a_i q^n$$

$$S_n = \frac{a_0}{(1 - q)}$$