# 3
# Fundamentação Teórica de Modelos Bayesianos de Previsão

## 3.1.
## Abordagem Bayesiana para Estimação

A informação que se tem acerca de um parâmetro de interesse $\theta$ é crucial na ciência estatística. Se o valor verdadeiro de $\theta$ não é conhecido então, a idéia é buscar métodos para reduzir o desconhecimento sobre esse parâmetro. Do ponto de vista bayesiano, o grau de incerteza a respeito de $\theta$ é representado através de um modelo probabilístico para $\theta$.

A abordagem clássica para estimação de um parâmetro populacional consiste em tomar uma amostra aleatória de tamanho $n$, proveniente de uma população cuja distribuição de probabilidades da variável aleatória, por exemplo, $Y$ seja função de um parâmetro $\theta$, distribuição essa denotada por $f(y \mid \theta)$ e, posteriormente, estimar o parâmetro $\theta$ através dos valores amostrais $y_i, \mathrm{i} = 1, \cdots, \mathrm{n}$. Esse estimador poderia ser construído através do método da máxima verossimilhança, por exemplo.

A abordagem bayesiana para estimação de um parâmetro populacional usa uma distribuição a priori de $\theta$, $f(\theta)$, e a distribuição de probabilidades conjunta da amostra, $f(y_1, y_2, \ldots, y_n \mid \theta)$, para determinar a distribuição a posteriori de $\theta$, denotada por $f(\theta \mid y_1, y_2, \ldots, y_n)$. Essa distribuição a posteriori para $\theta$ contém informação proveniente da amostra e da distribuição a priori de $\theta$. Sendo assim, o parâmetro $\theta$ é visto como uma variável aleatória na abordagem bayesiana, (MONTGOMERY & RUNGER, 2003, p. 437).

Pole et al. (1994) comentam que a análise bayesiana para a formulação de um modelo começa pela quantificação do conhecimento inicial acerca desse estado inicial. Essas informações a priori são então combinadas com a informação dos dados observados e quantificados probabilisticamente através da função de verossimilhança.

O mecanismo de combinar informações a priori e verossimilhança é o teorema de Bayes. Portanto, pode-se escrever:

$$P(\theta_t \mid D_t) = \frac{P(Y_t = y_t \mid \theta_t) P(\theta_t \mid D_{t-1})}{P(Y_t = y_t)} \tag{12}$$

O denominador $P(Y_t = y_t)$ não é função de $\theta$ e atua como uma constante. Sendo $D_t$ o estado do conhecimento no instante *t,* a verossimilhança $L(\theta_t \mid Y_t = y_t) \propto P(Y_t = y_t \mid \theta_t)$, $P(\theta_t \mid D_t)$ é distribuição a posteriori e $P(\theta_t \mid D_{t-1})$ a distribuição a priori.

O resultado de combinar informações a priori e verossimilhança é a distribuição a posteriori. Então, a distribuição a posteriori é proporcional à distribuição a priori e à verossimilhança: $posteriori \propto priori \times verossimilhança$.

Esse processo de passagem da priori para a posteriori é referida como aprendizado bayesiano, sendo o mecanismo formal através do qual as incertezas são modificadas quando da obtenção de uma nova informação, (POLE et al., 1994, p. 16).

## 3.2. Abordagem Bayesiana para Previsão

No que se refere a modelos de previsão de uma forma geral, Pole et al. (1994, p. 9) declaram que

> Whenever we make a forecast we actually make a statement of probability, or, more generally, state a probability distribution that quantifies the nature of our uncertainty. Any and every forecast is predicated upon a fount of knowledge; forecasts are therefore conditional probability statements, the conditioning being on the existing state of knowledge. If we change our knowledge base, then, typically, our forecast will change.[…] The Bayesian paradigm provides a rational, coherent, formal framework for combining information: routine model forecast (or individual model component quantifications, and more generally components forms too) are adjusted by subjective intervention to reflect so called 'external' information.

A abordagem bayesiana para previsão permite que informações não contidas nos dados históricos sejam incorporadas ao modelo. A possibilidade de se obter informações externas é o centro da inferência bayesiana. Os detalhes dessa abordagem são discutidos no decorrer desse trabalho.

## 3.3.
## Características Fundamentais dos Modelos Bayesianos de Previsão de Harrison & Stevens (HS)

Segundo Souza & Farias Neto (1980), os fundamentos essenciais dos modelos bayesianos de Harrison & Stevens (HS) são:

(i) Formulação Paramétrica (ou de espaço de estados) - Interpretação física dos parâmetros, isto é, as séries são representadas através de componentes que são diretamente interpretáveis, facilitando o entendimento do modelo por parte do analista.

(ii) Informação Probabilística – parâmetros são variáveis aleatórias (com distribuição de probabilidade associada) e não simplesmente quantidades desconhecidas. A abordagem bayesiana explicita não apenas as estimativas dos componentes a cada instante de tempo, mas também indica automaticamente uma medida de incerteza associada a tais estimativas.

(iii) Modelo Seqüencial – formulação através do Modelo Linear Dinâmico (MLD) que descreve como as estimativas dos parâmetros mudam temporalmente.

(iv) Incerteza do Modelo – o modelo adequado em certo instante de tempo pode não ser mais apropriado em um instante seguinte.

Segundo Harrison & Stevens (1976), a incerteza associada ao modelo pode ser de dois tipos distintos:

<u>Classe I</u> – assume-se que o processo pode ser representado adequadamente por um único modelo (desconhecido).

<u>Classe II</u> – supõe-se que o modelo que melhor representa o processo pode mudar com o tempo. Dessa maneira, a qualquer instante de tempo considera-se que o modelo atuante é aquele de maior probabilidade associada.

A Tabela 1 faz uma comparação entre os modelos bayesianos (HS) versus os modelos tradicionais (MT) para a previsão.

Tabela 1 - Comparação dos Modelos Bayesianos (HS) com Modelos Tradicionais (MT)

| **Características** | **Modelos Tradicionais (MT)** | **Modelos Bayesianos (HS)** |
|---|---|---|
| Estacionariedade do modelo | Existe um modelo constante para todo $t = 1, 2, \ldots$ | Pode existir mais de um modelo adequado para todo $t = 1, 2, \ldots$ |
| Informação externa | Utilizam somente os dados históricos. | Utilizam a série histórica e/ou outras informações relevantes. |
| Tamanho da série histórica | Tamanho suficientemente grande. | Tamanho qualquer, HS podem ser utilizados, inicialmente, mesmo na ausência de dados históricos. |
| Estacionariedade da série | O processo gerador dos dados deve ser estacionário e ergódico. | O processo pode ser qualquer. |

## 3.4.
## Modelo Linear Dinâmico – MLD

West & Harrison (1997) mostram que o modelo linear dinâmico normal pode ser escrito de forma geral como:

**Equação das Observações:** $\mathbf{Y}_t = \mathbf{F}_t' \ \boldsymbol{\theta}_t + \boldsymbol{\upsilon}_t, \qquad \boldsymbol{\upsilon}_t \sim N[\mathbf{0}, \mathbf{V}_t]$ (13)

**Equação do Sistema:** $\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_\mathbf{t} \sim \mathbf{N}[\mathbf{0}, \mathbf{W_t}]$ (14)

onde $\mathbf{Y}_t$ denota um vetor (r x1) das observações da série no tempo *t*; $\mathbf{F}_t$ é uma matriz (n x r) de constantes conhecidas – vetor de regressão; $\boldsymbol{\theta}_t$ é o vetor (n x 1) dos parâmetros de estado do modelo – vetor de estado; $\boldsymbol{\upsilon}_t$ é o ruído das observações, vetor (r x1), tendo distribuição normal com média zero e variância $\mathbf{V}_t$, que é uma matriz (r x r); $\mathbf{G}_t$ é uma matriz (n x n) de coeficientes conhecidos que determina a evolução sistemática dos parâmetros no tempo; e $\boldsymbol{\omega}_t$ é o ruído do sistema, vetor (n x 1), tendo distribuição normal com média zero e matriz de covariância $\mathbf{W_t}$, matriz (n x n). Supõe-se que $\boldsymbol{\upsilon}_t$ e $\boldsymbol{\omega}_t$ são temporalmente e mutuamente independentes, ou seja, para todo *t*, *s*, tais que $t \neq s$, $\text{cov}[\boldsymbol{\upsilon}_t, \boldsymbol{\upsilon}_s] = 0$ e $\text{cov}[\boldsymbol{\omega}_t, \boldsymbol{\omega}_s] = 0$; e, para todo *t*, *s*, $\text{cov}[\boldsymbol{\upsilon}_t, \boldsymbol{\omega}_s] = 0$. Outra suposição importante é que $\mathbf{G}_t \equiv \mathbf{G}$, ou seja, o interesse se encontra em MLD´s que possuem uma matriz de coeficientes que não se modifica com o tempo.

A eq. (13) descreve a forma como cada observação é gerada para um dado estado do sistema no instante *t,* é chamada equação das observações. A eq. (14), por conseguinte, descreve o modelo dinâmico, ou seja, a evolução paramétrica entre o instante *t-1* e o instante *t*, e é chamada equação do sistema.

O problema de estimação do MLD refere-se a encontrar estimativas para o vetor de estados $\boldsymbol{\theta}_t$, a cada instante t, com base nas observações disponíveis até aquele instante, inclusive.

### 3.4.1. Algoritmo do MLD com variância $V_t$ conhecida

Em muitas situações é mais simples utilizar o $Y_t$ (caso escalar) no lugar do vetor de observações $\mathbf{Y}_t$, esse procedimento caracteriza o MLD como univariado. West & Harrison (1997) definem o MLD univariado geral com variância $V_t$ conhecida, como:

Equação das Observações: $Y_t = \mathbf{F}_t' \; \boldsymbol{\theta}_t + \upsilon_t, \qquad \upsilon_t \sim N[0, V_t]$ (15)

Equação do Sistema: $\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N[\mathbf{0}, \mathbf{W}_t]$ (16)

Especificações iniciais do usuário:

$(\boldsymbol{\theta}_0 \mid D_0) \sim N(\mathbf{m}_0, \mathbf{C}_0)$ e $\{\mathbf{F}_t, \mathbf{G}_t, \mathbf{V}_t \text{ e } \mathbf{W}_t\}$

$D_0 \rightarrow$ estado do conhecimento no instante inicial.

$D_t \rightarrow$ estado do conhecimento no instante *t*.

Informação:

$(\boldsymbol{\theta}_{t-1} \mid D_{t-1}) \sim N[\mathbf{m}_{t-1}, \mathbf{C}_{t-1}] \rightarrow$ **posteriori** no instante *t-1*.

$(\boldsymbol{\theta}_t \mid D_{t-1}) \sim N[\mathbf{a}_t, \mathbf{R}_t] \rightarrow$ **priori** no instante *t*.

$$\mathbf{a}_t = \mathbf{G}_t \mathbf{m}_{t-1} \tag{17}$$

$$\mathbf{R}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t' + \mathbf{W}_t \tag{18}$$

Previsão no instante *t*:

$(Y_t \mid D_{t-1}) \sim N[f_t, Q_t]$

$$f_t = \mathbf{F}_t' \; \mathbf{a}_t \tag{19}$$

$$Q_t = \mathbf{F}_t' \; \mathbf{R}_t \mathbf{F}_t + V_t \tag{20}$$

Atualização:

$(\boldsymbol{\theta}_t \mid D_t) \sim N[\mathbf{m}_t, \mathbf{C}_t] \rightarrow$ **posteriori** no instante *t*

$$\mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t e_t \tag{21}$$

$$\mathbf{C}_t = \mathbf{R}_t - \mathbf{A}_t Q_t \mathbf{A}'_t \tag{22}$$

$$e_t = Y_t - f_t \tag{23}$$

$$\mathbf{A}_t = \mathbf{R}_t \mathbf{F}_t Q_t^{-1} \tag{24}$$

Previsões na origem $t$, $k$ passos à frente ($k \geq 1$):

$$(\boldsymbol{\theta}_{t+k} \mid D_t) \sim N[\mathbf{a}_t(k), \mathbf{R}_t(k)]$$

$$(Y_{t+k} \mid D_t) \sim N[f_t(k), Q_t(k)]$$

$$\mathbf{a}_t(k) = \mathbf{G}_{t+k}\mathbf{a}_t(k-1) \tag{25}$$

$$\mathbf{R}_t(k) = \mathbf{G}_{t+k}\mathbf{R}_t(k-1)\mathbf{G}'_{t+k} + \mathbf{W}_{t+k} \tag{26}$$

$$f_t(k) = \mathbf{F}'_{t+k}\mathbf{a}_t(k) \tag{27}$$

$$Q_t(k) = \mathbf{F}'_{t+k}\mathbf{R}_t(k)\mathbf{F}_{t+k} + V_{t+k} \tag{28}$$

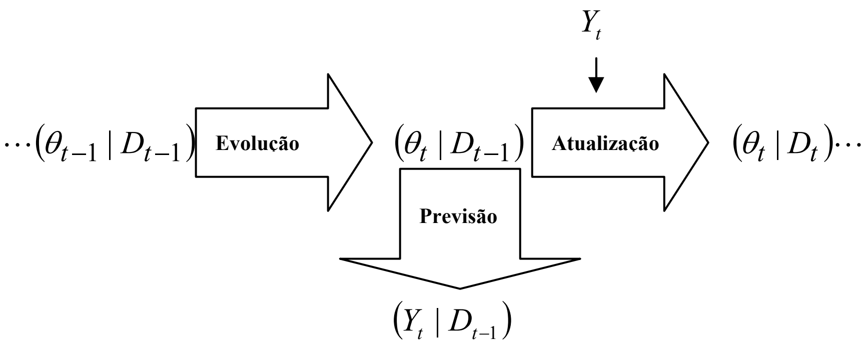
$$\mathbf{a}_t(0) = \mathbf{m}_t \tag{29}$$

$$\mathbf{R}_t(0) = \mathbf{C}_t \tag{30}$$

Como pôde ser observado, o processo de previsão envolve a estimação de três distribuições condicionais:

1. Distribuição a priori de $\theta_t$, denotada por $(\theta_t \mid D_{t-1})$: representa a melhor idéia acerca do estado do sistema após observar $\langle Y_1, \cdots, Y_{t-1} \rangle$, porém antes de observar $Y_t$.
2. Distribuição da previsão no instante $t$, $(Y_t \mid D_{t-1})$: é derivada da distribuição a priori $(\theta_t \mid D_{t-1})$.
3. Distribuição a posteriori de $\theta_t$, $(\theta_t \mid D_t)$: esta resulta da revisão da distribuição a priori com a inclusão da observação $Y_t$, utilizando o Teorema de Bayes.

De forma esquemática:



## 3.4.2.
## MLD - Modelo Estático de HS

É o modelo mais simples dentro da classe de Modelos Bayesianos de HS, correspondendo ao problema clássico de estimação seqüencial da média $\mu$ de uma população normal. Na formulação MLD, tem-se que:

$$\theta_t = \mu; \quad \mathbf{F}_t = 1; \quad \mathbf{G}_t \equiv \mathbf{G} = 1; \quad \upsilon_t \sim N[0, V]; \quad \boldsymbol{\omega}_t = 0 \qquad (31)$$

A formulação matemática para o modelo estático (Figura 5) se resume a:

**Equação das Observações:** $Y_t = \theta_t + \upsilon_t; \qquad \upsilon_t \sim N[0, V]$ (32)

**Equação do Sistema:** $\theta_t = \text{constante} = \mu$ (33)

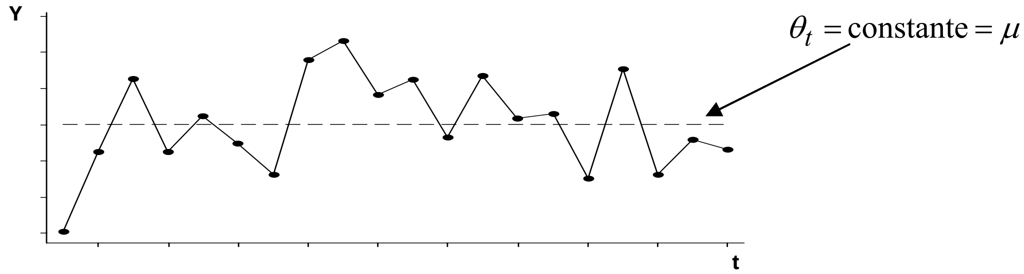



Figura 5 - um processo com comportamento do modelo estático.

## 3.4.3.
## MLD - Modelo Estacionário de HS

Esse modelo descreve a evolução dinâmica do nível, segundo um passeio aleatório. Nesse caso:

$$\theta_t = \mu_t; \quad \mathbf{F}_t = 1; \quad \mathbf{G}_t \equiv \mathbf{G} = 1; \quad \upsilon_t \sim N[0, V]; \quad \omega_t \sim N[0, W] \qquad (34)$$

A formulação matemática para o modelo estacionário (Figura 6) então se resume a:

**Equação das Observações:** $Y_t = \theta_t + \upsilon_t; \qquad \upsilon_t \sim N[0, V]$ (35)

**Equação do Sistema:** $\theta_t = \theta_{t-1} + \omega_t; \qquad \omega_t \sim N[0, W]$ (36)

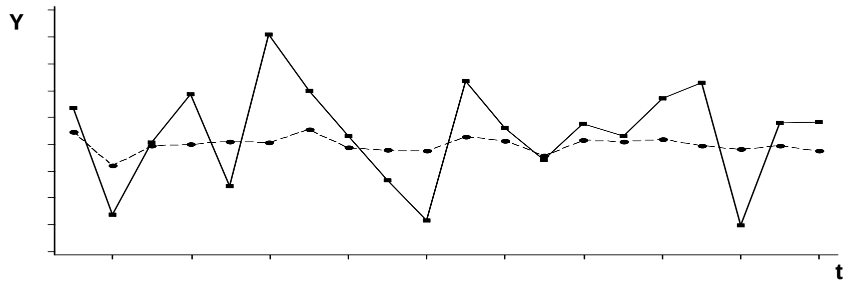



Figura 6 - um processo com comportamento do modelo estacionário.

### 3.4.4.
### MLD – Modelo de Crescimento Linear (MCL) de HS

No modelo de crescimento linear (MCL), o processo é composto de um nível e de uma inclinação que variam no tempo segundo as equações:

**Equação das Observações:** $Y_t = \mu_t + \varepsilon_t; \qquad \varepsilon_t \sim N[0, V_t]$ (37)

**Equações do Sistema:**

$$\mu_t = \mu_{t-1} + \beta_t + \delta_{\mu_t}; \quad \delta_{\mu_t} \sim N[0, V_\mu] \tag{38}$$

$$\beta_t = \beta_{t-1} + \delta_{\beta_t}; \qquad \delta_{\beta_t} \sim N[0, V_\beta] \tag{39}$$

onde,

$\mu_t$ - Nível no instante *t*.

$\beta_t$ - Inclinação no instante *t*.

$\varepsilon_t$ - Ruído das observações no instante *t*.

$\delta_{\mu_t}$ - Ruído do nível no instante *t*.

$\delta_{\beta_t}$ - Ruído da inclinação no instante *t*.

O vetor de estado compreende dois elementos, $\theta_t = [\mu_t \quad \beta_t]'$: o primeiro representa o nível e o segundo a inclinação. Na formulação MLD, pode-se mostrar que:

$$\mathbf{F}_t = \begin{bmatrix} 1 \\ 0 \end{bmatrix}; \qquad \mathbf{G}_t \equiv \mathbf{G} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}; \qquad \upsilon_t = \varepsilon_t; \qquad \boldsymbol{\omega}_t = \begin{bmatrix} \delta_{\mu_t} + \delta_{\beta_t} \\ \delta_{\beta_t} \end{bmatrix} \tag{40}$$

O modelo estático e o modelo estacionário são casos particulares do modelo de crescimento linear. Se $\beta_t = 0$ e $\delta_{\beta_t} = 0$ para todo *t*, então fica caracterizado o modelo estacionário anteriormente citado.

Quando $\beta_t = 0$ , $\delta_{\beta_t} = 0$ , $\delta_{\mu_t} = 0$ e $\mu_t = \mu_{t-1}$ para todo $t$, então o modelo estático entra em uso.

Farias Neto (1981, p. 47-50) mostra que a matriz de covariância pode ser dada por:

$$\mathbf{W}_t = \begin{bmatrix} V_\mu + V_\beta & V_\beta \\ V_\beta & V_\beta \end{bmatrix} \tag{41}$$

Existem outros modelos, como o MCL sazonal, mas que não seriam úteis no contexto desse trabalho, ou seja, CEP para dados autocorrelacionados. Maiores detalhes da operacionalização de um modelo bayesiano geral de HS podem ser encontrados em Pole et al. (1994, p. 43 e p.63).

### 3.4.5.
### Fatores de Desconto para $\mathbf{W}_t$

Ameen e Harrison (1985) introduziram o conceito de *fatores de desconto* com intuito de substituir a matriz de variância $\mathbf{W}_t$ do MCL-HS. Citam que os fatores de desconto proporcionam simplicidade ao modelo.

O fator de desconto controla o grau de “envelhecimento” do conteúdo informativo de uma observação. O uso de fatores de desconto baseia-se na idéia de descontar a informação contida nas observações mais antigas; quanto maior o fator de desconto, menor será a perda de informações com o passar do tempo (o sistema é mais conservador). A recíproca também é verdadeira, quanto menor o fator desconto, menor será a relevância das observações antigas.

Utilizando a notação adotada por Baratojo (1989), $\beta_x$ denota os fatores de desconto, onde $x$ é a especificação da componente não observável. O valor de $\beta_x$ é definido entre $0 < \beta_x \leq 1$, sendo que valores próximos de 1 fazem com que o sistema seja menos sensível a novas observações (conservador). Já valores próximos de 0 fazem com que o sistema seja mais sensível. Tabela 2 resume o uso de fatores de desconto.

Tabela 2 - Fatores de Desconto

| Fatores de Desconto $\beta_x$ | Variância do Sistema | Confiança nas Observações Antigas | Sensibilidade do Modelo |
|---|---|---|---|
| Próximos de 1 | Pequena | Alta | Baixa |
| Próximos de 0 | Grande | Baixa | Alta |

Fonte: adaptado de Baratojo (1989)

Como pode ser visto em Pole et al. (1994, p.52), a matriz $\mathbf{W}_t$ pode ser calculada através da formulação de fatores de desconto como:

$$\mathbf{W}_t = \left(\beta_x^{-1} - 1\right)\mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}_t' \tag{42}$$

$$\mathbf{R}_t = \beta_x^{-1}\mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}_t' \tag{43}$$

Desta forma, a variância da priori no instante $t$ ($R_t$) é calculada como função da variância da posteriori ($C_{t-1}$) no instante *t-1*, de acordo com o fator de desconto $\beta_x$.

Em modelos com múltiplos componentes, Pole et al. (1994) recomendam que o procedimento seja feito componente por componente diferentes. Fatores de desconto são especificados para cada componente e a evolução de componentes individuais da matriz de variância é calculada. Então na forma matricial tem-se que:

$$\mathbf{R}_t = \boldsymbol{\beta}\mathbf{G}\mathbf{C}_{t-1}\mathbf{G}'\boldsymbol{\beta} \tag{44}$$

$$\boldsymbol{\beta} = Diag\left[\left(\frac{1}{\sqrt{\beta_1}}\right)\mathbf{I}_1, \left(\frac{1}{\sqrt{\beta_2}}\right)\mathbf{I}_2, \cdots, \left(\frac{1}{\sqrt{\beta_k}}\right)\mathbf{I}_k\right]$$

$\mathbf{I}_i \rightarrow$ Matriz identidade de ordem $n_i$ X $n_i$.

O uso de $\boldsymbol{\beta}$ acima está associado à possibilidade de se escrever a matriz de transição $\mathbf{G}$ estruturada em blocos, um para cada componente do sistema, ou seja, $\mathbf{G} = Diag.[\mathbf{G}_1, \mathbf{G}_2, \cdots, \mathbf{G}_k]$.

Onde $\mathbf{G}_i$ = bloco matricial, de ordem $n_i$ X $n_i$ correspondente ao i-ésimo componente.

Para exemplificar, seja o modelo com duas componentes, tendência (T) e sazonalidade (S). Definindo $\beta_T$ e $\beta_S$ como sendo os fatores de desconto para os respectivos componentes, então:

$$\mathbf{W}_t = \begin{bmatrix} \mathbf{W}_T & 0 \\ 0 & \mathbf{W}_S \end{bmatrix}_t$$

$$\mathbf{W}_t = \begin{bmatrix} \left(\beta_T^{-1} - 1\right)\mathbf{G}_T \mathbf{C}_{T\,t-1} \mathbf{G}_T' & 0 \\ 0 & \left(\beta_S^{-1} - 1\right)\mathbf{G}_S \mathbf{C}_{S\,t-1} \mathbf{G}_S' \end{bmatrix}$$

### 3.4.6. Algoritmo do MLD com variância $V_t$ Desconhecida

Novamente, utilizando o procedimento descrito por Pole et al. (1994, p. 63) e sua notação, para MLD univariado e variância $V_t = k_t \phi^{-1}$ desconhecida e $k_t$ conhecido, tem-se que:

Equação das Observações: $Y_t = \mathbf{F}_t' \ \boldsymbol{\theta}_t + \upsilon_t, \qquad \upsilon_t \sim N\left[0, k_t \phi_t^{-1}\right]$ (45)

Equação do Sistema: $\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim t_{n_{t-1}}\left[\mathbf{0}, \mathbf{W}_t\right]$ (46)

Informação:

$\left(\boldsymbol{\theta}_{t-1} \mid D_{t-1}\right) \sim t_{n_{t-1}}\left[\mathbf{m}_{t-1}, \mathbf{C}_{t-1}\right] \rightarrow$ **posteriori** no instante *t-1*.

$\left(\boldsymbol{\theta}_t \mid D_{t-1}\right) \sim t_{\delta_t n_{t-1}}\left[\mathbf{a}_t, \mathbf{R}_t\right] \rightarrow$ **priori** no instante *t*.

$$\left(\phi_{t-1} \mid D_{t-1}\right) \sim G\left[\frac{n_{t-1}}{2}, \frac{d_{t-1}}{2}\right]$$

$$\left(\phi_t \mid D_{t-1}\right) \sim G\left[\frac{\delta_t n_{t-1}}{2}, \frac{\delta_t d_{t-1}}{2}\right]$$

$$\mathbf{a}_t = \mathbf{G}_t \mathbf{m}_{t-1} \quad (47)$$

$$\mathbf{R}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t' + \mathbf{W}_t \quad (48)$$

Previsão no instante *t*:

$$\left(Y_t \mid D_{t-1}\right) \sim t_{\delta_t n_{t-1}}\left[f_t, Q_t\right]$$

$$f_t = \mathbf{F}_t' \ \mathbf{a}_t \quad (49)$$

$$Q_t = \mathbf{F}_t' \ \mathbf{R}_t \mathbf{F}_t + k_t S_{t-1} \quad (50)$$

Atualização:

$\left(\theta_t \mid D_t\right) \sim t_{n_t}\left[m_t, C_t\right] \rightarrow$ **posteriori** no instante *t*

$$\left(\phi_t \mid D_t\right) \sim G\left[\frac{n_t}{2}, \frac{d_t}{2}\right]$$

$$\mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t e_t \tag{51}$$

$$\mathbf{C}_t = \left(\frac{S_t}{S_{t-1}}\right)\left[\mathbf{R}_t - \mathbf{A}_t\mathbf{A}_t'\, Q_t\right] \tag{52}$$

$$e_t = Y_t - f_t \tag{53}$$

$$\mathbf{A}_t = \frac{\mathbf{R}_t\mathbf{F}_t}{Q_t} \tag{54}$$

$$n_t = \delta_t n_{t-1} + 1$$

$$d_t = \delta_t d_{t-1} + \frac{S_{t-1}e_t^2}{Q_t}$$

$$S_t = \frac{d_t}{n_t}$$

Previsões na origem *t*, *k* passos à frente ($k \geq 1$):

$$(\boldsymbol{\theta}_{t+k} \mid D_t) \sim t_{\delta_t n_t}\left[\mathbf{a}_t(k), \mathbf{R}_t(k)\right]$$

$$(Y_{t+k} \mid D_t) \sim t_{\delta_t n_t}\left[f_t(k), Q_t(k)\right]$$

$$\mathbf{a}_t(k) = \mathbf{G}_{t+k}\mathbf{a}_t(k-1) \tag{55}$$

$$\mathbf{R}_t(k) = \mathbf{G}_{t+k}\mathbf{R}_t(k-1)\mathbf{G}_{t+k}' + \mathbf{W}_{t+k} \tag{56}$$

$$f_t(k) = \mathbf{F}_{t+k}'\mathbf{a}_t(k) \tag{57}$$

$$Q_t(k) = \mathbf{F}_{t+k}'\mathbf{R}_t(k)\mathbf{F}_{t+k} + k_{t+k}S_t \tag{58}$$

$$\mathbf{a}_t(0) = \mathbf{m}_t \tag{59}$$

$$\mathbf{R}_t(0) = \mathbf{C}_t \tag{60}$$

Desta forma, em qualquer instante de tempo o MLD pode ser caracterizado por:

$$M_t = \left\{\mathbf{F}_t; \mathbf{G}; V_t\ ; \mathbf{W}_t\right\} \tag{61}$$

O vetor $M_t$ é conhecido como vetor de caracterização do sistema. O vetor $\mathbf{F}_t$ é sempre conhecido a priori, assim como a matriz $\mathbf{G}$ também será conhecida. Já os elementos de $V_t$ e $\mathbf{W}_t$ são todos desconhecidos e terão de ser especificados a priori. Para $V_t$, usualmente considera-se a lei de variância para modelar o aspecto de que a variância do componente irregular tende a variar de acordo com o nível das séries. No caso da matriz $\mathbf{W}_t$, faz-se uso de fatores de desconto para caracterizar sua evolução temporal.

Além disso, antes de se iniciar o processo de estimação, também é necessário estipular alguns parâmetros:

- as condições iniciais para os componentes do modelo $(m_0)$ e suas respectivas variâncias $(R_0)$, ou seja, os parâmetros da distribuição a priori inicial;
- os fatores de desconto $\beta_x$ para cada componente e $\delta$ para a lei de variância;
- o expoente da lei de variância $i$.

quando não se tem informação sobre a distribuição a priori inicial, inicializa-se o filtro adotando-se uma "priori de referência", fazendo $m_0 = 0$ e $\mathbf{R}_0 = h.\mathbf{I}$, onde h é um número grande e $\mathbf{I}$ é a matriz identidade. Isto equivale a supor ignorância total a priori quanto ao vetor de estados $\boldsymbol{\theta}_t$.



## 3.5. Fator de Bayes

Segundo Pole et al. (1994) e Souza (2003), a ferramenta de avaliação de um MLD é o fator de Bayes, que lida com a razão da verossimilhança preditiva de dois modelos: o corrente e o alternativo.

Sendo:

$M$ : modelo corrente; preditiva $\Rightarrow p(y_t \mid D_{t-1}; M)$.

$M_A$: modelo alternativo; preditiva $\Rightarrow p(y_t \mid D_{t-1}; M_A)$.

Então o fator de Bayes é definido por:

$$H_t = \frac{p(y_t \mid D_{t-1}; M)}{p(y_t \mid D_{t-1}; M_A)} \tag{62}$$

Dessa forma, o modelo corrente será adequado quando $H_t >> 1$. Pole et al. (1994) e Souza (2003) sugerem que:

$H_t \geq 10 \Rightarrow$ indica evidência em favor de modelo corrente.

$H_t \geq 100 \Rightarrow$ indica forte evidência em favor de modelo corrente.

### 3.5.1. Fator de Bayes Acumulado

O fator de Bayes $H_t$ mede a evidência de um dado modelo em uma única observação no tempo t com relação a outro modelo. Observar uma seqüência de observações é uma outra forma de avaliar o desempenho da previsão segundo Pole et al. (1994). O fator de Bayes acumulado para uma seqüência de observações $Y_t, Y_{t-1}, \cdots, Y_{t-k+1}$ é definido como:

$$H_t(k) = \prod_{i=0}^{k-1} H_{t-i} \qquad (63)$$
$$H_t(k) = H_t \cdot H_{t-1}(k-1)$$

Souza (2003) afirma que o fator de Bayes acumulado pode ser utilizado para detectar, rapidamente, variações no comportamento do processo provocadas por descontinuidades dos tipos transientes ou mudanças estruturais.

### 3.5.2. Algoritmo para identificação de Descontinuidades

Passo 1 – Calcular seqüencialmente as quantidades $L_t$ e $l_t$:

$$L_t = H_t \cdot \min(1, L_{t-1})$$
$$l_t = \begin{cases} 1 + l_{t-1} & \Rightarrow \ L_{t-1} < 1 \\ 1 & \Rightarrow \ L_{t-1} \geq 1 \end{cases}$$

$l_t$ é conhecido como comprimento da descontinuidade e $L_t$ como fator de Bayes acumulado. Para efeito de inicialização, Pole et al. (1994) sugerem utilizar $L_0 = 1$ e $l_0 = 1$.

Passo 2 – Regra de decisão

Baratojo (1989) e Souza (2003) afrimam que se $L_t \geq \tau$, onde $\tau$ é um real positivo pré-estabelecido, o sistema será considerado em controle, ou seja, não é observada descontinuidade no instante *t*. Caso se $L_t < \tau$ no instante t, então:

(i) $l_t = 1 \ \rightarrow$ a descontinuidade detectada em *t* é do tipo transiente.

(ii) $l_t > 1 \ \rightarrow$ a descontinuidade detectada em *t* é do tipo mudança estrutural e ocorreu provavelmente a partir de $t - l_t + 1$.