LCF 280 – Métodos Quantitativos para a Gestão Ambiental

Modelagem de processos estocásticos

Apontamentos para estudo especial sobre as
as distribuições mais importantes para a Teoria de Filas

Referência: (Taha, R.A. 1997 *Operations Research – An Introduction*. Prentice Hall, New Jersey. 916p. – pág. 489-496; e 607-637)

Distribuição Binomial

Um processo produz em lotes com *n* itens. A fração *p* com itens defeituosos por lote é estimada a partir de dados históricos. A questão é determinar a função densidade de probabilidade (fdp)[1] do número de defeitos por lote. Existem $C_x^n = \frac{n!}{x!(n-x)!}$ combinações diferentes ao considerar a existência de *x* defeitos por lote de *n* itens, e a probabilidade de obter cada uma dessas combinações é $p^x (1-p)^{n-x}$. Pela lei da adição de probabilidades, deduz-se que:

$$P\{x=k\} = C_k^n \; p^k (1-p)^{n-k} \quad k = 1, 2, ..., n$$

Essa é a distribuição binomial com parâmetros *n* e *p*. A média e variância são:

$$E\{x\} = n\,p$$
$$var\,\{x\} = n\,p\,(1-p)$$

Distribuição Poisson

Clientes chegam de forma totalmente ao acaso (randomicamente), ou seja, é impossível prever quando alguém chegará. A fdp que descreve o número desse tipo de evento (chegadas) durante um período de tempo segue a distribuição Poisson. Seja *x* o número de eventos (p.ex.: chegadas) num determinado período de tempo (p.ex.: minuto ou hora), e a fdp Poisson pode ser definida da seguinte forma:

$$P\{x=k\} = \frac{\lambda^k e^{-\lambda}}{k!} \qquad k= 1, 2, ..., n$$

A média e variância são:

$$E\{x\} = \lambda$$
$$var\,\{x\} = \lambda$$

Intuitivamente, $E\{x\} = \lambda$ deve representar o número médio de eventos que ocorrem por unidade de tempo. Essencialmente, o parâmetro λ é definido como uma taxa (número por

---

[1] Para cada variável aleatória discreta ou contínua *x* corresponde uma função densidade de probabilidade (fdp) – *p(x)* ou *f(x)* – que lhes atribui medidas de probabilidade. A probabilidade de acontecer todo o espaço possível de valores é 1. E função cumulativa de probabilidades (fcp) P{x ≤ X} pode ser definida como:

$$P(X) = \sum_{x=a}^{X} p(x) \text{, se discreta, e } F(X) = \int_a^X f(x)dx \text{, se contínua.}$$

unidade de tempo) à qual o evento ocorre. Esta distribuição é fundamental para a teoria de filas.

Distribuição Exponencial Negativa

Se o número de chegadas a um centro de serviços durante um período específico ocorre de acordo com a distribuição Poisson, então, automaticamente, a distribuição dos intervalos entre chegadas sucessivas segue uma distribuição exponencial negativa (ou, simplesmente, exponencial). Especificamente, se λ é a taxa à qual o evento com distribuição Poisson ocorre, então a distribuição do tempo, x, entre chegadas sucessivas é dado por:

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0$$

A média e variância são:

$$E\{x\} = \frac{1}{\lambda}$$

$$var\{x\} = \frac{1}{\lambda^2}$$

A média E{x} é consistente com a definição de λ. Se a taxa à qual o evento ocorre, então $1/\lambda$ é o intervalo médio entre eventos sucessivos.

Distribuição Normal

A distribuição normal descreve muitos fenômenos aleatórios que ocorrem diariamente. A pdf de uma distribuição normal é definida como:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

onde,

$$E\{x\} = \mu$$

$$var\{x\} = \sigma^2$$

A notação N(μ, σ) é geralmente utilizada para representar uma distribuição normal com média μ e desvio padrão σ. A importância da distribuição normal se dá também pelo fato da média das amostras tiradas de quaisquer distribuições seguirem sempre distribuição normal (para mais informação sobre a distribuição normal, ver ED1).

Filas

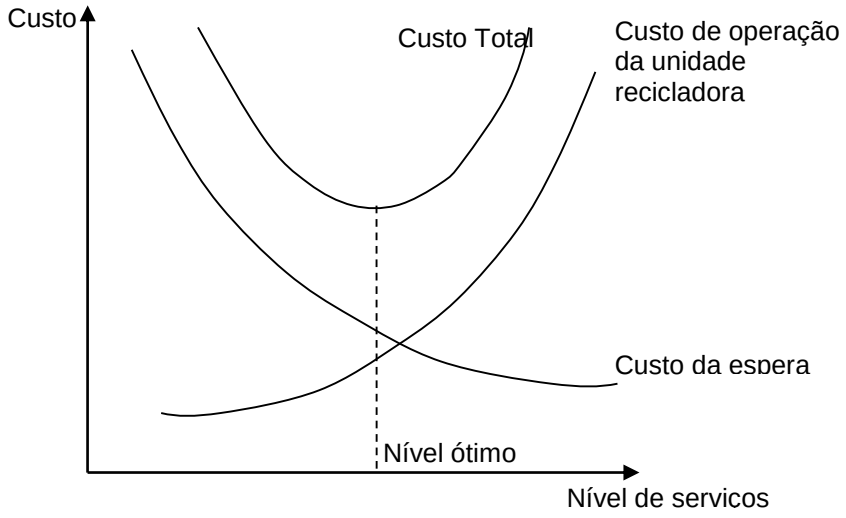
Por que estudar filas?
Dimensionar corretamente os sistemas para amenizar o impacto negativo de eventuais esperas.

O que se estuda?
Medidas de desempenho de uma determinada situação, como tempo médio de espera, comprimento da fila etc. Essas informações são depois usadas para definir o nível adequado no centro de serviços.

Exemplo:
Voluntários em um programa de reciclagem começaram a reclamar da lentidão do serviço. A unidade de reciclagem emprega atualmente três atendentes de recebimento de lixo reciclável. Um estudo foi contratado para investigar a reclamação. O estudo detectou a seguinte relação entre o número de atendentes e o tempo de espera por atendimento:

Atendentes	1	2	3	4	5	6	7
Espera média (min)	16,2	10,3	6,9	4,8	2,9	1,9	1,3

Examinando os resultados, e admitindo o desejo de se reduzir o tempo de espera em pelo menos 3 minutos, é fácil concluir que a unidade de reciclagem terá que aumentar o número de atendentes de 3 para 5 atendentes.
Os resultados também poderiam ser usados num contexto de análise de custos. Vide figura abaixo:



fundamentais
Os principais atores são o ***cliente*** e o ***servidor***. Clientes surgem de uma ***fonte***, que pode ser finita ou infinita. Na chegada ao ***centro de serviços***, podem ser atendidos imediatamente ou esperar numa ***fila*** se o centro de serviços estiver ocupado. Quando o centro completa um serviço, recebe imediatamente um cliente da fila. Se a fila estiver vazia, o centro permanece ocioso até a chegada de um novo cliente.
Do ponto de vista do analista de filas, a chegada de clientes é representada pelo ***tempo entre chegadas*** e o serviço é descrito pelo ***tempo de serviço*** por cliente, sendo que ambos podem ser probabilísticos ou determinísticos. O tamanho da fila pode se considerado de tamanho finito ou infinito.

A ***disciplina da fila***, que descreve a ordem como os clientes são selecionados da fila, é importante. São situações comuns: “primeiro a chegar, primeiro a servir (PCPS)”; “último a chegar, primeiro a servir (UCPS)”, “serviço em ordem ao acaso (SOA)”, “atendimento por prioridade”, “atendimento por caso” etc.
O desenho do centro de serviço pode prever a existência de um único servidor ou vários dispostos em paralelo. Variações nesses conceitos dão origem a diferentes modelos de fila, para os quais se definem técnicas de análise específicas.

O papel da distribuição exponencial
Em muitas situações, a chegada de clientes ocorre totalmente ao acaso, que neste caso significa que a ocorrência de um evento (chegada de um cliente ou término de um serviço) não é influenciada pelo intervalo de tempo decorrido desde a ocorrência do último evento. ***Tempo entre chegadas*** e ***tempo de serviço*** são quantitativamente descritos em modelos de filas pela distribuição exponencial:

$$f(t) = \lambda e^{-\lambda t}, \quad t > 0$$

onde E{t} = 1 / λ. O fato de que essa distribuição é totalmente ao acaso pode ser observado no seguinte caso: se agora fossem 8:20 e a última chegada tivesse ocorrido às 8:02, a probabilidade de que a próxima chegada ocorra até as 8:29 é função do intervalo entre 8:20 e 8:29 somente, e é totalmente independente do período de tempo decorrido desde a ocorrência do último evento (8:02 e 8:20). A este resultado nos referimos como falta de memória da distribuição exponencial.

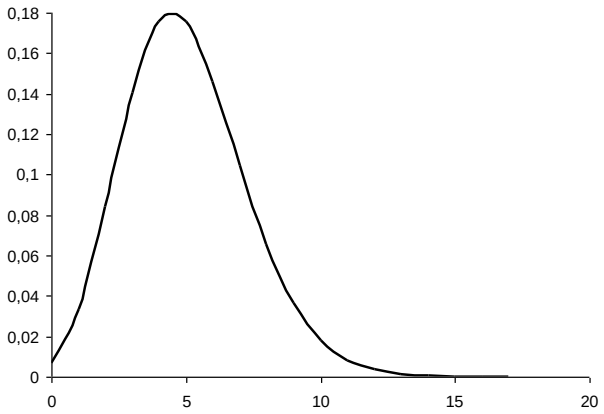
Relação entre a distribuição exponencial e Poisson
É possível provar que o número de chegadas durante certo período *t* de tempo seguirá a distribuição Poisson se o intervalo entre chegadas seguir a distribuição exponencial com média 1 / λ. Sendo o inverso também verdadeiro.

Exemplo: suponha uma taxa de nascimentos de um bebe a cada 12 minutos e que o intervalo entre nascimentos segue uma distribuição exponencial. Calcule:

a) Número médio de nascimentos por ano
Eventos médios por dia: $\lambda = (24x60)/12 = 120$ *nascimentos / dia*
Número de nascimentos por ano: $\lambda t = 120 \times 365 = 43.800$ *nascimentos / ano*

b) A probabilidade de que nenhum nascimento ocorrerá em qualquer dia
A probabilidade de nenhum nascimento ocorrer em um dia é calculado pela distribuição Poisson: $P_0(1) = (120 \times 1)^0 e^{-120 \times 1} / 0! \approx 0$

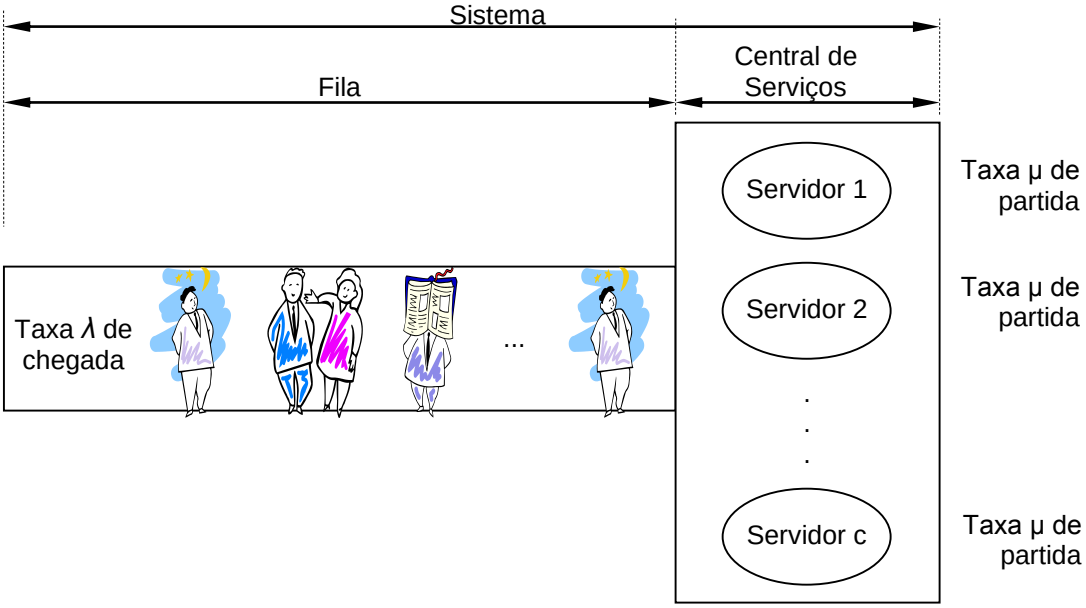
c) A probabilidade de emitir 50 certificados decorridas 3 horas, dado que foram emitidos 40 certificados nas últimas 2 horas
Procura-se, de fato, a probabilidade de ocorrerem 10 nascimentos na próxima hora. Sabemos que ocorre um nascimento a cada 12 minutos, ou seja, $\lambda = 60 / 12 = 5$ nascimentos por hora. Portanto, $P_{10}(1) = (5 \times 1)^{10} e^{-5 \times 1} / 10! = 0.01813$



Probabilidade de ocorrer *n* nascimentos em hora dado que *λ= 5*

Classificação dos modelos de filas

A literatura especializada classifica os modelos de filas como o representado na figura abaixo de acordo com uma conveniente notação.



(a / b / c) : (d / e / f)

Onde:

a: descreve a distribuição das chegadas[2]
b: descreve a distribuição das partidas (tempo de serviço)[2]
c: número de servidores em paralelo (1, 2, ..., ∞)
d: disciplina da fila (FCFS, LCFS, SIRO, GD)[3]
e: número (finito ou infinito) de clientes no sistema
f: tamanho da fonte de clientes (finita ou infinita)

---

[2] M: chegadas ou partidas markovianas (Poisson, ou equivalentemente tempo de serviço exponencial) D: determinístico; E_K: distribuição Erlang ou gamma de tempos; GI: distribuição genérica de tempo de serviços; G: distribuição geral de tempo de serviços.

[3] First Come, First Served; Last Come, First Served; Service In Random Order; General Discipline

Medidas de desempenho mais comuns
As mais comuns medidas de desempenho ao se analisar um modelo de filas são:
L_s: Número esperado de clientes no sistema
L_q: Número esperado de clientes na fila
W_s: Tempo médio de espera no sistema
W_q: Tempo médio de espera na fila
c: Número esperado de servidores ocupados

Exemplo:
Visitantes em uma unidade de conservação ocupam um estacionamento que possui apenas 5 vagas. Os automóveis que chegam nesse local seguem uma distribuição de Poisson com taxa de 6 carros por hora. O tempo de estacionamento é exponencialmente distribuído com média de 30 minutos. Visitantes que não encontram uma vaga podem temporariamente aguardar em fila numa área restrita. Nessa fila cabem apenas três carros. Os carros que não encontrarem espaço na fila devem procurar outra atração. Determine:

a) a probabilidade p_n de n carros ocuparem no sistema todo

b) a taxa efetiva de chegada de carros ao estacionamento

c) o número médio de carros no estacionamento

d) o tempo médio que um carro espera por um espaço no estacionamento

e) o número médio de vagas ocupadas no estacionamento

Cada vaga se comporta de fato como servidor e o estacionamento como um centro de serviços. No caso c = 5 servidores paralelos, e a capacidade máxima do sistema é de 8 carros (5 servidos mais 3 aguardando).

O modelo generalizado de filas assume que tanto a taxa de chegada como de partida dependem do estado do sistema, ou seja, dependem do número de clientes no centro de serviços. Assim precisamos definir:
n = número de clientes no sistema (na fila mais servidos)
λ_n = taxa de chegadas dado que n estão no sistema
μ_n = taxa de partidas dado que n estão no sistema

Sabe-se que p_n é função de λ_n e μ_n e que, sob condições estáveis (taxas esperadas de entrada e saída são iguais), obedece a seguinte forma:

$$p_n = \left(\frac{\lambda_{n-1}\lambda_{n-2}\ldots\lambda_0}{\mu_n\mu_{n-1}\ldots\mu_1} \right) p_0 \qquad n = 1,2,\ldots \qquad \text{e que} \qquad \sum_{n=0}^{\infty} p_n = 1$$

Para determinar a probabilidade p_n, precisamos primeiro determinar:

λ_n = 6 carros/hora, n = 0, 1, 2, ..., 8

$$\mu_n = \begin{cases} n\left(\dfrac{60}{30}\right) = 2n \text{ carros/hora}, \ \mathrm{n} = 1,2,\ldots,5 \\ 5\left(\dfrac{60}{30}\right) = 10 \text{ carros/hora}, \quad \mathrm{n} = 6,7,8 \end{cases}$$

E, portanto,

$$p_n = \begin{cases} \dfrac{\left(\dfrac{6}{2}\right)^n}{n!} p_0 \quad \mathrm{n} = 1,2,\ldots,5 \\ \dfrac{\left(\dfrac{6}{2}\right)^n}{5!5^{n-5}} p_0 \qquad \mathrm{n} = 6,7,8 \end{cases}$$

O valor de p_0 é computado substituindo-se p_n , n=1, 2, ... 8, na seguinte equação:

$$p_0 + p_1 + p_2 + p_3 + \ldots + p_8 = 1$$

ou

$$p_0 + p_0\left(\frac{3}{1!} + \frac{3^2}{2!} + \frac{3^3}{3!} + \frac{3^4}{4!} + \frac{3^5}{5!} + \frac{3^6}{5!5^1} + \frac{3^7}{5!5^2} + \frac{3^8}{5!5^3}\right) = 1$$

que resulta em p_0 = 0,04812, e nos seguintes valores de p_n

n	1	2	3	4	5	6	7	8
p_n	0,14436	0,21654	0,21654	0,16240	0,09744	0,05847	0,03508	0,02105

Para calcular a taxa efetiva de chegada ($\lambda_{efetiva}$) precisamos calcular primeiro a taxa de desistência (aqueles que encontram o sistema com 8 carros e não consegue entrar):

$\lambda_{desistentes} = \lambda\, p_8$ = 6 x 0,02105 = 0,1263 carros por hora

$\lambda_{efetiva} = \lambda - \lambda_{desistentes}$ = 6 - 0,1263 = 5,8737 carros por hora

O número L_s médio de carros no estacionamento (aqueles esperando ou ocupando um espaço) é calculado da seguinte forma:

$$L_s = 0\, p_0 + 1\, p_1 + 2\, p_2 + \ldots + 8\, p_8 = 3{,}1286$$

O tempo W_q de um carro esperando na fila temporária é calculado da seguinte forma:

$$W_q = W_s - 1/\mu$$

e sabendo que $W_s = L_s / \lambda_{efetiva}$ = 3,1286 / 5,8737 = 0,5265 horas, temos

$$W_q = 0{,}5265 - ½ = 0{,}03265 \text{ horas}$$

A média de vagas ocupadas equivale ao conceito de servidores ocupados, que é computado da seguinte forma: $c = L_s - L_q = \lambda_{efetiva} / \mu$ = 5,8737 / 2 = 2,9368

Mas, e se o sistema não é estável (entrada e saída irregulares), parâmetros variam com o tempo etc. → **S I M U L A Ç Ã O** (fica o convite para uma pós-graduação!!)