

UNIVERSIDADE FEDERAL DE JUIZ DE FORA  
DEPARTAMENTO DE ESTATÍSTICA  
CURSO DE ESTATÍSTICA

Daniel Morais de Souza

**Modelos ocultos de Markov: uma Abordagem em Controle de  
Processos**

JUIZ DE FORA

2013

Daniel Morais de Souza

MODELOS OCULTOS DE MARKOV: UMA ABORDAGEM EM CONTROLE DE  
PROCESSOS

Monografia apresentada ao curso de Estatística da Universidade Federal de Juiz de Fora, como requisito para a obtenção do grau de Bacharel em Estatística.

Orientador: Lupércio França Bessegato

Doutor em Estatística – UFMG

JUIZ DE FORA

2013

MODELOS OCULTOS DE MARKOV: UMA ABORDAGEM EM CONTROLE DE  
PROCESSOS

Monografia apresentada ao curso de  
Estatística da Universidade Federal de Juiz  
de Fora, como requisito para a obtenção do  
grau de Bacharel em Estatística.

Aprovada em \_\_\_\_ de \_\_\_\_\_ de 2013.

BANCA EXAMINADORA

---

Lupércio França Bessegato (orientador)

Doutor em Estatística – Universidade Federal de Minas Gerais

---

Camila Borelli Zeller

Doutora em Estatística – UNICAMP

---

Alfredo Chaoubah

Doutor em Engenharia Elétrica- Pontifícia Universidade Católica do Rio de Janeiro

Souza, Daniel Morais – Juiz de Fora, 2013

Modelos ocultos de Markov: uma Abordagem em Controle de Processos/  
Daniel Morais de Souza

72.p

Monografia – Universidade Federal de Juiz de Fora e Instituto de  
Ciências Exatas

Orientador: Lupércio França Bessegato

*À Deus por tudo.  
E à minha família.*

*“Quem já passou por essa vida e não viveu,  
pode saber mais,  
mas sabe menos do que eu”.*

*Vinicius de Moraes*

## AGRADECIMENTOS

Primeiramente queria agradecer à minha família por estar sempre ao meu lado, cada um de sua maneira única, sempre tornando momentos simples da vida, como um jogo de cartas, momentos muito especiais pra mim.

Agradeço ao meu pai por sempre lutar comigo e, conseqüentemente, por mim. Agradeço minha mãe e minha vó por sempre me protegerem. Agradeço minha tia Gina e meu vô por sempre me ensinar lindas lições. Agradeço meu irmão e minha cunhada Fa pelo companheirismo e por me ajudarem a concretizar este trabalho.

Devo agradecer muito, mais muito mesmo meus amigos Guilherme, Leandro e Thalita por tudo. Tenho certeza que sem vocês não estaria me formando hoje.

Não posso deixar de agradecer aos meus amigos de Ciências Exatas: Bárbara, Dadalti, Gi, Ju, Marcel, Vinícius e Yamashita. São amigos que levarei para o resto da vida.

Aos meus amigos de colégio: Banana, Eduardim, Fefe, Micoline, Mônica, Thiago e Vitão por me ajudarem em momentos difíceis e compartilhar momentos felizes.

Aos amigos Bebeto, Ni, Tuzim, Ronny e todos da banda, por sempre estarem ali na hora da tão esperada cerveja do final de semana.

Aos professores do departamento, por serem ótimos profissionais e grandes amigos. Em especial ao Joaquim e Clécio, hoje grandes amigos.

Ao meu orientador acadêmico e também orientador da vida, Lupércio, por toda a paciência e pelas horas de conversa.

Aos membros da banca, Camila e Alfredo, pelo tempo e dedicação.

À Deus, que mesmo eu esquecendo Ele algumas vezes, tenho certeza que Ele não fez o mesmo.

À todos vocês.

## RESUMO

A metodologia do modelo oculto de Markov está sendo amplamente utilizada em todos os campos do conhecimento nos dias de hoje. Com sua forte estrutura matemática, esta metodologia é capaz de modelar diversos problemas por ser bastante maleável e irrestrita, fato que não ocorre na metodologia de cadeia de Markov. Pode-se enxergar esta metodologia como uma extensão de cadeia de Markov, uma vez que ela pode conter mais de uma cadeia de Markov. Problemas de monitoramento de cartas de controle e controle *on-line* podem ser modelados via cadeia de Markov, visto que o intervalo de amostragem é regular e que a condição do processo ser Markoviano geralmente é assumida. Dorea *et al.* (2012) assumiram a existência de um sistema interno não observável de modo que o monitoramento por controle *on-line* é apenas uma parte do processo. Com base nesta suposição, pode-se modelar este processo via modelo oculto de Markov, já que o processo compreende duas cadeias de Markov (uma observável e outra não observável). Este trabalho apresentará a metodologia do modelo oculto de Markov para, em seguida, aplicar a metodologia no planejamento do controle estatístico *on-line* a fim de estabelecer uma estimativa para as probabilidades de alarme falso e a de não detecção.

**Palavras-chave:** Modelo Oculto de Markov, Controle estatístico *On-line*, Alarme Falso, Não Detecção.

## Sumário

1-	Introdução.....	12
2-	Controle estatístico de qualidade.....	14
2.1-	Introdução .....	14
2.2-	Cartas de controle $x$ .....	17
2.3-	Planejamento econômico de cartas de controle .....	23
2.4-	Controle <i>On</i> -line em cartas de controle.....	26
3-	Modelos Ocultos de Markov (HMM) .....	28
3.1-	Cadeias de Markov .....	28
3.1.1-	Processos Estocástico e Markoviano .....	28
3.1.2-	Principais Conceitos.....	29
3.1.3-	Irredutibilidade .....	31
3.1.4-	Periodicidade.....	32
3.1.5-	Distribuição Invariante .....	32
3.2-	Introdução do Modelo Oculto de Markov (HMM).....	33
3.3-	Elementos do HMM .....	35
3.4-	Os três problemas básicos do HMM .....	37
3.4.1-	Probabilidade de uma sequência de observações .....	38
3.4.2-	Melhor sequência de estados.....	44
3.4.3-	Estimação dos parâmetros.....	47
4-	HMM no controle estatístico <i>on-line</i> .....	50
4.1-	Introdução .....	50
4.2-	Alarme falso e Não detecção.....	53
5-	Análise de Sensibilidade.....	59
6-	Conclusão e recomendações futuras.....	69
	Referências Bibliográficas.....	70

## Lista de Figuras

Figura 1 - Ilustração de uma carta de controle .....	15
Figura 2 - Tempo até o sinal .....	23
Figura 3 - Modelo de urnas e bolas.....	35
Figura 4 - Ilustração da sequência de operações requeridas para a computação da variável Forward $\alpha_{(t+1)}(j)$ .....	41
Figura 5 - Implementação da computação de $\alpha_t(i)$ em termos das observações t e dos estados i .....	42
Figura 6 - Ilustração da sequência de operações requeridas para a computação da variável Backward $\beta_t(i)$ .....	43
Figura 7 - Ilustração da sequência de operações requeridas para a computação do evento conjunto do sistema estar no estado $S_i$ no tempo t e de estar no estado $S_j$ no tempo t+1.....	48
Figura 8 - Descrição do processo de inspeção .....	51

## Lista de Gráficos

Gráfico 1 - Sensibilidade da probabilidade de alarme falso/ largura dos limites de controle (k).....	60
Gráfico 2 – Sensibilidade da probabilidade de não detecção/ largura dos limites de controle (k).....	61
Gráfico 3 - Sensibilidade da probabilidade de alarme falso / intervalo entre amostras (m).....	62
Gráfico 4 - Sensibilidade da probabilidade de não detecção / intervalo entre amostras (m).....	63
Gráfico 5 - Sensibilidade da probabilidade de alarme falso / parâmetro p .....	64
Gráfico 6 - Sensibilidade da probabilidade de não detecção / parâmetro p .....	65
Gráfico 7 - Sensibilidade da probabilidade de alarme falso / parâmetro $\epsilon$ .....	66
Gráfico 8 - Sensibilidade da probabilidade de não detecção / parâmetro $\epsilon$ .....	67
Gráfico 9 - Sensibilidade da probabilidade de não detecção / parâmetro $\delta$ .....	68

## Lista de Tabelas

Tabela 1 - Valores de $d_2$ e $d_3$ .....	19
Tabela 2 – Valores das probabilidades para diferentes larguras de limites de controle	60
Tabela 3 - Valores das probabilidades para diferentes intervalos entre amostras .....	62
Tabela 4 - Valores das probabilidades para diferentes valores de parâmetro $p$ .....	63
Tabela 5 - Valores das probabilidades para diferentes valores de parâmetro $\varepsilon$ .....	65
Tabela 6 - Valores das probabilidades para diferentes valores do parâmetro $\delta$ .....	67

## Lista de Quadros

Quadro 1 - Teste de hipóteses do monitoramento .....	21
--	----

## 1- Introdução

O controle estatístico de qualidade tem sido amplamente utilizado por pesquisadores e indústrias com o intuito de otimizar os processos e garantir um nível de qualidade à sua produção. Uma de suas principais ferramentas é o controle de processos a partir de cartas de controle. Esta ferramenta monitora processos a fim de garantir qualidade à produção e avisar quando o processo não opera normalmente.

Há vários modos de planejar estas cartas de controle, dentre as quais podem-se destacar o planejamento econômico, que tem como objetivo minimizar os custos da produção. Taguchi *et al.* (1989) propuseram um modelo econômico para monitorar, em tempo real, uma determinada característica de qualidade que pode ser aplicado tanto em cartas de controle por variáveis (característica de qualidade é uma variável contínua) quanto em cartas de controle por atributos (característica de qualidade qualitativa, geralmente dicotômica, ex: conforme e não conforme). Este método é chamado de modelo de Taguchi para controle *on-line* de processos, e seu objetivo é determinar um intervalo ótimo de inspeção que visa minimizar o custo por item de um ciclo de produção.

De acordo com Dorea *et al.* (2012), é razoável assumir a existência de um sistema interno não observável de modo que o monitoramento por controle *on-line* é apenas uma parte do processo. Neste trabalho, sob a perspectiva dos modelos ocultos de Markov, e assumindo que a evolução do sistema é governada por uma cadeia de Markov de dois estágios, serão apresentadas as estimativas para o alarme falso e a não detecção do mau funcionamento do processo para controle estatístico *on-line*.

O modelo oculto de Markov (HMM) é utilizado para modelar processos que são governados por um processo Markoviano embutido, cujo sistema não pode ser diretamente observado. Este processo evolui pelo tempo por meio de transições entre seus estados, sendo que os estados são responsáveis pela emissão de sinais observáveis. Pode-se visualizar o HMM como uma extensão da cadeia de Markov, pois o modelo inclui o caso onde a observação é uma função de probabilidade.

O objetivo deste trabalho é apresentar a metodologia dos modelos ocultos de Markov e posteriormente fazer uma aplicação da mesma no controle estatístico *on-line* para definir algumas probabilidades importantes no planejamento econômico. Apresentadas a metodologia do modelo oculto de Markov e a aplicação no controle estatístico *on-line*, será feita uma simulação baseada em um exemplo retirado de Costa *et al.* (2005) para a comprovação das probabilidades oriundas da metodologia do HMM.

A metodologia dos modelos ocultos de Markov está sendo cada vez mais utilizada em diversas áreas, tais como saúde, biologia, computação, entre outras. Souza (2010) propôs esta metodologia para resolver o problema da proliferação do mosquito transmissor do vírus da dengue; já Khouri (2012) conseguiu fazer uma comparação e classificação de proteínas por meio de HMM; Rabiner (1989) utilizou esta metodologia no reconhecimento de fala.

Esta metodologia foi introduzida nos anos 60 e dois são os motivos que levam à sua aplicação: i) estes modelos têm uma estrutura matemática que pode ser aplicada em inúmeros problemas; e ii) quando aplicada corretamente, esta metodologia fornece ótimos resultados.

Este trabalho será apresentado da seguinte forma: além da introdução, o capítulo 2 apresenta algumas ferramentas do controle estatístico de qualidade como cartas de controle, planejamento econômico e controle estatístico *on-line*; no capítulo 3 será feita uma breve revisão sobre a metodologia de cadeia de Markov e algumas de suas propriedades, para, logo em seguida, ser apresentada a metodologia do modelo oculto de Markov; no capítulo 4 será feita uma aplicação da metodologia dos modelos ocultos de Markov no controle estatístico *on-line* e; o capítulo 5 apresenta as conclusões e principais resultados.

## **2- Controle estatístico de qualidade**

Neste capítulo serão apresentados os conceitos do controle estatístico de qualidade e suas principais ferramentas.

### **2.1- Introdução**

O controle estatístico de qualidade vem sendo cada vez mais utilizado dentro das indústrias, independente de suas áreas de atuação. Esta técnica estatística tem como principal objetivo fornecer uma otimização do processo de produção tal que reduza os custos e maximize os lucros respeitando certo nível de qualidade. Pode-se dividir esta técnica em duas etapas: a primeira parte é onde o pesquisador/empresa busca parâmetros estatísticos e econômicos que reduzam os custos e maximizem os lucros; e a segunda parte funciona como o monitoramento do processo, ou seja, o pesquisador/empresa monitora o processo para que os produtos sempre estejam respeitando o nível de qualidade especificado.

Há diversas formas de definir qualidade: Para Montgomery (2004) qualidade é inversamente proporcional à variabilidade, ou seja, quanto maior a variabilidade menor a qualidade; segundo Deming (2000) qualidade significa atender, e se possível, exceder as expectativas do consumidor; Taguchi *et al.* (1989) definiu a qualidade de um produto como as perdas econômicas impostas à sociedade a partir do momento em que ele é liberado para venda; Crosby (1995) afirma que qualidade significa atender às especificações.

O controle de qualidade desenvolve sistemas que garantem que os produtos sejam projetados e produzidos para atender ou superar as expectativas dos clientes. Em processos de produção que há um controle estatístico de qualidade, geralmente é delimitado limites de tolerância com o intuito de controlar a qualidade do produto.

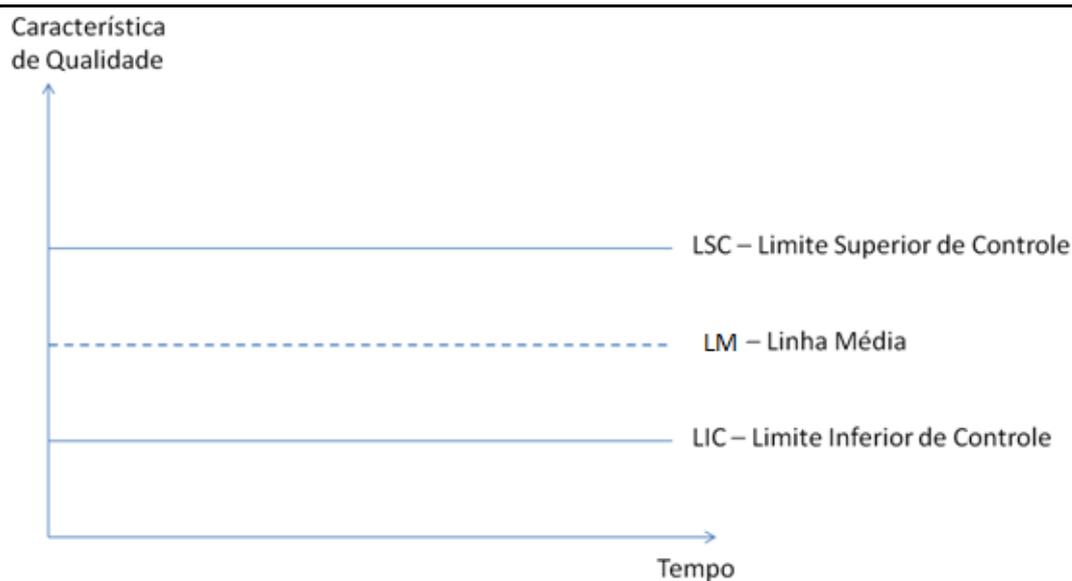
Uma das principais técnicas de controle estatístico de qualidade em processos de produção são as cartas de controle ou também, gráficos de controle. Desenvolvidas pelo Dr. Walter A. Shewhart (1920), as cartas de controle podem ser classificadas em dois grupos: cartas de controle por variáveis e cartas de controle por atributos. Segundo Costa *et al.* (2005) as cartas de controle por variáveis servem para o monitoramento de características de qualidade representadas por variáveis contínuas, e as cartas de controle por atributos servem para o monitoramento de processos qualitativos, em que os produtos podem ser classificados em defeituosos ou não

defeituosos. As cartas de controle por atributos também podem ser utilizadas para monitorar o número de não conformidades presentes em um produto, como por exemplo, em um chapa de metal.

As principais cartas de controle por variáveis são as cartas de controle  $\bar{x}$  e  $R$ . As cartas de controle  $\bar{x}$  são utilizadas para o controle da média de uma característica de qualidade, enquanto as cartas de controle  $R$  controlam a dispersão da característica de qualidade. Estas cartas de controle não são indicadas para mudanças pequenas e moderadas nos parâmetros que estão sendo controlados, para isso preferem-se as cartas de controle  $CUSUM$  e  $EWMA$ , que são as mais eficazes no monitoramento de processos sujeito a pequenas perturbações (MONTGOMERY, 2004; LEIRAS *et al.*, 2007).

As principais cartas de controle por atributos são as cartas de controle  $np$ ,  $p$ ,  $C$  e  $u$ . As cartas de controle  $np$  medem o número de produtos defeituosos que um processo produz (ou que um lote tem), as cartas de controle  $n$  medem a fração defeituosa do processo, as cartas de controle  $C$  medem o número de não-conformidades em uma amostra (ex: o número de defeitos que uma chapa de metal apresenta), e as cartas de controle  $u$  medem o número de não-conformidades por unidade de inspeção.

As cartas de controle contêm uma linha média (LM), representando o valor médio da característica de qualidade, e duas outras linhas horizontais, chamadas de limite superior de controle (LSC) e limite inferior de controle (LIC), representadas pela figura a seguir:



Fonte: Elaboração própria.

**Figura 1 - Ilustração de uma carta de controle**

O valor médio das cartas de controle é o valor alvo da estatística da característica de qualidade que se deseja para o produto, seja ela sua média, seu desvio-padrão, ou qualquer outra estatística. Os limites superior e inferior de controle formam um intervalo de aceitação para a qualidade do processo, ou seja, se a estatística da característica de qualidade dos produtos estiver dentro deste intervalo, considera-se o processo sob controle, caso contrário o processo é considerado fora de controle.

Escolhe-se a amplitude do intervalo LSC e LIC de maneira que, quando o processo estiver operando sob controle, praticamente todos os pontos amostrais estejam em seu interior. Geralmente utilizam-se os limites de controle com três desvios-padrão de afastamento em relação à linha média (“limites de 3 sigma”), que foram propostos por Shewhart baseados no seguinte lema: “se o processo estiver em controle, evite ajustes desnecessários, que só tendem a aumentar a variabilidade”(COSTA *et al.*, 2005).

O procedimento para o monitoramento dos processos proposto por Duncan (1956) para cartas de controle é o seguinte: A cada intervalo de tempo  $h$  é coletada uma amostra de produtos de tamanho  $n$  e é calculada a estatística da característica de controle desejada (ex: média, desvio-padrão, entre outros); se a estatística calculada estiver dentro dos limites de controle, o processo é considerado sob controle e ele continua operando normalmente; se a estatística calculada estiver fora dos limites de controle, o processo é considerado fora de controle e, portanto inicia-se uma investigação e uma ação corretiva para encontrar e eliminar a causa (ou causas) atribuível responsável por este comportamento.

Além das cartas de controle por variáveis e por atributos, Taguchi *et al.* (1989) propuseram um método que consiste em retirar apenas um item da linha de produção a cada intervalo fixo de  $m$  itens produzidos. Se o item for classificado não conforme, o processo é interrompido para a investigação e ajuste. Este método, chamado de controle *On-line*, pode ser aplicado tanto em cartas de controle por variáveis, quanto em cartas de controle por atributos, sendo que o problema é determinar o intervalo ótimo  $m$  e, no caso de controle de variáveis, o parâmetro dos limites de controle  $k$ . De acordo com Taguchi, este sistema deve ser empregado de um modo que os valores alvos da característica de qualidade possam ser economicamente planejados e controlados. Este assunto será abordado no final deste capítulo

Na próxima seção será discutida a metodologia de cartas de controle  $\bar{x}$  com o intuito de demonstrar como que as cartas de controle funcionam. Para a metodologia

das cartas de controle  $\bar{x}$  ser apresentada, é preciso ser introduzidos alguns conceitos das cartas de controle  $R$  necessários para a utilização de ambas as cartas, pois estas cartas são geralmente utilizadas em conjunto.

## 2.2- Cartas de controle $\bar{x}$

Se a característica de qualidade a ser monitorada é uma variável contínua, como por exemplo, o número de mililitros de uma lata de refrigerante, comumente a monitoração do processo é dada por um par de cartas de controle: um para o monitoramento da média da característica de qualidade e outro para a monitoração da variabilidade da mesma. Geralmente as cartas empregadas são o da média amostral  $\bar{x}$  para monitorar a média da característica e o da amplitude amostral  $R$  para monitorar a variabilidade.

A linha média  $LM_{\bar{x}}$  para as cartas de controle  $\bar{x}$  é localizada na média (valor esperado) de  $\bar{X}$ , e os limites de controle superior  $LSC_{\bar{x}}$ , e inferior  $LIC_{\bar{x}}$ , são geralmente estabelecidos a três desvios-padrão dessa média, dados por:

$$LSC_{\bar{x}} = \mu_{\bar{x}} + 3\sigma_{\bar{x}} \quad (01)$$

$$LIC_{\bar{x}} = \mu_{\bar{x}} - 3\sigma_{\bar{x}} \quad (02)$$

Pelo teorema central do limite pode-se afirmar que o valor esperado da estatística  $\bar{X}$  é igual ao valor esperado da característica de qualidade  $\mu_x$ , da variável aleatória  $X$ , isto é:

$$\mu_{\bar{x}} = \mu_x \quad (03)$$

Ainda supondo independência entre os valores, pelo teorema central do limite pode-se afirmar que o desvio-padrão amostral da característica de qualidade é dado por:

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} \quad (04)$$

sendo que  $\sigma_x$  representa o desvio-padrão da característica de qualidade monitorada  $X$  e  $n$  representa o tamanho amostral.

A variabilidade dos valores de  $\bar{X}$  diminui à medida que aumenta o tamanho das amostras. Por exemplo, para amostras de tamanho 9, o desvio-padrão amostral é igual a um terço do desvio-padrão da característica de qualidade  $X$ .

Porém na prática, estes parâmetros não são conhecidos, e quando conhecidos, não apresentam uma boa precisão. Por isso são utilizadas estimativas que, sem perda de representatividade, podem ser utilizadas nos cálculos das cartas de controle. Adotando

estas estimativas como  $\hat{\mu}_0 = \mu_{\bar{x}}$  e  $\hat{\sigma}_0 = \sigma_{\bar{x}}$ , temos os seguintes parâmetros para as cartas de controle  $\bar{x}$ :

$$LSC_{\bar{x}} = \hat{\mu}_0 + 3 \frac{\hat{\sigma}_0}{\sqrt{n}} \quad (05)$$

$$LM_{\bar{x}} = \hat{\mu}_0 \quad (06)$$

$$LIC_{\bar{x}} = \hat{\mu}_0 - 3 \frac{\hat{\sigma}_0}{\sqrt{n}} \quad (07)$$

Para o cálculo destas estimativas, é necessário antecipar a definição dos parâmetros das cartas de controle  $R$ , pois a estimativa do desvio-padrão amostral depende destes parâmetros. Portanto, os limites de controle para as cartas de controle  $R$  são dados por:

$$LSC_R = \mu_R + 3\sigma_R \quad (08)$$

$$LM_R = \mu_R \quad (09)$$

$$LIC_R = \mu_R - 3\sigma_R \quad (10)$$

sendo que  $LSC_R$ ,  $LM_R$  e  $LIC_R$  consecutivamente representam o limite superior de controle, linha média e o limite inferior de controle das cartas de controle  $R$ ; e que  $\mu_R$  e  $\sigma_R$  consecutivamente representam a média da amplitude amostral e o desvio-padrão da amplitude amostral.

O fator que multiplica  $\sigma_R$  é definido como 3 pela literatura, porém pode assumir qualquer valor dentre os reais.

Segundo Costa *et al.* (2005), se a característica de qualidade de interesse seguir uma distribuição normal com desvio-padrão  $\sigma$ , a distribuição da amplitude amostral  $R$  terá média e desvio-padrão dados por:

$$\mu_R = d_2\sigma \quad (11)$$

$$\sigma_R = d_3\sigma \quad (12)$$

onde  $d_2$  e  $d_3$  são constantes que dependem apenas do tamanho amostral  $n$  conforme a tabela a seguir:

**Tabela 1 - Valores de  $d_2$  e  $d_3$**

Tamanho da amostra <b>n</b>	Parâmetros da carta de controle R	
	<b>d2</b>	<b>d3</b>
2	1,128	0,853
3	1,693	0,888
4	2,059	0,880
5	2,326	0,864
6	2,534	0,848
7	2,704	0,833
8	2,847	0,820
9	2,970	0,808
10	3,078	0,797
11	3,173	0,787
12	3,258	0,778
13	3,336	0,77
14	3,407	0,763
15	3,472	0,756

Fonte: Adaptado de Costa *et al.* (2005).

Com base em (11) e (12), e substituindo  $\sigma$  por sua estimativa  $\hat{\sigma}_0$  (obtida com o processo sob controle), pode-se então definir os limites de controle e a linha média para as cartas de controle R como:

$$LSC_R = d_2\hat{\sigma}_0 + 3d_3\hat{\sigma}_0 \quad (13)$$

$$LM_R = d_2\hat{\sigma}_0 \quad (14)$$

$$LIC_R = d_2\hat{\sigma}_0 - 3d_3\hat{\sigma}_0 \quad (15)$$

Porém deve-se fazer um lembrete: Como a amplitude é uma medida definida em  $\mathbb{R}^+$ , ela não pode assumir valores negativos, portanto, quando (15) for calculado e resultar em um número negativo, imediatamente adota-se  $LIC_R = 0$ , significando que não existe um limite inferior de controle.

A literatura revela que as estimativas da média, do desvio-padrão e da amplitude devem ser baseadas em pelo menos 20 amostras. Dado um conjunto de  $m$  amostras, a estimativa da média da característica de qualidade  $\mu_X$  geralmente é o valor médio das médias amostrais, ou seja, suponha que  $\bar{X}_1, \dots, \bar{X}_m$  sejam as médias amostrais das  $m$  amostras coletadas, o melhor estimador é dado por:

$$\mu_X = \bar{\bar{X}} = \frac{\sum_{i=1}^m \bar{X}_i}{m} \quad (16)$$

Como  $\mu_X = \bar{X}$ , pode-se afirmar que  $\mu_{\bar{X}} = \bar{X}$ , ou seja,  $\bar{X}$  será usada como a linha média das cartas de controle  $\bar{x}$ .

A estimativa do desvio-padrão da característica de qualidade  $\sigma_X$  é calculada pelas amplitudes das  $m$  amostras. Suponha que  $R_1, \dots, R_m$  sejam as amplitudes das  $m$  amostras. O estimador da amplitude da característica de qualidade será dado por:

$$\mu_R = \bar{R} = \frac{\sum_{i=1}^m R_i}{m} \quad (17)$$

e conseqüentemente o estimador do desvio-padrão da característica será dado por:

$$\hat{\sigma}_0 = \frac{\bar{R}}{d_2} \quad (18)$$

como  $\bar{R}$  é um estimador não viciado para a amplitude, pode-se afirmar que  $\bar{R}/d_2$  é um estimador não viciado para o desvio-padrão.

Note que as cartas de controle  $\bar{x}$  dependem da média e do desvio-padrão da característica de qualidade. Já as cartas de controle  $R$  são independentes da média da característica de qualidade, portanto, não conseguem identificar alterações na medida central do processo.

Finalmente, podem-se definir as estimativas para os parâmetros das cartas de controle  $\bar{x}$  como:

$$LSC_{\bar{X}} = \bar{X} + 3 \frac{\bar{R}}{d_2 \sqrt{n}} \quad (19)$$

$$LM_{\bar{X}} = \bar{X} \quad (20)$$

$$LIC_{\bar{X}} = \bar{X} - 3 \frac{\bar{R}}{d_2 \sqrt{n}} \quad (21)$$

Este intervalo de  $\pm 3 \bar{R}/d_2 \sqrt{n}$  engloba grande parte dos valores de  $\bar{X}$ , para ser mais preciso 99,73% deles, ou seja, se o processo estiver sob controle, raramente um ponto caíra fora da região de aceitação da carta de controle.

Como dito anteriormente, o monitoramento do processo por meio das cartas de controle  $\bar{x}$  é feito retirando amostras de tamanho  $n$  a cada  $h$  horas. Retirado estas amostras, é calculada para cada amostra a média amostral, e registrada graficamente em uma carta de controle como mostrada na figura (1). É possível perceber que este procedimento é constituído por sucessíveis testes de hipóteses, os quais podem ser definidos como:

$$H_0: \text{Processo está sob controle } (\mu = \mu_0)$$

Vs

$$H_1: \text{Processo está fora de controle } (\mu \neq \mu_0)$$

considere que quando o processo estiver sob controle, a média do processo  $\mu$  assumirá  $\mu_0$ , e quando estiver fora de controle, a média do processo  $\mu$  será diferente de  $\mu_0$ .

Não rejeita  $H_0$  se o valor da estatística  $\bar{x}$  estiver dentro dos limites de controle. Analogamente, rejeita  $H_0$  se o valor da estatística  $\bar{x}$  estiver fora dos limites de controle.

Como toda técnica estatística, este procedimento é passível de erros. Estes erros são chamados de erros de declaração, pois o erro ocorre quando a o pesquisador declara que o processo está sob controle ou fora de controle quando na verdade não está. Como todo teste de hipótese, este procedimento pode ter dois erros:

i) Erro do tipo I: Ocorre quando é declarado que o processo está fora de controle, quando na verdade o processo está sob controle, ou seja, a estatística  $\bar{x}$  calculada está fora do intervalo de aceitação, porém o processo está sob controle. Na prática, este erro resulta em intervir o processo na hora errada, quando ele está isento de causas especiais que modificam a média da característica de qualidade. Este erro é comumente chamado de alarme falso. A probabilidade do alarme falso é representada por  $\alpha$ , isto é:

$$\alpha = P(\bar{X} > LSC_{\bar{x}} | \mu = \mu_0) + P(\bar{X} < LIC_{\bar{x}} | \mu = \mu_0) \quad (22)$$

ii) Erro do tipo II: Ocorre quando é declarado que o processo está sob controle, quando na verdade o processo está fora de controle, ou seja, a estatística  $\bar{x}$  calculada está dentro do intervalo de aceitação, porém o processo está fora de controle. Na prática, este erro resulta em não intervir o processo na hora certa, quando a média da característica de qualidade do processo foi afetada por alguma causa especial. A probabilidade deste erro ocorrer é representada por  $\beta$ , isto é:

$$\beta = P(LIC_{\bar{x}} \leq \bar{X} \leq LSC_{\bar{x}} | \mu \neq \mu_0) \quad (23)$$

O quadro a seguir apresenta os possíveis resultados de um teste de hipóteses do monitoramento por cartas de controle  $\bar{x}$ :

**Quadro 1 - Teste de hipóteses do monitoramento**

Hipótese $H_0$	Decisão			
	Aceitar $H_0$	Probabilidade	Rejeitar $H_0$	Probabilidade
Verdadeira	Decisão correta	$1 - \alpha$	<b>Erro tipo I</b>	$\alpha$
Falsa	<b>Erro tipo II</b>	$\beta$	Decisão correta	$1 - \beta$

Fonte: Costa *et al.* (2005)

Um conceito importante nas cartas de controle é o poder  $Pd$ , dado por:

$$Pd = 1 - \beta \quad (24)$$

ou seja, o poder é a probabilidade do processo ser declarado corretamente fora de controle. Muitos pesquisadores utilizam o poder para efeito de comparação de cartas de controle, pois quanto mais rápido o processo detectar alguma causa especial que modificou a média da característica de qualidade, melhor é a carta de controle.

A partir do poder, pode-se determinar o número médio de amostras entre a ocorrência da mudança na média do processo e a detecção e declaração de uma causa especial, dado por:

$$NMA = \frac{1}{Pd} \quad (25)$$

Para a determinação do número de amostras até o primeiro sinal de que o processo desajustou, utiliza-se o número médio de amostras antes de um alarme falso. Pode-se afirmar que este número segue uma distribuição geométrica de parâmetro  $p = \alpha$ . A média da distribuição geométrica é  $1/p$ , logo, o número médio de amostras é dado por:

$$NMAF = \frac{1}{\alpha} \quad (26)$$

Estes dois conceitos são amplamente utilizados para a comparação de modelos por terem uma interpretação prática e fácil.

Outra medida de comparação de cartas de controle é o intervalo de tempo entre a alteração da média do processo e a declaração do processo estar fora de controle. Segundo Costa *et al.* (2005) este intervalo é chamado de tempo esperado até o sinal ( $TES$ ), ou seja, tempo esperado até a declaração do processo estar fora controle.

Este conceito depende do intervalo de tempo entre amostras  $h$ , do número de amostras até o sinal e do intervalo de tempo entre a última amostragem com o processo sob controle e o instante que a média se desloca. Portanto, pode-se definir o tempo esperado até o sinal como:

$$\begin{aligned} TES &= E(TS) \\ &= E(h \times NAS - Q) \\ &= h \times E(NAS) - E(Q) \end{aligned} \quad (27)$$

onde  $TS$  é o tempo até o sinal,  $h$  é o intervalo entre amostras,  $NAS$  é o número de amostras até o sinal e  $Q$  é o intervalo de tempo entre a última amostragem com o processo sob controle e o instante que a média se desloca.

Pode-se definir que o valor esperado do número de amostras até o sinal corresponde ao número médio de amostras entre a ocorrência da mudança na média do processo e a detecção e declaração de uma causa especial, ou seja:

$$E(NAS) = NMA = \frac{1}{Pd} \quad (28)$$

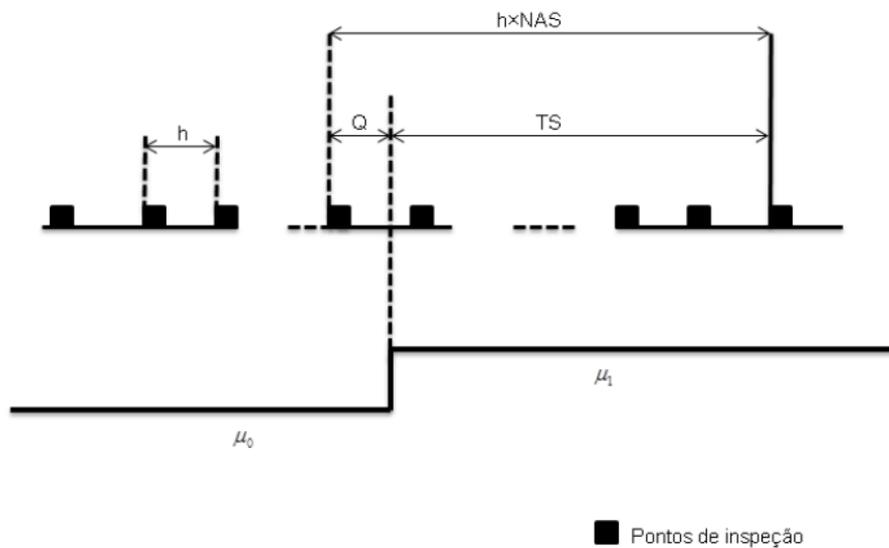
Como o deslocamento da média do processo pode ocorrer em qualquer tempo entre um intervalo entre amostras  $h$ , pode-se supor que o tempo que ocorre este descolamento  $Q$  segue uma distribuição uniforme, e, portanto, o seu valor esperado é dado por:

$$E(Q) = \frac{h}{2} \quad (29)$$

logo, chega-se a seguinte expressão:

$$\begin{aligned} TES &= h \times NMA - \frac{h}{2} \\ &= \frac{h}{Pd} - \frac{h}{2} \end{aligned} \quad (30)$$

A figura a seguir mostra como o  $TES$  é medido:



Fonte: Costa *et al.* (2005).

**Figura 2 - Tempo até o sinal**

### 2.3- Planejamento econômico de cartas de controle

Como dito anteriormente, as cartas de controle são amplamente utilizadas para o monitoramento da qualidade de processos, tanto na área industrial quando nos setores de serviço. Para implementar as cartas de controle, devem-se escolher

parâmetros estatísticos que minimizem os custos do processo, e para isto, deve-se fazer um planejamento.

O planejamento das cartas de controle pode ser feito utilizando quatro critérios diferentes:

- O heurístico de Shewhart
- O econômico-estatístico
- O semi-econômico
- O econômico

O planejamento de cartas de controle heurístico de Shewhart é pouco utilizado nos dias de hoje, pois ele não leva em consideração nenhum parâmetro de custo.

O planejamento econômico-estatístico foi proposto por Saniga (1989) com o intuito de criar um planejamento econômico levando em consideração a probabilidade dos erros de classificação, ou seja, os erros do tipo I e II. Este planejamento adota algumas restrições estatísticas no modelo econômico para as empresas obterem processos com uma menor variabilidade e conseqüentemente, produtos com maior qualidade. Portanto, este planejamento tem a vantagem de produzir produtos com maior qualidade, porém gera um custo esperado maior do que os planejamentos semi-econômico e econômico.

O planejamento semi-econômico considera algumas grandezas relacionadas aos custos, como por exemplo o custo de amostragem, a frequência de alarmes falsos e o tempo de detecção do processo fora de controle; porém este tipo de planejamento não considera todos os custos relevantes e com isso não aplica técnicas otimização a uma função de custo. Os principais trabalhos relacionados ao planejamento semi-econômico são de Girshick e Rubin (1952) e Weiler (1952).

O planejamento econômico de cartas de controle busca minimizar uma função que representa todos os custos relevantes do processo de produção. Segundo Woodall (1986), este procedimento apresenta algumas fragilidades, tais como:

- Maior incidência de alarmes falsos. Isto se deve ao fato do planejamento gerar riscos muito grandes ao produtor, ou seja,  $\alpha$  muito grande.
- Não detecta rapidamente os desvios padrões esperados do processo, causado pelo fato de que na maioria das vezes o risco do produtor também é grande, ou seja,  $\beta$  muito grande.
- Não é eficaz na detecção de pequenos desvios não esperados pelo processo. Este problema ocorre quando as cartas de controle utilizadas no planejamento econômico

são de  $\bar{x}$ , pois elas não são sensíveis a pequenas mudanças. Podem-se utilizar outras cartas de controle que captam essas pequenas mudanças, como por exemplo, as cartas *CUSUM*.

Em geral, os modelos econômicos buscam selecionar parâmetros de tal modo que maximize o lucro total esperado do processo de produção (WOODALL, 1987). O principal trabalho sobre planejamento econômico de cartas de controle foi apresentado por Duncan (1956), no qual foi proposto um planejamento econômico de cartas de controle  $\bar{x}$ . Neste planejamento, Duncan considerou as seguintes suposições para o modelo econômico de cartas de controle:

- i) O processo é caracterizado por ter somente um estado sob controle.
- ii) As causas especiais que o processo pode sofrer, resultando em uma mudança do valor desejado, ocorrem de acordo com um Processo de Poisson.
- iii) A transição entre os estados é instantânea.
- iv) O processo não é autocorretivo, ou seja, só com intervenção humana que o processo volta ao estado sob controle.

O objetivo é encontrar três parâmetros que minimizem o custo esperado do processo, que são:

- O tamanho de amostra  $n$ ;
- O intervalo entre amostras  $h$ ;
- Largura dos limites de controle (que nas seções anteriores, foi considerada igual a três)  $k$ .

A média do processo sob controle é dada por  $\mu_0$ , e que uma única causa especial de magnitude  $\delta$ , provoca uma mudança da média do processo de  $\mu_0$  para  $\mu_1 = \mu_0 \pm \delta\sigma$ . No modelo apresentado, considera-se que os parâmetros  $\mu_0$ ,  $\delta$  e  $\sigma$  são conhecidos, e que os parâmetros  $n$ ,  $h$  e  $k$  devem ser decididos.

Após o trabalho de Duncan, vários autores fizeram pesquisas na área de planejamento econômico em cartas de controle. Ladany (1973) desenvolveu o planejamento econômico para cartas de controle  $p$ , e logo após, vieram vários trabalhos sobre o planejamento econômico para as cartas de controle de Shewhart (incluindo cartas de controle  $\bar{x}$  e  $R$ ). Gibra (1975) apresentou uma lista de referências de trabalhos sobre técnicas de cartas de controle. Montgomery (1980) fez uma revisão da literatura do planejamento econômico em cartas de controle. Vance (1983) apresenta uma bibliografia das técnicas de controle estatístico em cartas de controle.

Lorenzen e Vance (1986) desenvolveram um método de planejamento econômico unificado que englobasse todas as cartas de controle. Este modelo busca encontrar o tamanho amostral, o intervalo amostral e os limites de controle que minimize uma função de custo calculada sob algumas suposições (Como o de Duncan (1956)).

Este tipo de planejamento também pode ser utilizado em controle estatístico *on-line*. Taguchi (1981), Taguchi (1985) e Taguchi *et al.* (1989) propuseram técnicas de controle estatístico *on-line* e como aplicá-lo no planejamento econômico de processos. Na próxima seção será abordada a metodologia do controle estatístico *on-line* e os principais trabalhos relacionados a este assunto.

## **2.4- Controle *On-line* em cartas de controle**

Controle *on-line* de processos está sendo cada vez mais utilizado nos últimos anos. Vários estudiosos e pesquisadores vêm utilizando este método pois ele apresenta resultados interessantes no campo quantitativo, principalmente na área de produção.

Este procedimento pode ser utilizado em cartas de controle por variáveis e cartas de controle por atributo.

O controle *on-line* por atributos primeiramente proposto por Taguchi *e al.* (1989) consiste em retirar e analisar apenas um único item da linha de produção a cada intervalo de  $m$  itens produzidos. Esse método é conhecido como modelo de Taguchi para controle *on-line* de processos, e seu objetivo é determinar o intervalo ótimo de inspeção, minimizando o custo por item em um ciclo de produção.

Muitos autores desenvolveram modelos de controle *on-line* com determinadas particularidades para atender as necessidades de cada área: Nayebpour e Woodall (1993) consideram um mecanismo de falha para o sistema, isto é, a mudança do processo sob controle para o processo fora de controle por meio das amostras segue uma distribuição geométrica; Borges *et al.* (2001) propuseram um método que considera presença de erros de classificação, ou seja, considera que pode haver erro no julgamento do estado de um item; Bessegato (2009), Bessegato (2011) e Bessegato (2012) incorporou algumas características no modelo como intervalo de inspeção variável, amostra não unitária e decisões não dicotômicas; Quinino e Ho (2004) e Ho e Trindade (2007) consideram o uso de classificações repetidas visando minimizar os custos.

Ho e Trindade (2007) garantem que esta metodologia pode ser utilizada em sistemas de produção com controles automáticos como soldagem, produção de diodos para placas de circuitos elétricos, produção de semicondutores e processos químicos em geral.

O controle *on-line* por variáveis é um método proposto por Taguchi *et al.* (1985, 1989) baseado na inspeção de um único item a cada  $m$  itens produzidos com o intuito de minimizar os custos médios por meio de mudanças nos parâmetros do processo. O item selecionado é inspecionado e classificado dentro da região de aceitação ou não, sendo que se o item for classificado fora da região de aceitação, o processo é interrompido e corrigido, caso contrário o processo continua operando normalmente.

No final deste trabalho, será feita uma aplicação dos modelos ocultos de Markov em controle *on-line* de cartas de controle.

No próximo capítulo será revisada toda a metodologia de cadeias de Markov e consequente, será apresentada a metodologia dos modelos ocultos de Markov.

### 3- Modelos Ocultos de Markov (HMM)

Neste capítulo será feita uma breve revisão do método de cadeias de Markov para, em seguida, apresentar a metodologia dos modelos ocultos de Markov.

#### 3.1- Cadeias de Markov

Antes de iniciar as definições do método de cadeias de Markov, será introduzido o conceito dos processos estocástico e Markoviano para melhor entendimento.

##### 3.1.1- Processos Estocástico e Markoviano

Um processo estocástico  $\{X(t), t \in T\}$  é uma sequência de variáveis aleatórias indexadas por algum elemento  $T$ , que geralmente é o tempo; ou seja, para cada  $t \in T$ ,  $X(t)$  é uma variável aleatória.

O conjunto  $T$  é chamado de conjunto de índices de um processo estocástico, e se este for contável, é dito que o processo estocástico é um processo de tempo discreto (Este trabalho será voltado apenas para processos de tempos discretos). O espaço de estados  $E$  de um processo estocástico é definido como o conjunto de todos os valores possíveis que as variáveis aleatórias  $X(t)$  podem assumir (ROSS, 2007).

Para fixar ideias, considere o seguinte exemplo: Imagine que nos dias do mês de janeiro, a probabilidade de um dia estar nublado é 0,2; de estar chuvoso é 0,3; e de estar ensolarado é 0,5. Este exemplo pode ser considerado como um processo estocástico de tempo discreto, onde  $X_t$  é o clima do dia  $t$  no mês de janeiro. Neste caso, o conjunto de índices seria  $T = \{1, 2, 3, \dots\}$ , admitindo que 32 seria o dia primeiro de janeiro no ano seguinte e assim consecutivamente, e o espaço de estados seria  $E = \{\text{nublado}, \text{chuvoso}, \text{ensolarado}\}$ .

Portanto, um processo estocástico nada mais é que uma coleção de variáveis aleatórias que descrevem o comportamento de algum processo com o passar do tempo. Uma vez definido o que é um processo estocástico, é de fácil compreensão o conceito de um processo Markoviano.

Um processo Markoviano é um processo estocástico que possui a propriedade de Markov, ou seja, possui a seguinte propriedade (ATUNCAR, 2011):

$$P(X_{t+1} \in A | X_1 = x_1, X_2 = x_2, \dots, X_t = x_t) = P(X_{t+1} \in A | X_t = x_t) \quad (31)$$

Esta propriedade, que também pode ser chamada de memória Markoviana, possui a peculiaridade de serem irrelevantes os estados anteriores para a predição do estado seguinte, desde que o estado atual seja conhecido.

Agora considere o exemplo anterior reformulado: Sabe-se que no mês de janeiro, considerando sempre o clima de determinado dia, há três probabilidades de ocorrências de clima no dia seguinte: i) dado que determinado dia é nublado, a probabilidade de o dia seguinte ser nublado é de 0,4, de ser chuvoso é de 0,3, e de ser ensolarado é de 0,3; ii) dado que determinado dia é chuvoso, a probabilidade de o dia seguinte ser nublado é de 0,3, de ser chuvoso é de 0,6, e de ser ensolarado é de 0,1; e iii) dado que determinado dia é ensolarado, a probabilidade de o dia seguinte ser nublado é de 0,2, de ser chuvoso é de 0,2, e de ser ensolarado é de 0,6. Este é um clássico exemplo de processo Markoviano, já que o clima de um dia qualquer depende apenas do clima do dia anterior. O conjunto de índices e o espaço de estados continua o mesmo do exemplo antecedente a esse.

A partir destes conceitos, será introduzido na próxima seção o conceito de cadeia de Markov, e conseguinte algumas de suas ferramentas e conceitos que possibilitarão o seu uso no campo do controle estatístico *on-line* para o controle estatístico de qualidade por meio de modelos ocultos de Markov. Esta seção será voltada para cadeias de Markov em que o parâmetro do tempo é discreto e o espaço de estados é finito e discreto.

### **3.1.2- Principais Conceitos**

Cadeia de Markov é um processo estocástico com espaço de estados discretos ou contínuos que apresenta a propriedade Markoviana, ou seja, em intervalos de tempos regulares e discretos, este processo estocástico evolui de um estado para outro dependentemente apenas do seu último estado, e conseqüentemente independentemente dos demais.

Considere um problema no qual em qualquer tempo o processo pode estar em um estado de  $N$  estados distintos,  $S_1, S_2, \dots, S_N$ . Em espaços de tempos regulares e discretos, este processo evolui de um estado para outro (ou possivelmente permanece no mesmo estado) de acordo com uma matriz de probabilidades associada com cada estado. Será denotado os instantes de tempo associados a cada mudança como  $t = 1, 2, \dots$ , e será denotado  $q_t$  como o atual estado do processo. Será dito que o processo  $q = \{q_t: t = 0, 1, \dots\}$  é uma cadeia de Markov se para todo  $t$  (RABINER, 1989):

$$P(q_t = S_j | q_{t-1} = S_i, q_{t-2} = S_k, \dots) = P(q_t = S_j | q_{t-1} = S_i) \quad (32)$$

Sem formalidades, pode ser visto que (32) impõe que a distribuição condicional de  $q_t$  dado o passado do processo dependa apenas do estado imediatamente anterior. Portanto, pode-se definir a probabilidade de transição  $a_{ij}$  do estado  $i$  para o estado  $j$  no instante de tempo  $t$  como:

$$a_{ij} = P(q_t = S_j | q_{t-1} = S_i) \quad t = 0, 1, 2, \dots \quad (33)$$

Estas probabilidades de transição podem ser organizadas em uma matriz quadrada de ordem  $N$  que comumente é chamada de Matriz de probabilidades de transição  $\{a_{ij}\}$ , ou seja:

$$\{a_{ij}\} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \dots & a_{NN} \end{pmatrix} \quad (34)$$

Pode-se observar que para cada estado  $i$ ,  $\{a_{ij}, j = 1, 2, \dots, N\}$  define-se uma função de distribuição de probabilidades. Com base nesta afirmação é possível determinar duas propriedades para estas probabilidades de transição:

$$a_{ij} \geq 0 \quad (35)$$

$$\sum_{j=1}^N a_{ij} = 1 \quad (36)$$

Uma matriz  $\{a_{ij}\}$  que respeita (35) e (36) é chamada de Matriz Estocástica.

Agora será definido um vetor de probabilidades que determinará a distribuição de probabilidades iniciais de uma cadeia de Markov, que será chamada de Distribuição inicial:

$$\pi_i = P[q_1 = S_i] \quad 1 \leq i \leq N \quad (37)$$

É possível modelar o último exemplo do clima dos dias de Janeiro por cadeia de Markov. Veja que o processo é Markoviano com  $N = 3$  (Nublado, chuvoso, ensolarado) e que as probabilidades de transição respeitam as duas propriedades anteriores. O clima será denotado para a melhor compreensão do exemplo como:

- Estado 1: Nublado

- Estado 2: Chuvoso

- Estado 3: Ensolarado

A partir das probabilidades de transição determinadas, e da notação definida anteriormente para o clima do dia, a matriz de probabilidades de transição é dada por:

$$\{a_{ij}\} = \begin{bmatrix} 0,4 & 0,3 & 0,3 \\ 0,3 & 0,6 & 0,1 \\ 0,2 & 0,2 & 0,6 \end{bmatrix}$$

Formulado este modelo, é possível responder a várias perguntas como, por exemplo: Dado que o clima do dia 4 de Janeiro foi ensolarado, qual é a probabilidade de nos próximos cinco dias o clima ser “ensolarado, chuvoso, nublado, nublado e chuvoso”?

Definindo formalmente o problema, a sequência de observações  $O$  pode ser determinada como  $O = \{S_3, S_3, S_2, S_1, S_1, S_3\}$  sendo que o tempo corresponde é  $T = 4, 5, 6, 7, 8, 9$ . A probabilidade desta sequência dado o modelo pode ser expressa por:

$$\begin{aligned} P(O|Modelo) &= P[S_3, S_3, S_2, S_1, S_1, S_3|Modelo] \\ &= P[S_3] \cdot P[S_3|S_3] \cdot P[S_2|S_3] \cdot P[S_1|S_2] \cdot P[S_1|S_1] \cdot P[S_3|S_1] \\ &= \pi_3 \cdot a_{33} \cdot a_{23} \cdot a_{12} \cdot a_{11} \cdot a_{31} \\ &= 1 \cdot (0,6) \cdot (0,2) \cdot (0,3) \cdot (0,4) \cdot (0,3) \\ &= 4,32 \cdot 10^{-3} \end{aligned}$$

Agora serão apresentadas algumas definições de cadeia de Markov necessárias para o estudo dos modelos ocultos de Markov no controle estatístico *on-line*. Primeiro serão apresentadas as definições de irredutibilidade e periodicidade para conseguinte apresentar o conceito da distribuição invariante.

### 3.1.3- Irredutibilidade

É dito que o estado  $j$  é acessível desde o estado  $i$  se existe uma trajetória do estado  $i$  ao estado  $j$ , ou seja, se existe um número inteiro não negativo finito  $n_{ij}$  tal que  $\{a_{ij}\}^{(n_{ij})} > 0$ . Se  $j$  é acessível desde  $i$  e  $i$  é acessível desde  $j$ , é dito que  $i$  e  $j$  comunicam. Uma cadeia é irredutível quando todos os estados comunicam entre eles (ATUNCAR 2011).

Em problemas de controle estatístico de qualidade, como por exemplo, inspeção de itens de uma linha de produção, a literatura revela que geralmente quando os pesquisadores modelam estes problemas por meio de cadeias de Markov essas cadeias

se mostram irredutíveis, o que os possibilitam aplicar a técnica da distribuição invariante para os cálculos desejáveis.

### 3.1.4- Periodicidade

É definido como período de um estado  $i$ , representado por  $d(i)$ , o máximo divisor comum de todos os inteiros  $t$  tais que  $\{a_{ii}\}^{(t)} > 0$ , ou seja:

$$d(i) = m. d. c\{t: \{a_{ii}\}^{(t)} > 0\} \quad (38)$$

Por exemplo, o processo começando no estado  $i$ , é possível que o processo apenas possa retornar ao estado  $i$  nos tempos  $2, 4, 6, 8, \dots$ , neste caso o estado  $i$  tem período 2. Um estado com o período 1 é chamado de aperiódico (ROSS, 2007).

É natural que os modelos probabilísticos por cadeia de Markov no controle estatístico *on-line* tenham cadeias de Markov aperiódicas, uma vez que todos estados se comunicam (o que implica em uma periodicidade igual a 1 de toda a cadeia).

### 3.1.5- Distribuição Invariante

Como neste presente trabalho não é tratado nenhuma aplicação de Modelos Ocultos de Markov com espaço de estados infinitos, não foi necessário tratar de recorrência, uma vez que esta definição é importante apenas no caso em que o modelo tenha espaço de estados infinitos.

Se uma cadeia de Markov for finita, irredutível e aperiódica, pode-se afirmar que existe a distribuição invariante e esta tem a seguinte forma:

$$\lim_{t \rightarrow \infty} \{a_{ij}\}^t = \bar{\pi} = \begin{pmatrix} \bar{\pi}_1 & \bar{\pi}_2 & \dots & \bar{\pi}_N \\ \bar{\pi}_1 & \bar{\pi}_2 & \dots & \bar{\pi}_N \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\pi}_1 & \bar{\pi}_2 & \dots & \bar{\pi}_N \end{pmatrix} \quad (39)$$

Outra forma de calcular a distribuição invariante é dada a seguir:

$$(\{a_{ij}\}' - I)\bar{\pi} = \mathbf{0} \quad (40)$$

em que  $\{a_{ij}\}'$  é a matriz de transição de probabilidades transposta de tamanho  $N \times N$ ,  $I$  é a matriz identidade de tamanho  $N \times N$ ,  $\mathbf{0}$  é o vetor nulo de tamanho  $N \times 1$  e  $\bar{\pi}$  é a distribuição invariante de tamanho  $N \times 1$ .

Se  $\bar{\pi} = (\bar{\pi}_1, \bar{\pi}_2, \dots, \bar{\pi}_N)$  é a distribuição invariante de uma cadeia de Markov,  $\bar{\pi}_i$  representa a probabilidade de que a cadeia se encontra no estado  $i$  após ela entrar em

equilíbrio. Outra interpretação importante para a distribuição invariante no controle estatístico de qualidade é que ela representa a proporção de tempo em que o sistema de produção visita cada estado após um número suficientemente grande de inspeções.

Nesta seção foi feita uma breve revisão dos principais conceitos de cadeia de Markov para introduzir os modelos ocultos de Markov, que pode ser visto como uma extensão de cadeia de Markov. Nas próximas seções serão discutidos: introdução à metodologia dos modelos ocultos de Markov, alguns exemplos e aplicações, elementos do modelo e os três problemas básicos e suas soluções.

### **3.2- Introdução do Modelo Oculto de Markov (HMM)**

Em cadeia de Markov é considerado que cada estado corresponde a uma observação. A literatura mostra que este modelo é muito restrito para ser aplicado em muitos problemas. O modelo oculto de Markov (HMM) vem como uma extensão da cadeia de Markov, pois agora o modelo inclui o caso no qual a observação é uma função de probabilidade do estado, isto é, o modelo resultante é um processo estocástico de dupla camada, na qual um processo estocástico é subjacente e não observável (oculto) que pode ser apenas observado pelo outro processo estocástico que produz as sequências de observações (RABINER 1989).

Os HMM's formam uma classe de processos estocásticos que estão sendo aplicados com sucesso em uma ampla variedade de problemas (COUVREUR 1996). Por exemplo, em problemas de reconhecimento de fala, os modelos ocultos de Markov conseguem caracterizar os sinais do mundo real em sinais do modelo; em dependências no DNA, segundo Silva (2003), estes modelos conseguem capturar a heterogeneidade que pode estar presente em diferentes regiões da mesma sequência de DNA; segundo Souza (2010) este modelo pode ser aplicado na classificação de comportamentos animais; em controle estatístico de qualidade no monitoramento de cartas de controle e no controle *on-line*, que é o principal tema deste trabalho, entre outros.

Serão mostrados dois exemplos didáticos com o intuito de exemplificar o HMM:

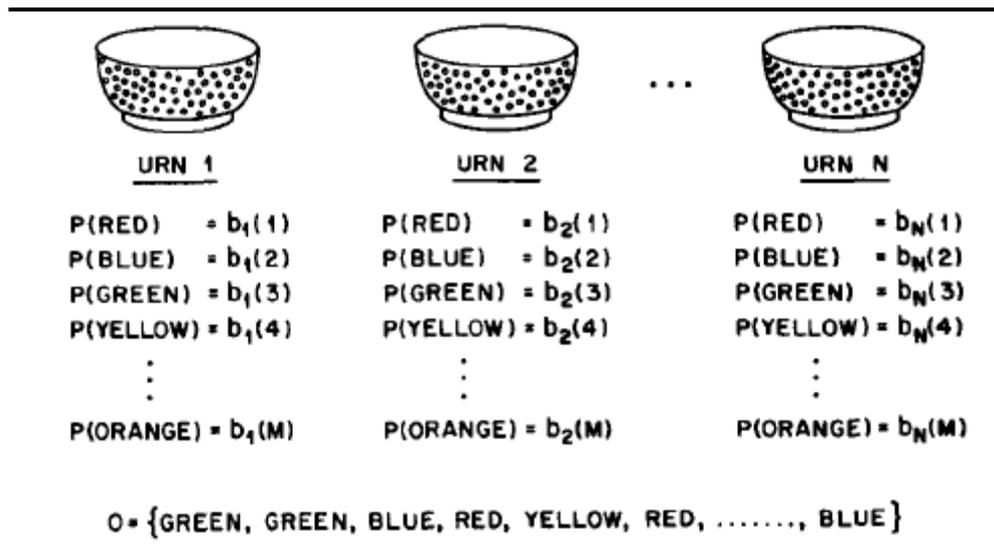
i) Suponha que há dois grupos mistos de pessoas, compostos por pessoas que conseguem assobiar e pessoas que não conseguem assobiar. Contudo, todas as pessoas conseguem bater palmas. No primeiro grupo, há três pessoas que conseguem assobiar e uma que não consegue e, no segundo grupo, duas conseguem assobiar e duas não conseguem. São gravadas algumas sequências de assobios e palmas, sendo que cada

pessoa assobia ou bate palma e, em seguida, escolhe aleatoriamente outra pessoa (ou a si próprio) para fazer o mesmo. Se uma sequência qualquer de assobios e palmas for coletada, é de interesse saber qual a probabilidade de uma observação ter sido gerada pelo primeiro grupo ou pelo segundo grupo. Este processo pode ser facilmente modelado por HMM, sendo que o grupo a qual é escolhida a pessoa é o processo não observável e a sequência de assobios e palmas (observações) é o processo observável.

O termo “observações” pode ter diferentes interpretações quando utilizado em cadeia de Markov ou em modelo oculto de Markov; no primeiro, significa que são os estados percorridos pelo modelo; enquanto que no segundo são os símbolos gerados nos estados do modelo (SOUZA, 2010).

ii) Considere que em uma sala há  $N$  urnas com diferentes números de bolas em cada uma delas. Assume-se que há  $M$  cores de bolas distintas. O processo é dado pela seguinte maneira: Primeiro a pessoa escolhe uma urna inicial, e desta urna uma bola é escolhida ao acaso; observa-se a cor da bola e conseqüentemente essa cor se torna a observação do processo. A bola é recolocada em sua urna e então uma nova urna é escolhida aleatoriamente de acordo com alguma distribuição de probabilidades associada com a atual urna; uma bola é escolhida aleatoriamente, observa-se a cor e a recoloca em sua urna. O processo é repetido muitas vezes e, portanto gera uma sequência observações constituída pelas cores das bolas.

Este processo pode ser modelado por HMM, já que são nítidas as duas camadas de um processo estocástico: o processo observável, que representa a cor da bola; e o processo não observável, que representa a escolha da urna. Como a sequência das cores não revela a sequência de urnas pode-se considerar que o processo das urnas é não observável (RABINER 1989). A figura a seguir ilustra este exemplo:



Fonte: Rabiner (1989).

Figura 3 - Modelo de urnas e bolas

### 3.3- Elementos do HMM

Os dois exemplos apresentados anteriormente mostram basicamente o que é um HMM e como ele pode ser aplicado em diferentes e simples cenários. Agora serão apresentados formalmente os elementos que constituem um HMM e como estes modelos geram as sequências de observações. Sempre que apresentado algum elemento do HMM será remetido ao exemplo ii da seção 3.3 para facilitar a compreensão. Será preservada a notação utilizada no primeiro tópico deste capítulo.

Um HMM é caracterizado pelos seguintes elementos:

- i)  $N$  representa o número de estados do modelo. Embora os estados sejam ocultos, em muitos problemas práticos há um significado importante para os estados ou para uma sequência de estados. No modelo das urnas e bolas, os estados corresponderiam às urnas.
- ii)  $S = \{S_1, S_2, \dots, S_N\}$  corresponde ao conjunto dos estados individuais do modelo. No modelo das urnas e bolas seria cada urna, como por exemplo,  $S_2$  seria a segunda urna da sala.
- iii)  $q_t$  corresponde ao estado no tempo  $t$ . No modelo das urnas e bolas, por exemplo,  $q_3 = S_1$  significa que a terceira urna escolhida foi a primeira urna.
- iv)  $M$  representa o número de observações distintas por estado. No modelo das urnas e bolas, o número de observações distintas corresponderia ao número de bolas distintas.

v)  $\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_M\}$  corresponde ao conjunto de observações individuais. No modelo das urnas e bolas seria o conjunto de bolas de uma urna.

vi)  $\mathbf{A} = \{\mathbf{a}_{ij}\}$  corresponde a distribuição de probabilidades de transição dos estados, dada por:

$$a_{ij} = P(q_{t+1} = S_j | q_t = S_i) \quad 1 \leq i, j \leq N \quad (41)$$

Para os problemas relacionados ao controle estatístico de qualidade, essas probabilidades seguem um caso especial onde qualquer estado pode ser alcançado por outro estado com um único passo, ou seja,  $a_{ij} > 0$  para todo  $i$  e  $j$ ; para outros casos de HMM's alguns estados não podem ser alcançados ou é necessário mais de um passo para serem alcançados. No modelo das urnas e bolas, supondo que a escolha das urnas tenha a mesma probabilidade, independente do estado anterior, essa distribuição de probabilidades respeitaria um distribuição uniforme discreta, resultando em uma probabilidade igual para as urnas:  $a_{ij} = 1/N$  para todo  $i$  e  $j$ .

vii)  $\mathbf{B} = \{\mathbf{b}_j(\mathbf{k})\}$  corresponde a distribuição de probabilidade da observação no estado, dada por:

$$b_j(k) = P(v_k \text{ em } t | q_t = S_j) \quad 1 \leq j \leq N \quad (42)$$

$$1 \leq k \leq M.$$

No modelo das urnas e bolas, supondo que a urna escolhida foi a urna dois na primeira escolha tenha duas bolas brancas, uma preta e duas vermelhas, a distribuição de probabilidades das observações neste estado são:

$$b_2(\text{branca}) = P(v_{\text{branca}} \text{ em } t = 1 | q_1 = S_2) = 2/5$$

$$b_2(\text{preta}) = P(v_{\text{preta}} \text{ em } t = 1 | q_1 = S_2) = 1/5$$

$$b_2(\text{vermelha}) = P(v_{\text{vermelha}} \text{ em } t = 1 | q_1 = S_2) = 2/5$$

viii)  $\boldsymbol{\pi} = \{\boldsymbol{\pi}_i\}$  corresponde a distribuição inicial dos estados, dada por:

$$\pi_i = P(q_1 = S_i) \quad 1 \leq i \leq N \quad (43)$$

No modelo das urnas e bolas, supondo que a probabilidade de escolher inicialmente uma urna fosse igual, a distribuição inicial dos estados respeitaria uma distribuição uniforme discreta, resultando em uma probabilidade igual para as urnas:  $\pi_i = 1/N$  para todo  $i$  e  $j$ .

Dados os valores apropriados dos elementos do HMM, este pode ser usado para gerar uma sequência de observações  $O = O_1 O_2 \dots O_T$ , onde cada observação  $O_t$  é um

dos símbolos de  $V$ , e  $T$  é o número de observações da sequência. Para gerar essa sequência de observações basta seguir os seguintes passos:

- 1) Escolha um estado inicial  $q_1 = S_i$  de acordo com a distribuição inicial dos estados  $\pi$ .
- 2) Compute  $t = 1$ .
- 3) Escolha  $O_t = v_k$  de acordo com a distribuição de probabilidade de observação do estado  $S_i$ , isto é,  $b_i(k)$ .
- 4) Transite para um novo estado  $q_{t+1} = S_j$  de acordo com a distribuição de probabilidade de transição do estado para o estado  $S_j$ , isto é,  $a_{ij}$ .
- 5) Compute  $t = t + 1$ ; retorne para o passo 3) se  $t < T$ ; caso contrário termine o procedimento.

O procedimento pode ser usado tanto para gerar uma sequência de observações quanto para mostrar como que uma dada sequência de observações foi gerada pelo apropriado HMM (RABINER 1989).

Claramente pode-se constatar que um HMM pode ser completamente definido pela distribuição de probabilidades de transição dos estados  $A$ , pela distribuição de probabilidade da observação no estado  $B$ , e pela distribuição inicial dos estados  $\pi$ ; logo é definida uma notação compacta para representar um HMM, dada por:

$$\lambda = (A, B, \pi) \tag{44}$$

Nesta seção foram apresentados os elementos que constituem um HMM. Na próxima seção estes elementos serão utilizados para resolver os três problemas básicos dos HMM.

### 3.4- Os três problemas básicos do HMM

Na última seção foram apresentados os elementos do HMM. Porém definidos estes elementos, aparecem três problemas que devem ser solucionados para que este tipo de modelo seja útil em aplicações do mundo real. Os três problemas do HMM são (RABINER 1989):

- Problema 1: Dado uma sequência de observações  $O = O_1 O_2 \dots O_T$  e o HMM  $\lambda = (A, B, \pi)$ , como de maneira eficiente se pode calcular a probabilidade da sequência dado o modelo  $P(O|\lambda)$ ?

- Problema 2: Dado uma sequência de observações  $O = O_1 O_2 \dots O_T$  e o HMM  $\lambda = (A, B, \pi)$ , como escolher a sequência de estados que maximize a sequência de observações?

- Problema 3: Como ajustar os parâmetros de um HMM com o intuito de maximizar a probabilidade da sequência  $P(O|\lambda)$ ?

O problema 1 se trata de definir a probabilidade de uma sequência dado o HMM, ou também, este problema pode ser interpretado como quão bem um modelo se ajusta a uma sequência de observações. Esta última interpretação é extremamente útil, como por exemplo, no reconhecimento de fala. Quando se utiliza HMM em reconhecimento de fala, ao se produzir um fonema qualquer, essa entrada pode ser classificada como pertencente ao modelo que indicar a maior probabilidade  $P(\lambda|O)$ , ou seja, um método eficaz para fazer comparação de modelos (SOUZA 2010).

O problema 2 se trata de definir a sequência de estados que melhor explica a sequência de observações, ou seja, a sequência de estados cuja probabilidade seja a maior dentre todas as sequências de estados possíveis do modelo. Informalmente, este problema busca desvendar a parte oculta do modelo, isto é, achar a sequência de estados “correta”. Porém deve ficar claro que este problema não determina a sequência de estados correta, apenas a sequência que tem maior probabilidade de corresponder a uma determinada sequência de observações.

O problema 3 se trata da otimização dos parâmetros do modelo, ou seja, a configuração de  $\lambda$ . As sequências de observações utilizadas para ajustar o modelo são chamadas de sequências de treinamento, pois que são usadas para treinar o HMM. Esta etapa é de suma importância, pois ela permite adaptar os parâmetros do modelo, criando os melhores modelos para o mundo real.

Nos próximos tópicos serão apresentadas as soluções destes problemas.

### **3.4.1- Probabilidade de uma sequência de observações**

O objetivo desta seção é calcular a probabilidade de uma sequência de observações  $O = O_1 O_2 \dots O_T$  dado o HMM  $\lambda$ , isto é,  $P(O|\lambda)$ . O modo mais trivial e direto para calcular essa probabilidade é através da verificação de todas as sequências de estados possíveis de tamanho  $T$  que o modelo pode gerar, e, conseguinte, efetuar o cálculo das sequências de observações terem sido geradas pelos respectivos estados dessas sequências.

Para efetuar o cálculo por este modo trivial, considere uma sequência de estados fixa  $Q$ :

$$Q = q_1 q_2 \dots q_T$$

Onde  $q_1$  e  $q_T$  representam os estados iniciais e finais da sequência de estados. Dado esta sequência de estados, é possível definir a probabilidade de uma sequência de observações  $O$  por:

$$P(O|Q, \lambda) = \prod_{t=1}^T P(O_t|q_t, \lambda) \quad (45)$$

Assumindo independência entre as observações, tem-se que:

$$P(O|Q, \lambda) = \prod_{t=1}^T b_{q_t}(O_t) \quad (46)$$

A probabilidade de uma sequência de estados pode ser escrita por:

$$P(Q|\lambda) = \pi_{q_1} \prod_{t=1}^{T-1} a_{q_t q_{t+1}} \quad (47)$$

Pelo lema da probabilidade condicional, a probabilidade conjunta de  $O$  e  $Q$  é dada por:

$$P(O, Q|\lambda) = P(O|Q, \lambda) \times P(Q|\lambda) \quad (48)$$

Por fim, pelo lema da probabilidade total, a probabilidade de uma sequência de observações  $O$  ter sido gerada pelo modelo é a soma de (48) para todas as sequências de estados possíveis em  $\lambda$ , dada por:

$$P(O, Q|\lambda) = \sum_Q P(O|Q, \lambda) \times P(Q|\lambda) \quad (49)$$

Que também pode ser escrita como:

$$P(O, Q|\lambda) = \sum_{q_1, q_2, \dots, q_T} \pi_{q_1} b_{q_1}(O_1) a_{q_1 q_2} b_{q_2}(O_2) \dots a_{q_{T-1} q_T} b_{q_T}(O_T) \quad (50)$$

Por esta fórmula, pode-se fazer a seguinte interpretação: i) Inicializa-se o processo ( $t = 1$ ) escolhendo o primeiro estado  $q_1$  com probabilidade  $\pi_{q_1}$ ; ii) gera-se uma observação  $O_1$  com probabilidade  $b_{q_1}(O_1)$ ; iii) o tempo muda para  $t = t + 1$  e o processo transita para o estado  $q_2$  com probabilidade  $a_{q_1 q_2}$ ; iv) gera-se uma observação  $O_2$  com probabilidade  $b_{q_2}(O_2)$ . Este processo continua até acontecer a última transição do estado  $q_{T-1}$  para  $q_T$  e gerar a última observação  $O_T$  com probabilidade  $b_{q_T}(O_T)$ .

O problema para efetuar este cálculo é que o número de possíveis sequências de estados de tamanho  $T$  utilizando os  $N$  estados é  $N^T$ . E se for reescrita apenas o termo dentro do somatório de (50), é obtido:

$$\pi_{q_1} \times \prod_{i=1}^{T-1} a_{q_i q_{i+1}} \times \prod_{i=1}^{T-1} b_{q_i}(O_i) \times b_{q_T}(O_T) \quad (51)$$

Pode-se visualizar que em cada produto realiza-se  $T - 2$  multiplicações, logo, para efetuar o cálculo de (51), deve-se realizar  $1 + (T - 2) + 1 + (T - 2) + 1 = 2T + 1$

multiplicações para cada sequência de estados. Dessa forma, para efetuar o cálculo da probabilidade de uma sequência de observações a partir das equações anteriores, é preciso realizar  $N^T(2T - 1)$  multiplicações e  $N^T - 1$  adições, o que torna o processo impraticável computacionalmente para diversos casos (SOUZA 2010).

Para a solução deste problema ser possível, é preciso um procedimento mais eficiente e computacionalmente possível. Baum e Egon (1967) e Baum e Shel (1968) propuseram dois procedimentos independentes e capazes de resolver este problema de uma maneira eficiente e computacionalmente praticável, chamados de *Forward Backward*.

Os procedimentos *Forward* e *Backward* utilizam técnicas de programação dinâmica, e com isso requer um número de multiplicações e adições consideravelmente menor do que o procedimento anterior para efetuar o cálculo da probabilidade da sequência de uma observação dado o HMM.

Considere a variável *Forward* definida por:

$$\alpha_t(i) = P(O_1 O_2 \dots O_t, q_t = S_i | \lambda) \quad (52)$$

Isto é, a variável *Forward* é a probabilidade conjunta da sequência de observações parcial  $O_1 O_2 \dots O_t$  e do estado ser  $S_i$  no tempo  $t$ , dado o HMM. Indutivamente pode-se usar a variável *Forward* para efetuar o cálculo de  $P(O|\lambda)$  como descrito abaixo:

1) Inicialização:

$$P(O_1, q_1 = S_i | \lambda) = \alpha_1(i) = \pi_i b_i(O_1) \quad 1 \leq i \leq N \quad (53)$$

ou seja, a variável *Forward* em  $t = 1$  é a probabilidade conjunta de observar a observação  $O_1$  e o sistema estar no estado  $S_i$  dado o modelo.

2) Indução:

$$\alpha_{t+1}(j) = \left[ \sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(O_{t+1}) \quad 1 \leq t \leq T - 1 \quad (54)$$

$$1 \leq j \leq N$$

Será mostrado o segundo passo deste procedimento com um intuito didático.

$$P(O_1 O_2, q_2 = S_j | \lambda) = \alpha_2(j) = \left[ \sum_{i=1}^N \pi_i b_i(O_1) a_{ij} \right] b_j(O_2) = \left[ \sum_{i=1}^N \alpha_1(i) a_{ij} \right] b_j(O_2) \quad (55)$$

A variável *Forward* em  $t = 2$  é a probabilidade conjunta de observar a sequência de observações  $O_1 O_2$  e o sistema estar no estado  $S_j$  dado o modelo. O produto  $\alpha_1(i) a_{ij}$  pode ser interpretado como  $P(O_1, q_1 = S_i, q_2 = S_j | \lambda)$ ; somando este último produto por todos os estados que  $S_i$  pode assumir em  $t = 1$ , ou seja,  $\sum_{i=1}^N \alpha_1(i) a_{ij}$ , obtêm-se  $P(O_1, q_2 = S_j | \lambda)$ ; e finalmente, se multiplicar este último

termo pela probabilidade da segunda observação ser  $O_2$  dado  $q_2 = S_j$ , ou seja,  $b_j(O_2)$ , resulta em  $\alpha_2(j)$ . Esta etapa do cálculo é repetido até  $t = T - 1$ .

3) Terminação:

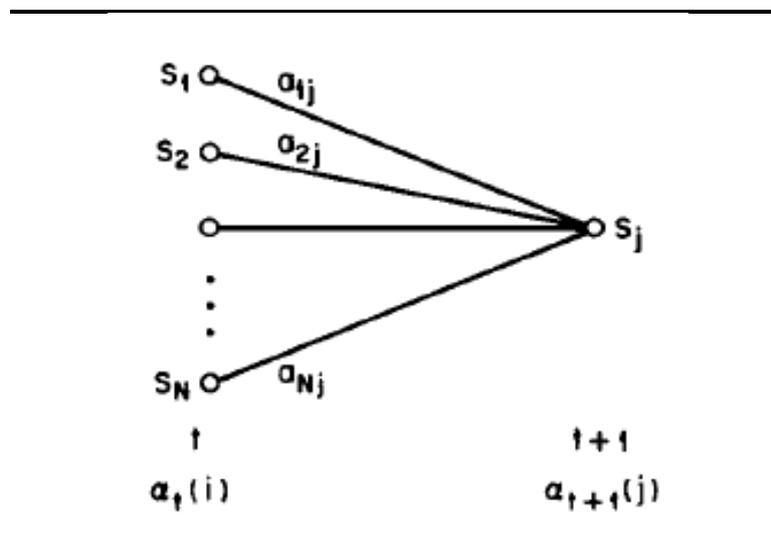
$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i) \quad (56)$$

Pela definição, pode-se afirmar que:

$$\alpha_T(i) = P(O_1 O_2 \dots O_T, q_T = S_i | \lambda) \quad (57)$$

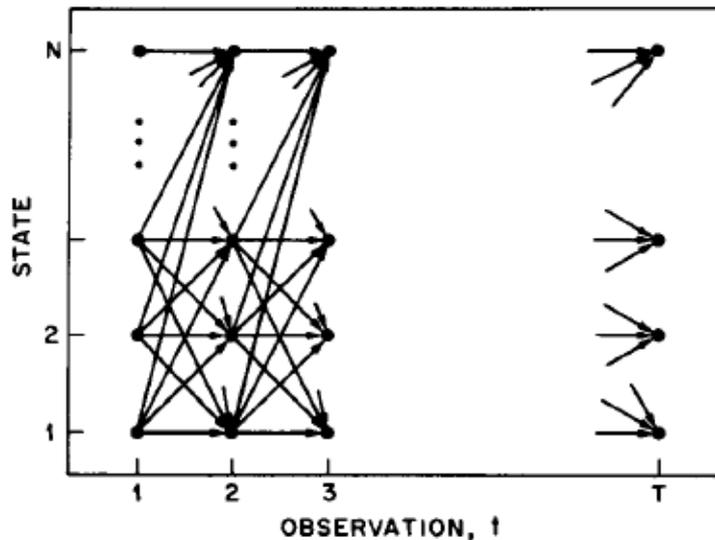
Pelo teorema da probabilidade total, basta somar todos os  $\alpha_T(i)$ 's para se obter a probabilidade desejada, ou seja,  $P(O|\lambda)$ .

Inicializa-se o processo definindo a variável *Forward* como a probabilidade conjunta do estado  $S_i$  e a observação inicial  $O_1$ . O segundo passo, que é a indução, é ilustrada nas figuras a seguir:



Fonte: Rabiner (1989).

**Figura 4 - Ilustração da sequência de operações requeridas para a computação da variável Forward  $\alpha_{(t+1)}(j)$**



Fonte: Rabiner (1989).

**Figura 5 - Implementação da computação de  $\alpha_t(i)$  em termos das observações  $t$  e dos estados  $i$**

Esta figura mostra como o estado  $S_j$  pode ser alcançado no tempo  $t + 1$  pelos  $N$  estados possíveis  $S_i$ , no tempo  $t$ .

O número de cálculos computacionais utilizados para realizar este procedimento é apenas  $N^2T$ , enquanto que no procedimento direto seria  $2TN^T$ .

De modo análogo e independente, o procedimento *Backward* também fornece a probabilidade de uma sequência de observações dado o modelo. Este procedimento também utiliza técnicas de programação dinâmica, porém vale ressaltar que ele é independente do procedimento *Forward* e que apenas um dos procedimentos é necessário para o cálculo da probabilidade desejada.

Considere a variável *Backward* definida por:

1) Inicialização:

$$\beta_T(i) = 1 \quad 1 \leq i \leq N \quad (58)$$

Este procedimento começa pela última observação da sequência, e assim sendo, a variável *Backward* assume um no tempo  $t = T$ , pois considera todas as sequências de observações possíveis.

2) Indução:

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j) \quad t = T - 1, T - 2, \dots, 1 \quad (59)$$

$$1 \leq i \leq N$$

Como este procedimento funciona “de trás para frente”, será considerado agora o segundo tempo para procedimento, que para o processo seria  $t = T - 1$ :

$$\beta_{T-1}(i) = \sum_{j=1}^N a_{ij} b_j(O_T) \beta_T(j) = \sum_{j=1}^N a_{ij} b_j(O_T) \quad (60)$$

Como  $\beta_T(j)$  é igual a um, chega-se a  $\sum_{j=1}^N a_{ij} b_j(O_T)$ . Esta expressão mostra que a variável *Backward* capta todas as mudanças dos possíveis  $N$  estados que  $S_j$  pode assumir fixando o estado  $S_i$  por meio das probabilidades de transição  $a_{ij}$  e das probabilidades das observações nos estados  $b_j(O_{t+1})$ .

3) Terminação:

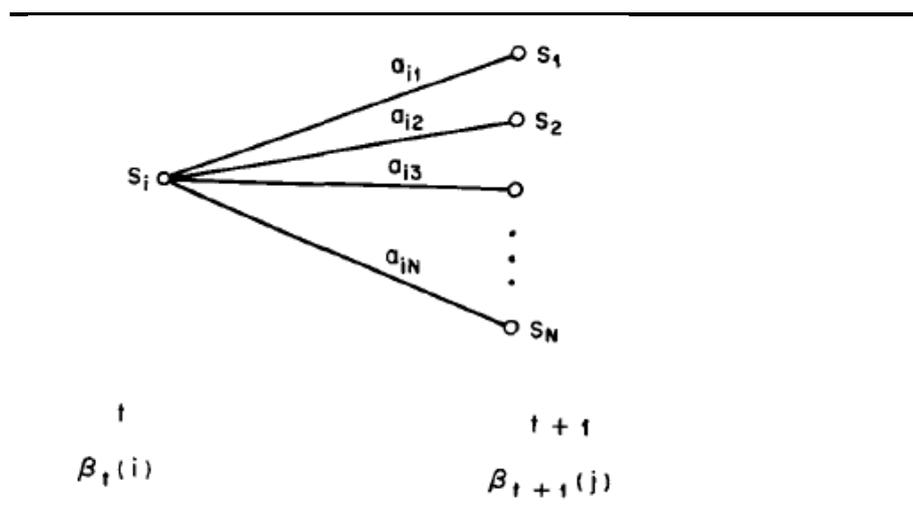
$$P(O|\lambda) = \sum_{i=1}^N \pi_i b_i(O_1) \beta_1(i) \quad (61)$$

Para visualizar como que o processo de indução termina, considere as expressões abaixo:

$$\begin{aligned} P(O|\lambda) &= \sum_{i=1}^N \pi_i b_i(O_1) \beta_1(i) \\ &= \sum_{i=1}^N \pi_i b_i(O_1) \left[ \sum_{j=1}^N a_{ij} b_j(O_2) \beta_2(j) \right] \\ &\quad \vdots \\ &= \sum_{i=1}^N \pi_i b_i(O_1) \left[ \sum_{j=1}^N a_{ij} b_j(O_2) \beta_2(j) \right] \cdots \sum_{l=1}^N a_{il} b_l(O_T) \beta_T(l) \end{aligned} \quad (62)$$

Por estes passos, pode-se visualizar que a variável *Backward* regride capturando todos os possíveis caminhos que a sequência de observações pode apresentar.

Inicializa-se o processo definindo como um a variável *Backward* no tempo  $t = T$ . O segundo passo, que é a indução, é ilustrada na figura a seguir:



Fonte: Rabiner (1989).

**Figura 6 - Ilustração da sequência de operações requeridas para a computação da variável Backward  $\beta_t(i)$**

Esta figura mostra que o processo estando no estado  $S_i$  no tempo  $t$ , deve-se considerar todos os estados possíveis  $S_j$  no tempo  $t + 1$  de acordo com as probabilidades de transição  $a_{ij}$ , com as probabilidades da observação no estado  $b_j(O_{t+1})$ , e com a sequência de observações parcial para o estado  $j$ ,  $\beta_{t+1}(j)$ .

O procedimento *Backward* requer o mesmo número de operações para efetuar o cálculo da probabilidade desejada, ou seja,  $N^2T$ . Na resolução dos próximos dois problemas do HMM, as variáveis estudadas nesta seção serão conjuntamente necessárias.

Como citado anteriormente, este problema é comumente utilizado para a comparação e escolha de modelos. A informação desejada é a probabilidade do modelo dada uma sequência de observações, ou seja,  $P(\lambda|O)$ . Com base no teorema de Bayes, pode-se reescrever essa probabilidade da seguinte forma (SOUZA 2010):

$$P(\lambda|O) = \frac{P(\lambda,O)}{P(O)} = \frac{P(O|\lambda)P(\lambda)}{P(O)} \quad (63)$$

Do ponto de vista Bayesiano,  $P(\lambda|O)$  representa a *posteriori*,  $P(O|\lambda)$  representa a verossimilhança da sequência de observações dado o modelo, e  $P(\lambda)$  seria a *apriori*; ou seja, geralmente a probabilidade de ocorrência de determinado modelo pode ser estimada matematicamente por meio de um conhecimento *apriori* sobre a aplicação em questão. A probabilidade da sequência das observações  $P(O)$  geralmente não é simples de ser calculada, porém, se o objetivo for a comparação de modelos e conseqüentemente, a comparação de  $P(\lambda|O)$ , pode-se considerar  $P(O)$  como uma constante, pois esta probabilidade não depende de nenhum parâmetro.

### 3.4.2- Melhor sequência de estados

Este problema, diferentemente do problema apresentado na seção anterior, pode ter vários modos de ser solucionado, isto é, de encontrar a sequência de estados que melhor explica a sequência de observações. Isto se deve ao fato de cada solução ter diferentes critérios para efetuar esta busca. Por exemplo, pode-se definir o critério de buscar sempre o estado  $q_t$  que individualmente é o mais provável para dada observação. Este método maximiza o número esperado de estados individuais corretos, porem pode apresentar alguns problemas como, por exemplo, sequências de estados impossíveis. Para implementar a solução deste problema, será definido a seguinte variável:

$$Y_t(i) = P(q_t = S_i|O, \lambda) \quad (64)$$

Isto é, a probabilidade do processo estar no estado  $S_i$  no tempo  $t$  dado a sequência de observações e o modelo. Pelo teorema da probabilidade condicional, pode-se mostrar que:

$$P(O, q_t = S_i | \lambda) = P(q_t = S_i | O, \lambda) \cdot P(O | \lambda) \quad (65)$$

Consequentemente pode-se reescrever (64) como:

$$\begin{aligned} P(q_t = S_i | O, \lambda) &= \frac{P(O, q_t = S_i | \lambda)}{P(O | \lambda)} \\ &= \frac{P(O_1, \dots, O_t, q_t = S_i | \lambda) \cdot P(O_{t+1}, \dots, O_T | q_t = S_i, \lambda)}{P(O_1, \dots, O_T | \lambda)} \\ &= \frac{\alpha_t(i) \cdot \beta_t(i)}{\sum_{i=1}^N \alpha_t(i) \cdot \beta_t(i)} \end{aligned} \quad (66)$$

Foi mostrado que esta variável pode ser expressa em termos de variáveis *Forward* e *Backward*, pois a variável *Forward* representa a sequência de observações parcial  $O_1 \dots O_t$  e o processo estar no estado  $S_i$  no tempo  $t$ , e a variável *Backward* representa a sequência de observações parcial  $O_{t+1} \dots O_T$  e o processo estar no estado  $S_i$  no tempo  $t$ . O fator de normalização  $\sum_{i=1}^N \alpha_t(i) \cdot \beta_t(i)$  faz a variável  $Y_t(i)$  se torna uma medida de probabilidade, e portanto, tem-se que:

$$\sum_{i=1}^N Y_t(i) = 1 \quad (67)$$

Para o critério do estado individualmente mais provável, pode-se usar esta variável para solucionar o problema como:

$$q_t = \operatorname{argmax}_{1 \leq i \leq N} [Y_t(i)] \quad 1 \leq t \leq T \quad (68)$$

Porém o foco deste problema não será por este critério, pois como já dito, pode criar sequências de estados impossíveis para o processo. O critério mais razoável e mais utilizado para solucionar este problema é buscar uma única sequência de estados que melhor explica a sequência das observações, isto é, maximiza  $P(Q | O, \lambda)$ , e consequentemente maximiza  $P(Q, O | \lambda)$ . Uma técnica para encontrar a única sequência de estados que maximiza esta probabilidade é baseada no método da programação dinâmica, e é chamada de algoritmo *Viterbi* (VITERBI, 1967 e FORNEY, 1973).

Dada um sequência de observações  $O = \{O_1 O_2 \dots O_T\}$ , para encontrar a melhor sequência de estados,  $Q = \{q_1 q_2 \dots q_T\}$ , deve-se definir a seguinte variável:

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} P[q_1 q_2 \dots q_t = i, O_1 O_2 \dots O_t | \lambda] \quad (69)$$

ou seja,  $\delta_t(i)$  é a maior probabilidade ao longo de uma única sequência no tempo  $t$ . Esta variável carrega a probabilidade das primeiras  $t$  observações e do processo no tempo  $t$  terminar no estado  $S_i$ . Por indução tem-se:

$$\delta_{t+1}(j) = [\max_i \delta_t(i) a_{ij}] \cdot b_j(O_{t+1}) \quad (70)$$

Para guardar a sequência de estados é necessário manter a faixa do argumento que maximiza (70) para cada  $t$  e  $j$ . É possível fazer tal feito pelo vetor  $\psi_t(j)$ .

Este procedimento utiliza a variável  $\delta_t(j)$  para o armazenamento das probabilidades da sequência dos estados e a variável  $\psi_t(j)$  para o armazenamento da sequência de estados. O procedimento completo para encontrar a melhor sequência de estados segue os seguintes passos (RABINER, 1989):

1) Inicialização:

$$\delta_1(i) = \pi_i b_i(O_1) \quad 1 \leq i \leq N \quad (71)$$

$$\psi_1(i) = 0 \quad (72)$$

Esta etapa do procedimento apenas serve para inicializa-lo, uma vez que  $\delta_1(j)$  não consegue fornecer o melhor estado para o tempo  $t = 1$ , logo, ele apenas guarda  $P(q_1, O_1 | \lambda)$  para o passo da recursão. Como nenhum estado é selecionado como o melhor, atribui-se zero para  $\psi_1(j)$ .

2) Recursão:

$$\delta_t(j) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] \cdot b_j(O_t) \quad 2 \leq t \leq T \quad (73)$$

$$1 \leq j \leq N$$

$$\psi_t(j) = \operatorname{argmax}_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] \quad 2 \leq t \leq T \quad (74)$$

$$1 \leq j \leq N$$

Esta etapa acumula as probabilidades das melhores sequências parciais a cada passo por  $\delta_t(j)$ , e acumula as melhores sequências parciais por  $\psi_t(j)$ . Será considerado agora o segundo tempo do procedimento afim de uma melhor visualização do mesmo:

$$\begin{aligned} \delta_2(j) &= \max_{1 \leq i \leq N} [\delta_1(i) a_{ij}] \cdot b_j(O_2) \\ &= \max_{1 \leq i \leq N} [\pi_i b_i(O_1) a_{ij}] \cdot b_j(O_2) \end{aligned} \quad (75)$$

ou seja,  $\delta_2(j)$  procura qual o estado que maximiza a probabilidade conjunta da primeira observação e o primeiro estado dado o modelo a partir da distribuição inicial  $\pi_i$ , da probabilidade da primeira observação dado o estado  $S_i$ ,  $b_i(O_1)$ , a probabilidade de transição  $a_{ij}$ , e a probabilidade da segunda observação dado o estado  $S_j$ ,  $b_j(O_2)$ .

3) Terminação:

$$P^* = \max_{1 \leq i \leq N} [\delta_T(i)] \quad (76)$$

$$q_T^* = \operatorname{argmax}_{1 \leq i \leq N} [\delta_T(i)] \quad (77)$$

Esta etapa fornece a melhor sequência de estados  $q_T^*$ , e sua probabilidade,  $P^*$ .

4) “Caminho de volta” (Sequência de estados):

$$q_t^* = \psi_{t+1}(q_{t+1}^*) \quad t = T - 1, T - 2, \dots, 1 \quad (78)$$

Por meio de (78), pode-se determinar a sequência de estados parcial, muitas vezes importante para o pesquisador.

### 3.4.3- Estimação dos parâmetros

Este último problema está atento em estimar os três parâmetros do HMM, os quais são: a matriz de transição de probabilidade  $A$ , a distribuição de probabilidades da observação no estado  $B$ , e a distribuição inicial do modelo  $\pi$ . O objetivo é determinar um método que ajusta os parâmetros do modelo que maximiza a probabilidade de uma sequência de observação, geralmente chamada de sequência de treinamento.

Não há analiticamente um método para estimar estes parâmetros de modo que maximize a probabilidade das sequências de observações, contudo, é possível determinar um modelo  $\lambda = (A, B, \pi)$  tal que  $P(O|\lambda)$  é maximizada localmente. Para a solução deste problema será utilizado o algoritmo de *Baum-Welch* (DEMPSTER *et al*, 1977), que se trata de uma particularização do algoritmo EM aplicado aos HMM's.

Com o intuito de apresentar o procedimento para a reestimação dos parâmetros do modelo, será definida a variável  $\xi_t(i, j)$ , que é a probabilidade conjunta do processo estar no estado  $S_i$  no tempo  $t$  e do processo estar no estado  $S_j$  no tempo  $t + 1$  dado o modelo e a sequência de observações, isto é:

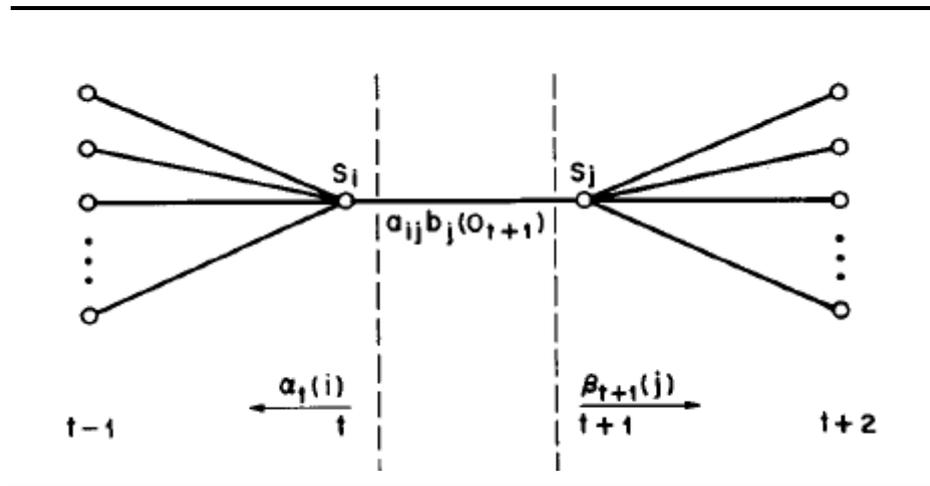
$$\xi_t(i, j) = P(q_t = S_i, q_{t+1} = S_j | O, \lambda) \quad (79)$$

Esta variável pode ser interpretada pela seguinte forma:

$$\begin{aligned} \xi_t(i, j) &= P(q_t = S_i, q_{t+1} = S_j | O, \lambda) \\ &\cong \frac{P(q_t = S_i, q_{t+1} = S_j | O, \lambda)}{P(O | \lambda)} \\ &= \frac{P(O_1, \dots, O_t, q_t = S_i | \lambda) a_{ij} b_j(O_{t+1}) P(O_{t+2}, \dots, O_T | q_{t+1} = S_j, \lambda)}{P(O | \lambda)} \\ &= \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)} \end{aligned} \quad (80)$$

ou seja, esta variável também pode ser expressa em termos das variáveis *Forward* e *Backward*. O termo  $P(O|\lambda)$  foi adicionado na expressão apenas para dar uma medida de

probabilidade na variável. Com base nestes resultados, é possível afirmar que esta variável pode ser ilustrada da seguinte maneira:



Fonte: Rabiner (1989).

**Figura 7 - Ilustração da sequência de operações requeridas para a computação do evento conjunto do sistema estar no estado  $S_i$  no tempo  $t$  e de estar no estado  $S_j$  no tempo  $t+1$**

Previamente vista na seção 2.4.2, a variável  $Y_t(i)$  pode ser relacionada à variável  $\xi_t(i, j)$  da seguinte maneira:

$$\begin{aligned}
 Y_t(i) &= P(q_t = S_i | O, \lambda) \\
 &= \sum_{j=1}^N P(q_t = S_i, q_{t+1} = S_j | O, \lambda) \\
 &= \sum_{j=1}^N \xi_t(i, j)
 \end{aligned} \tag{81}$$

Se somar  $Y_t(i)$  pelo índice de tempo  $t$  chega-se a quantidade  $\bar{Y}$  que pode ser interpretada como o número esperado de vezes que o estado  $S_i$  é visitado pelo processo, ou seja:

$$\bar{Y} = \sum_{t=1}^{T-1} Y_t(i) \tag{82}$$

Similarmente, se somar  $\xi_t(i, j)$  pelo índice de tempo  $t$  (de  $t = 1$  a  $t = T - 1$ ) chega-se a quantidade  $\bar{\xi}$  que pode ser interpretada como o número esperado de transições do estado  $S_i$  para o estado  $S_j$ , isto é:

$$\bar{\xi} = \sum_{t=1}^{T-1} \xi_t(i, j) \tag{83}$$

Utilizando (82) e (83) pode-se estabelecer um método de reestimação dos parâmetros do modelo,  $A$ ,  $B$  e  $\pi$ , dados por:

$$\bar{\pi}_i = \text{probabilidade de estar no estado } S_i \text{ no instante } 1 = Y_1(i) \tag{84}$$

$$\bar{a}_{ij} = \frac{\text{número de transições do estado } S_i \text{ para o estado } S_j}{\text{número de transições do estado } S_i \text{ para qualquer estado}}$$

$$= \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} = \frac{\bar{\xi}}{\bar{\gamma}} \quad (85)$$

$$\bar{b}_j(k) = \frac{\text{número de vezes que foi observado a observação } O_k \text{ no estado } S_i}{\text{número de vezes no estado } S_i}$$

$$= \frac{\sum_{t=1}^T \gamma_t(i) o_{t=v_k}}{\sum_{t=1}^T \gamma_t(i)} \quad (86)$$

Dado o modelo  $\lambda = (A, B, \pi)$ , o procedimento de reestimação estabelece novos parâmetros  $\bar{\lambda} = (\bar{A}, \bar{B}, \bar{\pi})$ , e foi provado por Baker (1975), Baum e Sell (1968) que ou o modelo antigo  $\lambda$  é o ponto máximo da função de verossimilhança, que neste caso  $\bar{\lambda} = \lambda$ , ou o modelo novo  $\bar{\lambda}$  é mais provável que o modelo antigo  $\lambda$ , ou seja,  $P(O|\bar{\lambda}) > P(O|\lambda)$ .

Com o intuito de reestimar os parâmetros do modelo, o algoritmo *Baum-Welch* realiza muitas iterações sobre a sequência de observações dada, por isso deve-se criar um critério para finalizar a reestimação dos parâmetros. O resultado final do procedimento de reestimação é chamado de estimativa de máxima verossimilhança do modelo.

Uma vez apresentados os conceitos de controle estatístico de qualidade, de controle estatístico *on-line* e dos modelos ocultos de Markov, o próximo capítulo será apresentado uma aplicação do HMM no controle estatístico *on-line* para a estimação das probabilidades de alarme falso e a não detecção, que são importantes para o planejamento do processo.

## 4- HMM no controle estatístico *on-line*

No capítulo 2 foi apresentado o controle estatístico de qualidade e uma de suas ferramentas mais utilizadas hoje em dia, o controle estatístico *on-line*, que serve para monitorar processos de produção a fim de estabelecer níveis de qualidade adequados para os produtos. No capítulo 3 foi apresentado o modelo oculto de Markov e alguns exemplos.

O controle *on-line* tem a característica de ter intervalos de amostragem iguais, o que possibilita o uso de cadeias de Markov para sua modelagem. Porém, pode-se considerar que este processo tem um sistema não observável importante para o modelo, o qual é a real situação da linha de produção, isto é, se o processo está operando sob controle ou fora dele. Com base nesta informação, pode-se concluir que este sistema é regido por duas cadeias de Markov, sendo elas dependente entre si, ou seja, um modelo oculto de Markov.

Visto que é possível modelar o controle estatístico *on-line* via modelo oculto de Markov, este capítulo apresenta esta metodologia com o intuito de realizar a estimação das probabilidades de alarme falso e a de não detecção.

Este capítulo está disposto em 4.1 Introdução, 4.2 Alarme falso e Não detecção.

### 4.1- Introdução

Considere o seguinte monitoramento por controle *on-line* por variável para processos de produção: Um intervalo de aceitação  $C$  é proposto baseado nos requerimentos de qualidade da indústria e na distribuição da característica de qualidade. A cada  $m$  itens produzidos, um único item é retirado do processo de produção e examinado. Se o item examinado estiver dentro deste intervalo de aceitação o processo é dito estar sob controle estatístico e o processo continua, caso contrário, o processo é declarado fora de controle estatístico e é interrompido para o ajuste. Após o ajuste do processo um novo ciclo começa.

Um ciclo, como definido antes, é o intervalo de tempo que o processo está sob controle e a detecção e eliminação de alguma causa especial, ou seja, a cada parada para ajuste um novo ciclo é iniciado.

Agora suponha que este processo possa ser modelado por dois processos estocásticos distintos, porém dependentes. O primeiro processo seria a verdadeira situação do processo (não observável), isto é, se realmente está sob controle ou fora de controle. O segundo processo seria a declaração do processo, ou seja, se o item

examinado estiver dentro do intervalo de aceitação o processo é considerado sob controle, caso contrário, é considerado fora de controle.

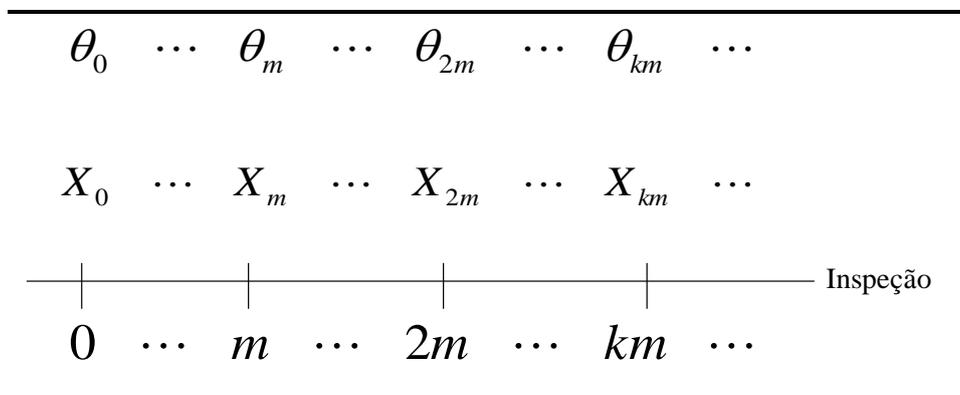
Ho *et al.* (2007) propuseram um modelo para este tipo de processo que é governado por uma cadeia de Markov de dois estágios, ou melhor, um modelo oculto de Markov (HMM).

Como todo HMM, este modelo possui dois processos estocásticos, sendo que o primeiro é não observável, e o segundo, que está interligado ao primeiro, é observável. O primeiro processo estocástico é definido por  $\{\theta_t\}_{t \geq 0}$ , onde  $\theta_t = 0$  o processo está sob controle e  $\theta_t = 1$  o processo está fora de controle. Como nunca se sabe qual é o verdadeiro estado em que o processo se encontra, este processo estocástico é não observável. O segundo processo estocástico é definido pela variável observável  $X_t$ , relativa a característica de qualidade que está se medindo. A cada  $m$  amostras seleciona um item e examina o produto com o intuito de saber se a estatística calculada está no intervalo de aceitação  $C$  definido por:

$$C = (\mu_0 - k\sigma, \mu_0 + k\sigma) \tag{87}$$

que pela literatura é definido considerando uma função de custo a ser minimizada.

A figura a seguir descreve o monitoramento do processo via HMM:



Fonte: Adaptado de Ho *et al.* (2007).

**Figura 8 - Descrição do processo de inspeção**

Portanto, pode-se notar que uma cadeia de Markov  $\{\theta_t\}_{t \geq 0}$  descreve a evolução da real situação do processo e, associada a esta cadeia, pode-se observar uma sequência de variáveis aleatórias independentes condicionadas  $\{X_t\}_{t \geq 1}$ , com a distribuição de cada variável  $X_t$  dependendo de cada estado correspondente  $\theta_t$ . Como este processo representa um HMM, têm-se as seguintes propriedades (DOREA *et al.*, 2012):

$$P(X_{t+1} \in A | X_1, \dots, X_t, \theta_0, \dots, \theta_t) = P((X_{t+1} \in A | \theta_t) \tag{88}$$

$$P(X_1 \in A_1, \dots, X_t \in A_t | \theta_1, \dots, \theta_t) = \prod_{i=1}^t P(X_i \in A_i | \theta_i) \quad (89)$$

sendo que  $A_i$  são eventos.

A expressão (88) representa a propriedade de Markov, ou seja, é irrelevante o conhecimento dos estados anteriores para a predição do estado seguinte, desde que o estado atual seja conhecido. (89) mostra que as observações são independentes uma das outras, uma propriedade muito importante para várias suposições e definições deste trabalho.

Utilizando as definições do capítulo 3 têm-se que:

i)  $\mathbf{N}$  representa o número de estados do modelo, ou seja, nesta aplicação  $N = 2$ , pois o processo pode se encontrar sob controle ou fora de controle.

ii)  $\mathbf{S} = \{S_0, S_1, \dots, S_{N-1}\}$  corresponde ao conjunto dos estados individuais do modelo, que neste caso serão definidos por:

$$S_0 = \textit{processo sob controle}$$

$$S_1 = \textit{processo fora de controle}$$

iii)  $q_t$  corresponde ao estado no tempo  $t$ .

iv)  $\mathbf{M}$  representa o número de observações distintas por estado, isto é,  $M = 2$ , pois o processo pode ser declarado sob controle ou fora de controle.

v)  $\mathbf{V} = \{v_1, v_2, \dots, v_M\}$  corresponde ao conjunto de observações individuais. Neste modelo as observações individuais são definidas por:

$$v_1 = \{X_t \in C\} = \textit{processo declarado sob controle}$$

$$v_2 = \{X_t \notin C\} = \textit{processo declarado fora de controle}$$

vi)  $\mathbf{A} = \{a_{ij}\}$  corresponde a distribuição de probabilidades de transição dos estados, dada por:

$$a_{ij} = P(q_{t+1} = S_j | q_t = S_i) \quad i, j = 0, 1 \quad (90)$$

Para esta aplicação, esta distribuição de probabilidades de transição é governada por uma distribuição geométrica.

vii)  $\mathbf{B} = \{b_j(k)\}$  corresponde a distribuição de probabilidade da observação no estado, dada por:

$$b_j(k) = P(v_k \textit{ em } t | q_t = S_j) \quad j = 0, 1 \quad (91)$$

$$k = 1, 2$$

No modelo em destaque, definem-se duas probabilidades das observações em cada estado, dado por:

- Probabilidade de declarar o processo sob controle sendo que ele está sob controle:

$$b_0(1) = P(v_1 \text{ em } t | q_t = S_0) \quad (92)$$

- Probabilidade de declarar o processo sob controle sendo que ele está fora de controle (Não detecção):

$$b_1(1) = P(v_1 \text{ em } t | q_t = S_1) \quad (93)$$

- Probabilidade de declarar o processo fora de controle sendo que ele está sob controle (Alarme falso):

$$b_0(2) = P(v_2 \text{ em } t | q_t = S_0) \quad (94)$$

- Probabilidade de declarar o processo fora de controle sendo que ele está fora de controle (Poder):

$$b_1(2) = P(v_2 \text{ em } t | q_t = S_1) \quad (95)$$

No modelo proposto, o alarme falso acontece quando a característica de qualidade observada caia fora do intervalo de aceitação  $C$ , ou seja,  $X_t \notin C$ , sendo que o processo está sob controle  $\theta_t = 0$ . A não detecção ocorre quando a característica de qualidade observada caia dentro do intervalo de aceitação, isto é,  $X_t \in C$ , sendo que o processo está fora de controle  $\theta_t = 1$ .

Na próxima seção serão definidas as probabilidades de alarme falso e de não detecção assumindo conhecidas as equações (88) e (89).

#### 4.2- Alarme falso e Não detecção

Para um HMM  $\{\theta_t, X_t\}$ , assume-se que a cadeia  $\{\theta_t\}$  tem distribuição de probabilidades conhecida, e dada por:

$$P = \begin{bmatrix} 1-p & p \\ \epsilon & 1-\epsilon \end{bmatrix} \quad 0 < p < 1 \quad (96)$$

$$\epsilon > 0$$

$$p + \epsilon < 1$$

Assume-se que a distribuição condicional de  $X_t$  dado  $\theta_t$  também é conhecida, e baseada na distribuição de probabilidades, uma região de controle  $C$  é pré-selecionada. Tem-se que (DOREA, 2012):

$$P(X_t \in A | \theta_t = i) = \int_A f(x|i) dx \quad (97)$$

A partir de (97), pode-se definir a probabilidade de um item selecionado cair fora da região de aceitação dado que o processo está sob cotrole, dada por:

$$0 < q_0 = \int_{C^c} f(x|0)dx < 1 \quad (98)$$

e podem-se definir as probabilidades do item selecionado cair fora da região de aceitação dado que o processo está fora de controle e do item selecionado cair dentro da região de aceitação dado que o processo está fora de controle, consecutivamente dadas por:

$$0 < q_1 = \int_{C^c} f(x|1)dx < 1 \quad (99)$$

$$0 < (1 - q_1) \int_C f(x|1)dx < 1 \quad (100)$$

Supondo que o intervalo amostral seja fixo e igual a um, ou seja,  $m = 1$ , o monitoramento de qualidade das cartas de controle on-line adota a seguinte estratégia:

i) Os itens  $X_1, X_2, \dots, X_t$  são inspecionados e verificados enquanto todos pertencem à região de aceitação  $C$ .

ii) A manutenção é necessária no tempo  $t$  se  $X_1 \in C, X_2 \in C, \dots, X_t \notin C$ .

Podem-se definir os tempos de alerta, ou seja, tanto quando acontece o alarme falso quanto o alarme é verdadeiro, como (DOREA *et al.*, 2012):

$$\tau_X = \inf \{k: k \geq 1, X_k \notin C\} \quad (101)$$

isto é, o tempo de alerta é definido como o menor valor de tempo  $k$  em que a característica de qualidade está fora da região de confiança. Portanto, o alarme falso acontece no tempo  $k$  se  $\tau_X = k$  e o processo não observável esteja sob controle, ou seja,  $\theta_k = 0$ . Agora será definido o primeiro tempo em que o processo atinge o estado fora de controle como:

$$\tau_\theta = \inf \{k: k \geq 1, \theta_k = 1\} \quad (102)$$

A partir destas variáveis, pode-se afirmar que o alarme falso acontece quando ( $\tau_X < \tau_\theta$ ), e que a não detecção acontece quando ( $\tau_X > \tau_\theta$ ).

Seja  $P_\nu$  a probabilidade quando a distribuição inicial de  $\theta_0$  é  $\nu$ , ou seja,  $P_0$  e  $P_1$  denota as probabilidades quando  $\theta_0 = 0$  e  $\theta_0 = 1$  respectivamente. Definindo  $\nu$  como a distribuição inicial de  $\theta_0$ , a probabilidade de alarme falso é dada por:

$$P_\nu(\tau_X < \tau_\theta) = \nu(0) \frac{(1-p)q_0}{1-(1-p)(1-q_0)} + \nu(1) \frac{\epsilon q_0}{1-(1-p)(1-q_0)} \quad (103)$$

**Prova:**

Para  $\theta_0 = 0$ , pode-se mostrar por (96), (97), (98), (99) e (100) que:

( $k = 1$ )

$$\begin{aligned} P_0(\tau_X = 1, \tau_\theta > 1) &= P_0(\theta_1 = 0, X_1 \notin C) \\ &= P_0(X_1 \notin C | \theta_1 = 0)P_0(\theta_1 = 0) \\ &= (1 - p)q_0 \end{aligned}$$

( $k \geq 2$ )

$$\begin{aligned} P_0(\tau_X = k, \tau_\theta > k) &= P_0(X_1 \in C, \dots, X_{k-1} \in C, X_k \notin C, \theta_1 = \theta_2 = \dots = \theta_k = 0) \\ &= P_0(\theta_1 = \dots = \theta_k = 0) \prod_{j=1}^{k-1} \{P_0(X_j \in C | \theta_j = 0)\} P_0(X_k \notin C | \theta_k = 0) \\ &= (1 - p)^k (1 - q_0)^{k-1} q_0 \\ &= (1 - p)[(1 - p)(1 - q_0)]^{k-1} q_0 \end{aligned}$$

E similarmente, se  $\theta_0 = 1$ , pode-se mostrar por (96), (97), (98), (99) e (100) que:

( $k = 1$ )

$$\begin{aligned} P_1(\tau_X = 1, \tau_\theta > 1) &= P_1(\theta_1 = 0, X_1 \notin C) \\ &= P_1(X_1 \notin C | \theta_1 = 0)P_1(\theta_1 = 0) \\ &= \epsilon q_0 \end{aligned}$$

( $k \geq 2$ )

$$\begin{aligned} P_1(\tau_X = k, \tau_\theta > k) &= P_1(X_1 \in C, \dots, X_{k-1} \in C, X_k \in C, \theta_1 = \theta_2 = \dots = \theta_k = 0) \\ &= P_1(\theta_1 = \dots = \theta_k = 0) \prod_{j=1}^{k-1} \{P_1(X_j \in C | \theta_j = 0)\} P_1(X_k \in C | \theta_k = 0) \\ &= \epsilon (1 - p)^{k-1} (1 - q_0)^{k-1} q_0 \\ &= \epsilon [(1 - p)(1 - q_0)]^{k-1} q_0 \end{aligned}$$

Com base nestes resultados, será feita uma analogia a uma simples probabilidade condicional com o teorema da probabilidade total, para provar (103):

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} = \frac{\sum_i P(A_i \cap B)}{P(B)} \\ &= \frac{P_0(\tau_X < \tau_\theta | \tau_X, \tau_\theta \geq 1)}{P(\tau_X, \tau_\theta \geq 1)} \\ &= \frac{P(\tau_X < \tau_\theta \cap \tau_X, \tau_\theta \geq 1)}{P(\tau_X, \tau_\theta \geq 1)} \\ &= \frac{P_0(\tau_X < \tau_\theta \cap \tau_X, \tau_\theta \geq 1) + P_1(\tau_X < \tau_\theta \cap \tau_X, \tau_\theta \geq 1)}{P(\tau_X, \tau_\theta \geq 1)} \\ &= \frac{v(0)(1 - p)q_0 + v(1)\epsilon q_0}{1 - (1 - p)(1 - q_0)} \end{aligned}$$

$$= v(0) \frac{(1-p)q_0}{1-(1-p)(1-q_0)} + v(1) \frac{\epsilon q_0}{1-(1-p)(1-q_0)}$$

Provado a probabilidade do alarme falso, agora será definida a probabilidade de não detecção, dada por:

$$P_v(\tau_X > \tau_\theta) = v(0) \frac{p(1-q_1)}{1-(1-p)(1-q_0)} + v(1) \frac{[q_0(1-\epsilon)+(1-q_0)p](1-q_1)}{1-(1-p)(1-q_0)} \quad (104)$$

**Prova:**

Note que:

( $k = 1$ )

$$\begin{aligned} P_0(\tau_X = \tau_\theta = 1) &= P_0(\theta_1 = 1, X_1 \notin C) \\ &= P(\theta_1 = 1 | \theta_0 = 0) P(X_1 \notin C | \theta_1 = 1) \\ &= pq_1 \end{aligned}$$

( $k = 2$ )

$$\begin{aligned} P_0(\tau_X = \tau_\theta = k) &= P_0(\theta_k = 1, X_k \notin C) \\ &= P_0(X_1 \in C, \dots, X_{k-1} \in C, X_k \notin C, \theta_1 = \theta_2 = \dots = \theta_{k-1} = 0, \theta_k = 1) \\ &= P_0(\theta_1 = \dots = \theta_{k-1} = 0) P_1(\theta_k = 1) \prod_{j=1}^{k-1} \{P_0(X_j \in C | \theta_j = 0)\} P_0(X_k \notin C | \theta_k = 1) \\ &= (1-p)^{k-1} p (1-q_0)^{k-1} q_1 \\ &= [(1-p)(1-q_0)]^{k-1} pq_1 \end{aligned}$$

E similarmente, para  $\theta_0 = 1$ , têm-se:

( $k = 1$ )

$$\begin{aligned} P_1(\tau_X = \tau_\theta = 1) &= P_1(\theta_1 = 1, X_1 \notin C) \\ &= P(\theta_1 = 1 | \theta_0 = 1) P(X_1 \notin C | \theta_1 = 1) \\ &= (1-\epsilon)q_1 \end{aligned}$$

( $k = 2$ )

$$\begin{aligned} P_1(\tau_X = \tau_\theta = k) &= P_1(\theta_k = 1, X_k \notin C) \\ &= P_1(X_1 \in C, \dots, X_{k-1} \in C, X_k \notin C, \theta_1 = \theta_2 = \dots = \theta_{k-1} = 0, \theta_k = 1) \\ &= P_1(\theta_1 = \dots = \theta_{k-1} = 0) P_1(\theta_k = 1) \prod_{j=1}^{k-1} \{P_1(X_j \in C | \theta_j = 0)\} P_1(X_k \notin C | \theta_k = 1) \\ &= \epsilon(1-p)^{k-2} p (1-q_0)^{k-1} q_1 \\ &= \epsilon(1-q_0) [(1-p)(1-q_0)]^{k-2} pq_1 \end{aligned}$$

A partir destas definições, tem-se que:

$$P_v(\tau_X = \tau_\theta = k) = v(0) \frac{pq_1}{1-(1-p)(1-q_0)} + v(1) \frac{[q_0(1-\epsilon)+(1-q_0)p]q_1}{1-(1-p)(1-q_0)} \quad (105)$$

Com esta expressão, tem-se:

$$\begin{aligned}
P_v(\tau_X > \tau_\theta) &= 1 - [P_v(\tau_X > \tau_\theta) + P_v(\tau_X = \tau_\theta = k)] \\
P_v(\tau_X > \tau_\theta) &= 1 - [v(0) \frac{(1-p)q_0}{1-(1-p)(1-q_0)} + v(1) \frac{\epsilon q_0}{1-(1-p)(1-q_0)} \\
&\quad + v(0) \frac{pq_1}{1-(1-p)(1-q_0)} + v(1) \frac{[q_0(1-\epsilon) + (1-q_0)p]q_1}{1-(1-p)(1-q_0)}] \\
&\quad \vdots \\
P_v(\tau_X > \tau_\theta) &= v(0) \frac{p(1-q_1)}{1-(1-p)(1-q_0)} + v(1) \frac{[q_0(1-\epsilon) + (1-q_0)p](1-q_1)}{1-(1-p)(1-q_0)}
\end{aligned}$$

Agora, assume-se que um único item é inspecionado a cada  $m$  itens produzidos.

Então, para  $m \geq 2$ , pode-se definir o tempo de alerta como:

$$\tau_X^{(m)} = \inf \{km: k \geq 1, X_{km} \notin C\} \quad (106)$$

Usando os mesmos argumentos utilizados para provar (103) e (104), analogamente pode-se mostrar que:

$$\begin{aligned}
P_0(\tau_X^{(m)} = km, \tau_\theta > km) &= P_0(X_m \in C, \dots, X_{km-m} \in C, X_{km} \notin C, \theta_m = \theta_{2m} = \dots \\
&= \theta_{km} = 0)
\end{aligned}$$

$$\begin{aligned}
P_0(\theta_m = \dots = \theta_{km} = 0) &\prod_{j=1}^{k-1} \{P_0(X_{jm} \in C | \theta_{jm} = 0)\} P_0(X_{km} \notin C | \theta_{km} = 0) \\
&= (1-p)^{km} (1-q_0)^{k-1} q_0 \\
&= (1-p)^m [(1-p)^m (1-q_0)]^{k-1} q_0
\end{aligned}$$

e conseqüentemente:

$$\begin{aligned}
P_1(\tau_X^{(m)} = k, \tau_\theta > k) &= P_1(X_m \in C, \dots, X_{km-m} \in C, X_{km} \in C, \theta_m = \theta_{2m} = \dots = \theta_{km} \\
&= 0)
\end{aligned}$$

$$\begin{aligned}
&= P_1(\theta_m = \dots = \theta_{km} = 0) \prod_{j=1}^{k-1} \{P_1(X_{jm} \in C | \theta_{jm} = 0)\} P_1(X_{km} \notin C | \theta_{km} = 0) \\
&= \epsilon (1-p)^{km-1} (1-q_0)^{k-1} q_0 \\
&= \epsilon (1-p)^{m-1} [(1-p)^m (1-q_0)]^{k-1} q_0
\end{aligned}$$

Além disso, têm-se que:

$$\begin{aligned}
P_0(\tau_X^{(m)} = \tau_\theta = k) &= P_0(\theta_{km} = 1, X_{km} \notin C) \\
&= P_0(X_m \in C, \dots, X_{km-m} \in C, X_{km} \notin C, \theta_m = \theta_{2m} = \dots = \theta_{km-m} = 0, \theta_{km} = 1)
\end{aligned}$$

$$\begin{aligned}
&= P_0(\theta_m = \dots = \theta_{km-m} = 0)P_1(\theta_{km} = 1) \prod_{j=1}^{k-1} \{P_0(X_{jm} \in C | \theta_{jm} = 0)\} P_0(X_{km} \\
&\quad \notin C | \theta_{km} = 1) \\
&\quad = (1-p)^{km-1} p (1-q_0)^{k-1} q_1 \\
&\quad = (1-p)^{m-1} [(1-p)^m (1-q_0)]^{k-1} p q_1
\end{aligned}$$

e conseqüentemente:

$$\begin{aligned}
&P_1(\tau_X^{(m)} = \tau_\theta = k) = P_1(\theta_k = 1, X_k \notin C) \\
&= P_1(X_m \in C, \dots, X_{km-m} \in C, X_{km} \notin C, \theta_m = \theta_{2m} = \dots = \theta_{km-m} = 0, \theta_{km} = 1) \\
&= P_1(\theta_m = \dots = \theta_{km-m} = 0)P_1(\theta_{km} = 1) \prod_{j=1}^{k-1} \{P_1(X_{jm} \in C | \theta_{jm} = 0)\} P_1(X_{km} \\
&\quad \notin C | \theta_{km} = 1) \\
&\quad = \epsilon (1-p)^{km-2} p (1-q_0)^{k-1} q_1 \\
&\quad = \epsilon (1-p)^{m-2} [(1-p)^m (1-q_0)]^{k-1} p q_1
\end{aligned}$$

A partir destes resultados e assumindo que a estratégia de monitoramento do processo tem um intervalo amostral  $m$ , pode-se determinar a probabilidade do alarme falso como:

$$P_v(\tau_X^{(m)} < \tau_\theta) = \nu(0) \frac{(1-p)^m q_0}{1-(1-p)^m (1-q_0)} + \nu(1) \frac{\epsilon (1-p)^{m-1} q_0}{1-(1-p)^m (1-q_0)} \quad (107)$$

e para  $m \geq 2$ , a probabilidade de não detecção é dada por:

$$\begin{aligned}
&P_v(\tau_X^{(m)} > \tau_\theta) = \\
&\nu(0) \frac{1-(1-p)^{m-1} + (1-p)^{m-1} p (1-q_1)}{1-(1-p)^m (1-q_0)} + \nu(1) \frac{1-(1-p)^m (1-q_0) - \epsilon (1-p)^{m-2} [(1-p)q_0 + p q_1]}{1-(1-p)^m (1-q_0)} \quad (108)
\end{aligned}$$

Estas probabilidades de alarme falso e de não detecção são importantes para o planejamento e monitoramento das cartas de controle On-line, pois elas estão diretamente ligadas aos custos e ao modelo estatístico.

No próximo capítulo será feita uma análise de sensibilidade destas probabilidades com o intuito de descobrir quais parâmetros que mais influenciam nesta estimação.

## 5- Análise de Sensibilidade

Neste capítulo será feita uma análise de sensibilidade das probabilidades de alarme falso e de não detecção baseada no exemplo numérico de Ho *et al.* (2007) a fim de constatar quais os parâmetros que influenciam mais em cada probabilidade. O programa que será utilizado para efetuar os cálculos será o *R*, um software estatístico de código livre.

Ho *et al.* (2007) considera o problema em que uma fabrica de circuitos integrados quer instalar um sistema de controle a fim de medir a dimensão de uma característica de qualidade. Esta variável segue uma distribuição normal com média  $\mu_0$  e desvio-padrão  $\sigma_0$ , sendo que o especialista na área de planejamento determinou um intervalo de itens entre amostras igual a 5 itens, e por questões de qualidade, definiu a largura dos limites de controle igual a 3.

A partir de alguns cálculos, constata-se que em média, o processo muda do estado sob controle para o estado fora de controle depois de 20 itens produzidos e que a magnitude desta mudança são de 1,5 desvio-padrão na média, sendo que a variabilidade do processo continua a mesma, ou seja, o processo fora de controle segue uma distribuição normal com média  $\mu_1 = \mu_0 + \delta\sigma_0$  e com o mesmo desvio-padrão  $\sigma_0$ . O processo fora de controle retorna para as condições normais sem intervenção humana com probabilidade igual a 0,5.

Logo as probabilidades de alarme falso e de não detecção são dadas por (107) e (108) respectivamente, com  $p = \frac{1}{20} = 0,05$ ,  $\varepsilon = 0,5$ ,  $v(0) = 0,95$ ,  $v(1) = 0,05$ ,  $m = 5$ ,

$$k = 3, \quad q_0 = \int_{-\infty}^{\mu_0 - 3\sigma_0} f(x|\mu = \mu_0, \sigma = \sigma_0)dx + \int_{\mu_0 + 3\sigma_0}^{\infty} f(x|\mu = \mu_0, \sigma = \sigma_0)dx,$$

$$q_1 = \int_{-\infty}^{\mu_0 - 3\sigma_0} f(x|\mu = \mu_1, \sigma = \sigma_1 = \sigma_0)dx + \int_{\mu_0 + 3\sigma_0}^{\infty} f(x|\mu = \mu_1, \sigma = \sigma_1 = \sigma_0)dx.$$

Com base nestes parâmetros, foram obtidas as seguintes probabilidades:

$$- P(\text{Alarme falso}) = 0,0089334$$

$$- P(\text{Não detecção}) = 0,9794312$$

Agora serão feitas análises dos seguintes parâmetros: largura dos limites de controle ( $k$ ), os parâmetros da cadeia da real situação do processo ( $p, \varepsilon$ ), do intervalo

entre amostras ( $m$ ) e da magnitude da mudança da média do processo ( $\delta$ ). Esta análise permitirá concluir quais parâmetros que mais influenciam nas probabilidades de alarme falso e de não detecção.

- Variação da largura dos limites de controle ( $k$ ):

A tabela (2) apresenta a variação das probabilidades de alarme falso e de não detecção com a variação da largura dos limites de controle:

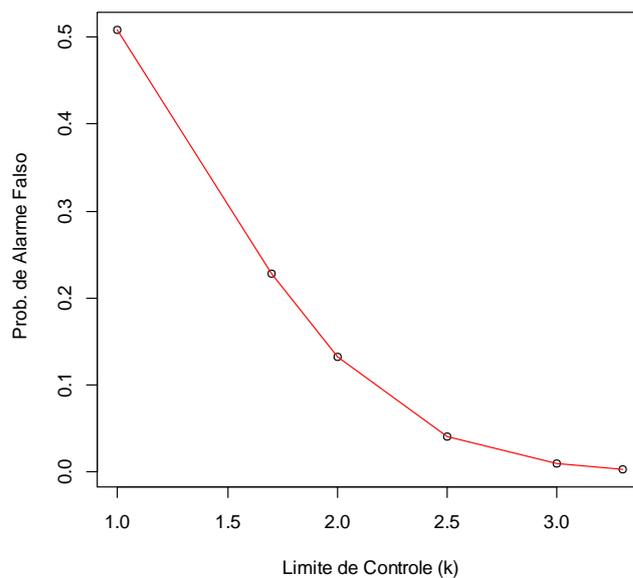
**Tabela 2 – Valores das probabilidades para diferentes larguras de limites de controle**

Largura dos limites de controle (k)	Probabilidades	
	Alarme Falso	Não detecção
1	0,5081394	0,4330581
1,7	0,2281075	0,7151274
2	0,1314841	0,8215546
2,5	0,0397841	0,9334613
<b>3</b>	<b>0,0089334</b>	<b>0,9794312</b>
3,3	0,0032181	0,9904874

Fonte: Elaboração própria com base nos resultados da pesquisa.

A partir desta tabela, pode-se afirmar que ambas as probabilidades são bastante sensíveis com a variação da largura dos limites de controle, pois sua variação provoca grandes mudanças nestas probabilidades.

O gráfico (1) apresenta a variação da probabilidade de alarme falso com a variação da largura dos limites de controle:

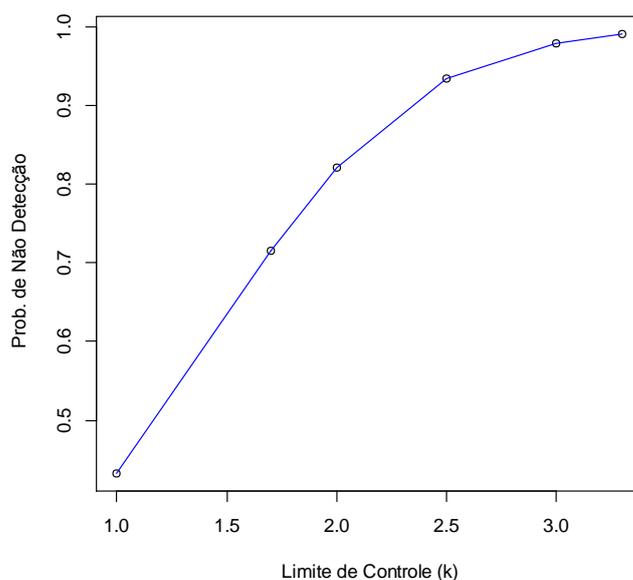


Fonte: Elaboração própria com base nos resultados da pesquisa.

**Gráfico 1 - Sensibilidade da probabilidade de alarme falso/ largura dos limites de controle (k)**

De acordo com o gráfico (1), quando a largura dos limites de controle aumenta, a probabilidade de alarme falso cai exponencialmente. Isto se deve ao fato de quanto maior for a região de aceitação, ou seja, a largura dos limites de controle, menor será a probabilidade de um item cair fora da região de aceitação dado que o processo está sob controle.

O gráfico (2) apresenta a variação da probabilidade de não detecção com a variação da largura dos limites de controle:



Fonte: Elaboração própria com base nos resultados da pesquisa.

### **Gráfico 2 – Sensibilidade da probabilidade de não detecção/ largura dos limites de controle (k)**

Ao contrário da probabilidade de alarme falso, pode-se afirmar a partir do gráfico (2) que a probabilidade de não detecção aumenta quando a largura dos limites de controle. Isto acontece pois quanto maior for a região de aceitação, mais itens cairão na região de aceitação enquanto o processo estiver fora de controle.

- Variação do intervalo entre amostras ( $m$ ):

A tabela (3) apresenta a variação das probabilidades de alarme falso e de não detecção com a variação do intervalo entre amostras:

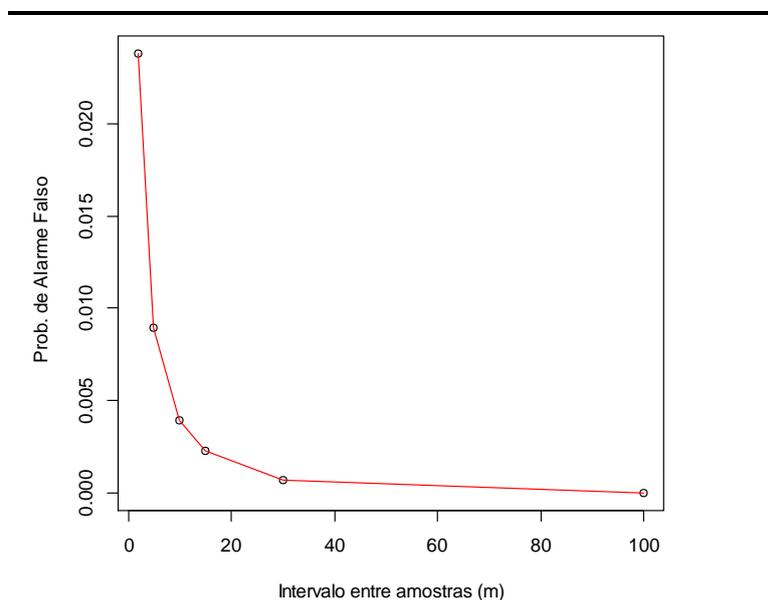
**Tabela 3 - Valores das probabilidades para diferentes intervalos entre amostras**

Intervalo entre amostras (m)	Probabilidades	
	Alarme Falso	Não detecção
2	0,0238036	0,9451932
<b>5</b>	<b>0,0089334</b>	<b>0,9794312</b>
10	0,0039172	0,9898077
15	0,0022699	0,9947734
30	0,0007198	0,9983426
100	0,000156	0,9999639

Fonte: Elaboração própria com base nos resultados da pesquisa.

A partir desta tabela, pode-se afirmar que ambas as probabilidades são bastante sensíveis com a variação do intervalo entre amostras, pois sua variação provoca grandes mudanças nestas probabilidades.

O gráfico (3) apresenta a variação da probabilidade de alarme falso com a variação do intervalo entre amostras:



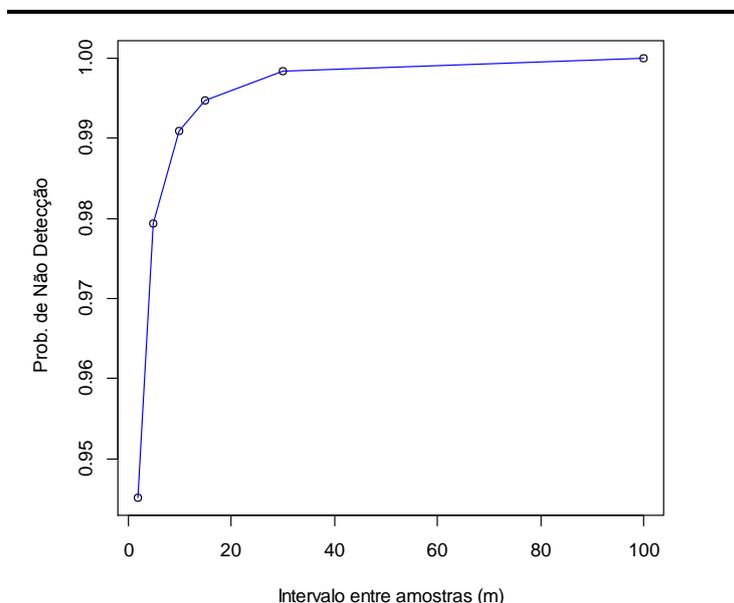
Fonte: Elaboração própria com base nos resultados da pesquisa.

**Gráfico 3 - Sensibilidade da probabilidade de alarme falso / intervalo entre amostras (m)**

A partir do gráfico (3), pode-se dizer que quando o intervalo entre amostras aumenta, a probabilidade de alarme falso cai exponencialmente. Esta diminuição se

deve ao fato do sistema coletar menos amostras, o que resulta conseqüentemente em mais alarmes verdadeiros do que alarmes falsos.

O gráfico (4) apresenta a variação da probabilidade de não detecção com a variação do intervalo entre amostras:



Fonte: Elaboração própria com base nos resultados da pesquisa.

**Gráfico 4 - Sensibilidade da probabilidade de não detecção / intervalo entre amostras (m)**

Pode-se afirmar a partir do gráfico (4) que quando o intervalo de amostras aumenta, a probabilidade de não detecção aumenta também. Isto se deve ao fato de quanto maior for o intervalo entre amostras, maior chance o processo tem de voltar para o estado fora de controle dado que ele está fora de controle. Esta análise está muito ligada aos parâmetros  $\delta$  e  $\varepsilon$ ; caso fossem diferentes, esta análise não seria válida.

- Variação do parâmetro  $p$ :

A tabela (4) apresenta a variação das probabilidades de alarme falso e de não detecção com a variação parâmetro  $p$  da cadeia:

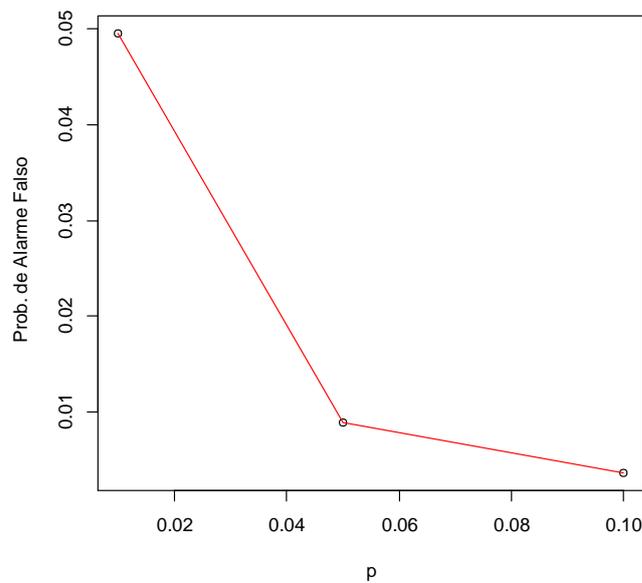
**Tabela 4 - Valores das probabilidades para diferentes valores de parâmetro  $p$**

Parâmetro $p$	Probabilidades	
	Alarme Falso	Não detecção
0,01	0,0495327	0,9380858
<b>0,05</b>	<b>0,0089334</b>	<b>0,9794312</b>
0,1	0,0037055	0,9861058

Fonte: Elaboração própria com base nos resultados da pesquisa.

A partir desta tabela, pode-se afirmar que ambas as probabilidades são bastante sensíveis com a variação do parâmetro  $p$ , pois sua variação provoca grandes mudanças nestas probabilidades.

O gráfico (5) apresenta a variação da probabilidade de alarme falso com a variação do parâmetro  $p$ :



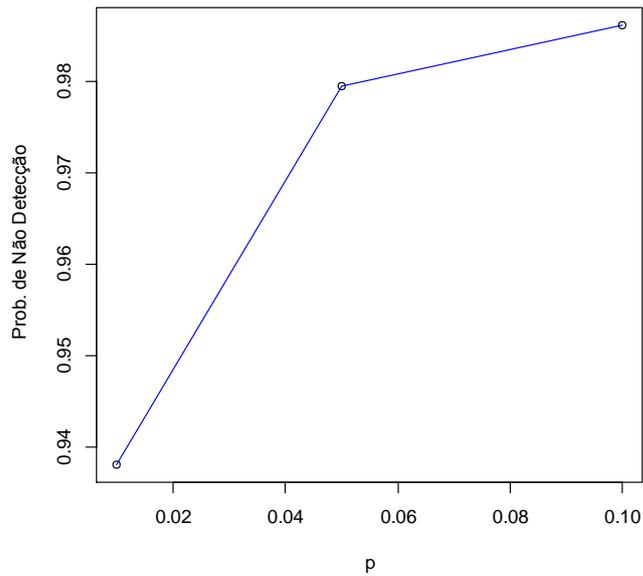
---

Fonte: Elaboração própria com base nos resultados da pesquisa.

### **Gráfico 5 - Sensibilidade da probabilidade de alarme falso / parâmetro $p$**

De acordo com o gráfico (5), tem-se que quando o parâmetro  $p$  aumenta, a probabilidade de alarme falso cai. Isto acontece pois a mudança da média do processo ocorre mais rápido, o que faz com que o período que o processo permanece sob controle seja menor.

O gráfico (6) apresenta a variação da probabilidade de não detecção com a variação do parâmetro  $p$ :



Fonte: Elaboração própria com base nos resultados da pesquisa.

### Gráfico 6 - Sensibilidade da probabilidade de não detecção / parâmetro $p$

Quando o parâmetro  $p$  aumenta, a probabilidade de não detecção aumenta também, pois o período que o processo permanece fora de controle é maior, o que resulta em média mais amostras que os itens estão dentro da região de aceitação dado que o processo está fora de controle.

Perceba que neste caso a mudança do parâmetro  $p$  afeta na distribuição inicial do processo  $v$ , pois esta distribuição depende deste parâmetro.

- Variação do parâmetro  $\varepsilon$ :

A tabela (5) apresenta a variação das probabilidades de alarme falso e de não detecção com a variação parâmetro  $\varepsilon$  da cadeia:

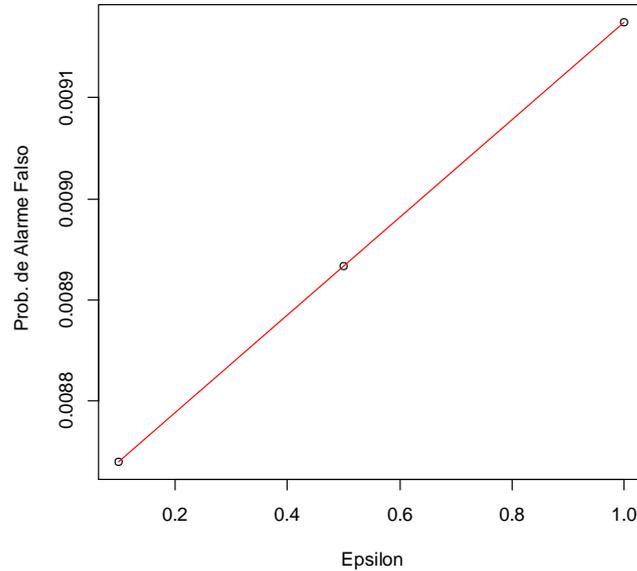
**Tabela 5 - Valores das probabilidades para diferentes valores de parâmetro  $\varepsilon$**

Parâmetro $\varepsilon$	Probabilidades	
	Alarme Falso	Não detecção
0,01	0,0087407	0,9798748
<b>0,5</b>	<b>0,0089334</b>	<b>0,9794312</b>
1	0,0091742	0,9788768

Fonte: Elaboração própria com base nos resultados da pesquisa.

A partir desta tabela, pode-se afirmar que ambas as probabilidades não são sensíveis com a variação do parâmetro  $\varepsilon$ , pois sua variação provoca pequenas mudanças nestas probabilidades.

O gráfico (7) apresenta a variação da probabilidade de alarme falso com a variação do parâmetro  $\varepsilon$ :

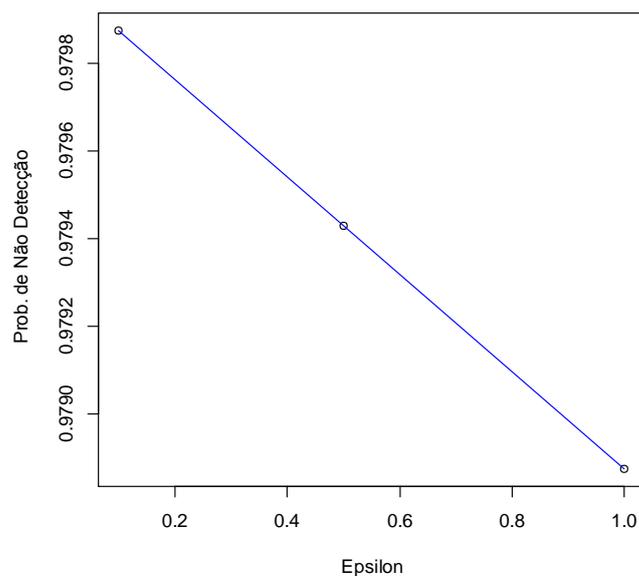


Fonte: Elaboração própria com base nos resultados da pesquisa.

#### **Gráfico 7 - Sensibilidade da probabilidade de alarme falso / parâmetro $\varepsilon$**

Pode-se visualizar a partir do gráfico (7) que há uma dependência linear positiva entre o parâmetro  $\varepsilon$  e a probabilidade de alarme falso, ou seja, à medida que o parâmetro  $\varepsilon$  aumenta, a probabilidade de alarme falso aumenta linearmente. Isto acontece pois o período sob controle do processo será maior, pois a probabilidade do processo retornar ao estado sob controle dado que ele está fora de controle é maior.

O gráfico (8) apresenta a variação da probabilidade de não detecção com a variação do parâmetro  $\varepsilon$ :



Fonte: Elaboração própria com base nos resultados da pesquisa.

### Gráfico 8 - Sensibilidade da probabilidade de não detecção / parâmetro $\epsilon$

Pode-se visualizar a partir do gráfico (8) que há uma dependência linear negativa entre o parâmetro  $\epsilon$  e a probabilidade de não detecção, ou seja, à medida que o parâmetro  $\epsilon$  aumenta, a probabilidade de não detecção diminui linearmente. Isto acontece pois o período fora controle do processo será menor, pois a probabilidade do processo retornar ao estado sob controle dado que ele está fora de controle é maior.

- Variação da magnitude da mudança da média do processo ( $\delta$ ):

A tabela (6) apresenta a variação das probabilidades de alarme falso e de não detecção com a variação do parâmetro  $\delta$ :

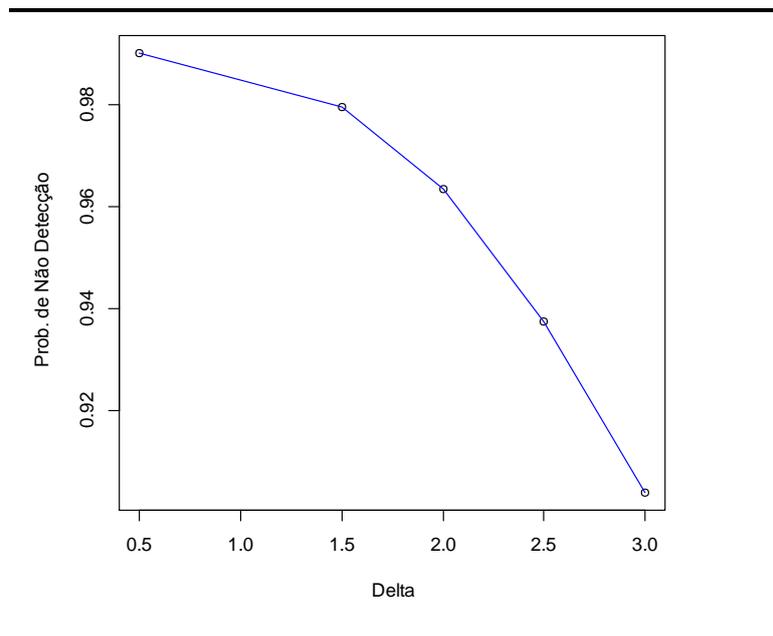
**Tabela 6 - Valores das probabilidades para diferentes valores do parâmetro  $\delta$**

Parâmetro $\delta$	Probabilidades	
	Alarme Falso	Não detecção
0,5	0,0089334	0,9899446
<b>1,5</b>	<b>0,0089334</b>	<b>0,9794312</b>
2	0,0089334	0,9634361
2,5	0,0089334	0,9373335
3	0,0089334	0,9039896

Fonte: Elaboração própria com base nos resultados da pesquisa.

A partir da tabela (6), pode-se notar que a probabilidade do alarme falso não é influenciada pela variação da magnitude da mudança da média do processo, já a probabilidade de não detecção é bastante sensível com a variação do parâmetro  $\delta$ .

O gráfico (9) apresenta a variação da probabilidade de não detecção com a variação do parâmetro  $\delta$ :



Fonte: Elaboração própria com base nos resultados da pesquisa.

### **Gráfico 9 - Sensibilidade da probabilidade de não detecção / parâmetro $\delta$**

A partir do gráfico (9) pode-se afirmar que a probabilidade de não detecção cai quando  $\delta$  aumenta. Isto é pelo fato de quanto maior for a mudança da média do processo, maior será a probabilidade de um item cair fora da região de aceitação dado que o processo está fora de controle.

De acordo com a análise de sensibilidade feita neste capítulo, pode-se afirmar que a probabilidade de alarme falso é bastante sensível às variações dos parâmetros  $k$ ,  $m$  e  $p$ ; contudo, não é sensível às variações do parâmetro  $\varepsilon$ . Vale ressaltar que a probabilidade de alarme falso não depende do parâmetro  $\delta$ .

Já a probabilidade de não detecção é bastante sensível às variações dos parâmetros  $k$ ,  $m$ ,  $p$  e  $\delta$ ; contudo, não é sensível às variações do parâmetro  $\varepsilon$ .

## 6- Conclusão e recomendações futuras

O objetivo inicial deste trabalho foi apresentar a metodologia do modelo oculto de Markov e aplicá-lo no controle estatístico *on-line* para a estimação das probabilidades de alarme falso e não detecção

Além da introdução, no capítulo 2 foi feita uma apresentação de algumas ferramentas do controle estatístico de qualidade para conseguinte, introduzir o problema de controle estatístico *on-line*. No capítulo 3 foi revisada a metodologia de cadeias de Markov e posteriormente apresentada a metodologia do modelo oculto de Markov, sendo principalmente abordados os elementos de um HMM e as soluções dos três problemas básicos.

No capítulo 4 foi apresentada a metodologia proposta por Dorea *et al.* (2012), que é uma aplicação do modelo oculto de Markov no controle estatístico *on-line* para a estimação das probabilidades de alarme falso e a de não detecção. No capítulo 5 foi aplicada a metodologia proposta por Dorea *et al.* (2012) em um exemplo retirado de Ho *et al.* (2007) com o intuito de realizar uma análise de sensibilidade de alguns parâmetros. Foi possível constatar que as probabilidades de alarme falso e de não detecção são influenciadas por diferentes parâmetros em diferentes proporções.

Como extensões para este trabalho é relevante considerar erros de classificação na inspeção dos itens do controle estatístico *on-line* por variáveis via modelo oculto de Markov. Deve-se explorar também o planejamento econômico do controle estatístico *on-line* por variáveis via modelo oculto de Markov.

## Referências Bibliográficas

- Atuncar, G. S. *Conceitos básicos de processos estocásticos*. UFMG. Belo Horizonte, 2011.
- Baker, J. K. The dragon system-An overview. *IEEE Trans. Acoust. Speech Signal Processing*, v. ASS-23, n. 1, p. 24-29, 1975.
- Baum, L. E. e Egon, J. A. An inequality with applications to statistical estimation for probabilistic functions of a Markov process and to a model for ecology. *Bull. Amer. Meterol. soc.*, v. 73, n. 3, p. 360-363, 1967.
- Baum, L. E. e Sell, G. R. Growth functions for transformations on manifolds. *Pac. J. Math*, v. 27, n. 2, p. 211-227, 1968.
- Bessegato, L. F. *Extensão para Controle On-line por atributo com erros de classificação: Intervalo de inspeção variável, Amostragem não-unitária*. 2009. 181f. tese (Doutorado em estatística) – Instituto de Ciências Exatas, UFMG, Belo Horizonte, 2009.
- Borges, W., Ho, L. L. e Turnes, O. An analysis of Taguchi's on-line quality monitoring procedure for attributes with diagnosis errors. *Applied Stochastic Models in Business and Industry*, v. 17, n.3, p. 261-276, 2001.
- Costa, A. F., Epprecht, E. K. e Carpinetti, L. C. *Controle Estatístico de Qualidade*. São Paulo: ATLAS, 2005.
- Couvreur, C. *Hidden Markov Models and Their Mixtures*. 1996. 158f. dissertação (Mestrado em Matemática) – Département de mathématiques, Université catholique de Louvain, Louvain, 1996.
- Crosby, P. B. *Quality without tears: the art of hassle free management*. New York: McGraw-Hill, 1995.
- Deming, W. E. *Out of crisis*. Cambridge: MIT Press, 2000.
- Dempster, A. P., Laird N. M. e Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc*, v. 39, n.1, p. 1-38, 1977.
- Dorea, C. C. Y., Gonçalves, C. R., Medeiros, P. G. E., Santos, W. B. False-Alarm and Non-Detection Probabilities for On-line Quality Control via HMM. *Int. Journal of Math. Analysis*, v. 6, n. 24, p. 1153-1162, 2012.
- Duncan, A. J., The economic design of  $\bar{X}$  charts used to maintain current control of a process. *Journal of American Statistical Association*, v. 51, n. 274, p. 228-242, 1956.
- Forney, G. D. The Viterbi algorithm. *Proc. IEEE*, v. 61, n. 3, p. 268-278, 1973.
- Gibra, I. N. Recent development in control chart techniques. *Journal of Quality Technology*, v.7, n. 4, p. 183-192.

- Girshick, M. e Rubin, H. A Bayes Approach to a Quality Control Model. *Annals of Mathematical Statistics*, v. 23, n. 1, p. 114-125, 1952.
- Ho, L. L., Medeiros, P. G. e Borges, W. Na alternative model for on-line quality monitoring for variables. *International Journal of production economics*, v. 107, n. 1, p. 202-222, 2007.
- Khoury, C. M. B. *Modelos Escondidos de Markov para Classificação de Proteínas*. 2004. 115f. dissertação (Mestrado em Ciência da Computação) - Centro da informática, UFPE, Recife, 2004.
- Ladany, S. P. Optimal use of control charts for controlling current production. *Management Science*, v.19, n. , p. 763-772.
- Leiras, A., Epprecht, E. K., Costa, A. F. B. Planejamento do controle estatístico de processos com baixa fração não conforme restrito a amostras pequenas. *Produção*, v. 17, n. 1, p. 139-150, 2007.
- Lorenzen, T. J., Vance, L. C., The economic design of control charts: A unified approach. *Technometrics*, v. 28, n. 1, p. 3-10.
- Montgomery, D. C. *Introdução ao controle estatístico da qualidade*. Rio de Janeiro: LTC, 2004.
- Montgomery, D. C. The economic design of control charts: A review and literarute survey. *Journal of Quality Technology*, v.12 n. 2 , p. 75-87.
- Nayebpour, M. R. e Woodall, W. H. An analysis of Taguchi's on-line quality-monitoring procedures for attributes. *Technometrics*, v. 35, n. 1, p. 53-60.
- Quinino, R. C. e Ho, L. L. Repetitive tests as na economic alternative procedure to control attributes with diagnosis errors. *European Journal of Operation Research.*, v.155, n. 1, 209-225, 2004.
- Rabiner, L. R. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, v. 77, n. 2, p. 257-286, 1989.
- Ross, S. M. *Introduction to probability models*, 9 edition. Academic Press: San Diego, 2007.
- Saniga, E. M. Economic statistical control-chat designs with an application to  $\bar{X}$  and R charts. *Technometrics*, v. 31, n. 3, p. 313-320, 1989.
- Souza, K. P. *Aplicação de Modelos de Markov Ocultos na Obtenção de Taxas de mortalidade das Larvas do Mosquito da Dengue*. 2010. 78f. dissertação (Mestrado em computação) - Centro de Ciências Exatas e Tecnologia, UFMGS, Campo Grande, 2010.
- Taguchi, G., Elsayed, E. A. e Hsiang, T. *Quality Engineering in Production Systems*. New York: McGraw-Hill, 1989.

Vance, L. C. A bibliography of statistical quality control chart techniques, 1970-1980. *Journal of Quality Technology*, v. 15, n. 2, p. 59-62.

Viterbi, A. J. Error bounds for convolutional codes and an asymptotically optimal decoding algorithm. *IEEE Trans Informat. Theory*, v. IT-13, n. 2, p.260-269, 1967.

Weiler, H. On the Most Economical Sample size for Controlling the Mean of a Population. *Annals of Mathematical Statistics*, v.23, n. 2, p. 247-254, 1952.

Woodall, W. H. Conflicts between Deming's philosophy and the economic design of control charts. In Lenz, H. J., Wetherill, G. B. e Wilrich, P. T., editors, *Frontiers in Statistical Quality Control*. Physica-Verlag, Heidelberg, 1987.

Woodall, W. H. Weaknesses of economic design of control charts (Letter to the Editor). *Technometrics*, v.28, n. 4, p. 408-410, 1986.