INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
UNIVERSIDADE DE SÃO PAULO

Modelos Dinâmicos Bayesianos em Marketing: Uma Revisão

Baseado em Migon, Alves, Menezes e Pinheiro (2023)

Alex Monito Nhancololo[1]

Seminários - MAE5748
01/12/2025

[1]amnhancololo@ime.usp.br || amnhancololo@gmail.com

Lab250

CONTEXTUALIZAÇÃO

$\Longrightarrow$ **Modelos Lineares Dinâmicos (DLM)**

- Criados por Harrison e Stevens em 1976
- Subutilizados no Marketing, mas extremamente poderosos

CONTEXTUALIZAÇÃO

$\Longrightarrow$ **Modelos Lineares Dinâmicos (DLM)**

- Criados por Harrison e Stevens em 1976
- Subutilizados no Marketing, mas extremamente poderosos

$\Longrightarrow$ **Modelos L. Dinâmicos vs Modelos Lineares Generalizados (GLM)**

- GLM os parâmetros $\boldsymbol{\theta}_{p \times 1} = [\alpha, \beta]^\top$ são fixos $\forall t \in \mathbb{R}$
- DLM $\boldsymbol{\theta}_{p \times 1}$ não são fixos $\forall t \in \mathbb{R}$
- GLM não se adapta aos "regime changes (RC)". **Ex. de RC:** Mudança repentina do comportamento do consumidor
- DLM se adapta rapidamente aos "regime changes"

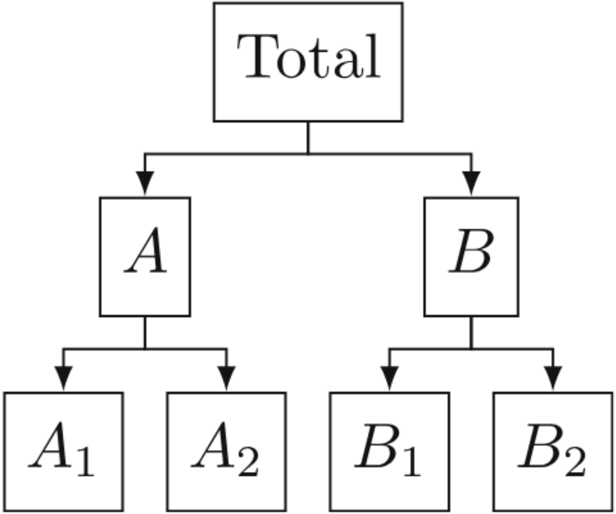
CONTEXTUALIZAÇÃO

$\Longrightarrow$ Modelos L. Dinâmicos vs Modelos Lineares Generalizados (GLM) em Big Data

- Múltiplos produtos, lojas e escalas temporais.
  - Em GLM, rodar modelos complexos (como VAR ou MCMC tradicional) é computacionalmente inviável e lento.
  - Em DLM flexível via Desacoplar/Reacoplar

CONTEXTUALIZAÇÃO

$\Longrightarrow$ DLM vs GLM em Big Data



CONTEXTUALIZAÇÃO

$\Longrightarrow$ **Regressão Linear**

$$
\begin{aligned}
y_i &= \alpha + \beta x_i + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2) \\
&\Downarrow \\
y_i &= \begin{bmatrix} 1 & x_i \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \varepsilon_i
\end{aligned} \tag{1}
$$

$$
y_i = \mathbf{X}_i^\top \boldsymbol{\theta} + \varepsilon_i, \quad \mathbf{X}_i^\top = \begin{bmatrix} 1 & x_i \end{bmatrix}, \quad \boldsymbol{\theta} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \tag{2}
$$

CONTEXTUALIZAÇÃO

$\Longrightarrow$ **Regressão Linear**

$$\begin{aligned} y_i &= \alpha + \beta x_i + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2) \\ &\Downarrow \\ y_i &= \begin{bmatrix} 1 & x_i \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \varepsilon_i \end{aligned} \tag{1}$$

$$y_i = \mathbf{X}_i^\top \boldsymbol{\theta} + \varepsilon_i, \quad \mathbf{X}_i^\top = \begin{bmatrix} 1 & x_i \end{bmatrix}, \quad \boldsymbol{\theta} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \tag{2}$$

$\Longrightarrow$ **Transição para Modelo Linear Dinâmico**

$$y_t = \mathbf{X}_t^\top \boldsymbol{\theta}_t + \varepsilon_t \tag{3}$$

Problema: y_t é escalar a cada instante t e só podemos estimar um parâmetro $\boldsymbol{\theta}, \forall t \in \mathbb{R}$.

Lab250

Modelos Lineares Dinâmicos

$\Longrightarrow$ **Solução do problema:** Impor uma restrição para que os parâmetros de regressão no tempo t dependam do tempo $t-1$ (Prop. Markov).

$$y_t = \mathbf{X}_t^\top \boldsymbol{\theta}_t + \varepsilon_t, \;\; \varepsilon_t \sim N_m\left[0, V_t\right]$$
$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N_p(\mathbf{0}, \mathbf{W}_t)$$

Modelos Lineares Dinâmicos

$\Longrightarrow$ **Solução do problema:** Impor uma restrição para que os parâmetros de regressão no tempo t dependam do tempo $t-1$ (Prop. Markov).

$$
\begin{aligned}
y_t &= \mathbf{X}_t^\top \boldsymbol{\theta}_t + \varepsilon_t, \quad \varepsilon_t \sim N_m\left[0, V_t\right] \\
\boldsymbol{\theta}_t &= \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N_p(\mathbf{0}, \mathbf{W}_t) \\
&\Downarrow \\
\mathbf{y}_t &= \mathbf{F}_t^\top \boldsymbol{\theta}_t + \boldsymbol{\varepsilon}_t \quad \boldsymbol{\varepsilon}_t \sim N_m\left[\mathbf{0}, \mathbf{V}_t\right] \quad \text{(Equação das observações)} \\
\boldsymbol{\theta}_t &= \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N_p(\mathbf{0}, \mathbf{W}_t), \quad \text{(Equação dos estados)}
\end{aligned}
$$

DLM$\Longrightarrow \{\boldsymbol{F}_t, \boldsymbol{G}_t, \mathbf{V}_t, \mathbf{W}_t\}(4)$

- $\{\boldsymbol{y}_t\}_{m\times 1}$, $\{\boldsymbol{\omega}_t\}_{p\times 1}$, $\{\mathbf{F}_t\}_{m\times n}$, $\boldsymbol{G}_t = \boldsymbol{G}_{p\times p}$, $\{\mathbf{W}_t\}_{p\times 1}$, $\{\mathbf{V}_t\}_{m\times 1}$, $\boldsymbol{\theta}_{p\times 1}$
- $\varepsilon_t \perp \varepsilon_{t'}$, $\omega_t \perp \omega_{t'}$, $\varepsilon_t \perp \omega_t \perp D_t$
- $D_t = \{D_{t-1}, y_1, \dots, y_n, I_t\}$, informação disponível até tempo t
- I_t: informação extra aos dados, introduzida manualmente pelo analista para corrigir algo (Intervenção).

Lab250

Exemplos de Modelos Lineares Dinâmicos ($\{\boldsymbol{F_t}, \boldsymbol{G_t}, \boldsymbol{V_t}, \boldsymbol{W_t}\}$)

- Modelo polinomial de 1ª ordem $\{1, 1, V_t, W_t\}$

$$\begin{aligned} y_t &= \theta_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, V_t) \\ \theta_t &= \theta_{t-1} + \psi_t, \quad \psi_t \sim N(0, W_t) \end{aligned} \tag{5}$$

$\theta_t = \mu_t \Longrightarrow$ Tendência do processo; $\boldsymbol{F}_t = \boldsymbol{G}_t = 1$

Obs: **Ruim** (a longo prazo) e **Bom** (a curto prazo)
Motivo: Seja $f_t(h) = \mathbb{E}[y_{t+h}|D_t] = m_t, \quad \forall h \geq 1$

Curto Prazo ($h = 1$)

- **Adaptação rápida:** se $\psi_t^* \uparrow \Longrightarrow \ m_t \to y_t$
- **Ex:**
  $y_{t-1} = 80 \xrightarrow{\psi_t} y_t = 100 \ \Longrightarrow \ f_t(1) \approx 100$
- **Inferência:** $\hat{y}_{t+1} \approx y_t$

Longo Prazo ($h \gg 1$)

- **Estacionaridade Projetada:**
  $\frac{\partial f_t(h)}{\partial h} = 0, \ \Longrightarrow \ f_t(h) = m_t \quad (\forall h)$
- **Viés:** Se $\exists \beta \neq 0$ (Tendência):
  $\mathbb{E}[y_{t+h} - f_t(h)] \propto \beta \cdot h$
- **Consequência:** $\lim_{h\to\infty} \text{MSE}(h) \to \infty$

Se variância de ψ_t (que é W_t) $\uparrow$, o Ganho de Kalman $(K_t) \to 1$:
$m_t = m_{t-1} + K_t(y_t - f_{t-1}) \approx y_t$

Lab250

Exemplos de Modelos Lineares Dinâmicos ($\{F_t, G_t, V_t, W_t\}$)

- Modelo polinomial de 2ª ordem $\{F_t, G_t, 1, V_t, W_t\}$

$$
\begin{gathered}
y_t = \theta_{1,t} + \varepsilon_t, \quad \varepsilon_t \sim N(0, V_t) \Longrightarrow \boldsymbol{y}_t = \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{\boldsymbol{F}_t^\top} \times \underbrace{\begin{bmatrix} \theta_{1,t} \\ \theta_{2,t} \end{bmatrix}}_{\boldsymbol{\theta}_t} + \varepsilon_t \sim N(0, V_t) \\
\vdots \\
\begin{cases} \theta_{1,t} = \theta_{1,t-1} + \theta_{2,t-1} + \psi_{1,t} \\ \theta_{2,t} = \theta_{2,t-1} + \psi_{2,t} \end{cases} \Longrightarrow \underbrace{\begin{bmatrix} \theta_{1,t} \\ \theta_{2,t} \end{bmatrix}}_{\boldsymbol{\theta}_t} = \underbrace{\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}}_{\boldsymbol{G}_t = \boldsymbol{G}} \times \underbrace{\begin{bmatrix} \theta_{1,t-1} \\ \theta_{2,t-1} \end{bmatrix}}_{\boldsymbol{\theta}_t} + \underbrace{\begin{bmatrix} \psi_{1,t} \\ \psi_{2,t} \end{bmatrix}}_{\boldsymbol{\psi}_t}, \quad \boldsymbol{\psi}_t \sim N(0, \boldsymbol{W}_t)
\end{gathered} \tag{6}
$$

Onde:
$\theta_{1,t}$ é por ex: nível de vendas na loja Atacadão, no dia t (ex: $\theta_{1,t=3} = 3000$);
$\theta_{1,t}$ é por ex: crescimento das vendas na loja Atacadão até o dia t (ex: $\{\theta_{1,t_i}\}_{i=1}^{3} = \{1000, 2000, 3000\}$, cres= 1000)

Lab250

Modelo Sazonal (Fourier)

- Para a componente de frequência $k\omega$ (com $\omega = 2\pi/p$):

$$
\begin{aligned}
y_t = \boldsymbol{F}_t^\top \boldsymbol{\theta}_t + \nu_t &\Longrightarrow y_t = \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{\boldsymbol{F}_t} \times \underbrace{\begin{bmatrix} S_{1,t} \\ S_{2,t} \end{bmatrix}}_{\boldsymbol{\theta}_t} + \nu_t \\
&\vdots \\
\boldsymbol{\theta}_t = \boldsymbol{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t &\Longrightarrow \underbrace{\begin{bmatrix} S_{1,t} \\ S_{2,t} \end{bmatrix}}_{\boldsymbol{\theta}_t} = \underbrace{\begin{bmatrix} \cos(k\omega) & \sin(k\omega) \\ -\sin(k\omega) & \cos(k\omega) \end{bmatrix}}_{\boldsymbol{G}_t (\text{Rotação})} \times \underbrace{\begin{bmatrix} S_{1,t-1} \\ S_{2,t-1} \end{bmatrix}}_{\boldsymbol{\theta}_{t-1}} + \boldsymbol{\omega}_t
\end{aligned}
\tag{7}
$$

Onde:
ω: Frequência angular. Se dados mensais ($p = 12$), $\omega = 2\pi/12$;
$\boldsymbol{G}_t$: Matriz de Rotação.

Modelo de Regressão Dinâmica

- Relação entre resposta y_t e covariável x_t com parâmetros variando no tempo.

$$\begin{aligned}
y_t = \alpha_t + \beta_t x_t + \nu_t &\Longrightarrow y_t = \underbrace{\begin{bmatrix} 1 & x_t \end{bmatrix}}_{\boldsymbol{F}_t^\top} \times \underbrace{\begin{bmatrix} \alpha_t \\ \beta_t \end{bmatrix}}_{\boldsymbol{\theta}_t} + \nu_t \\
&\vdots \\
\boldsymbol{\theta}_t = \boldsymbol{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t &\Longrightarrow \underbrace{\begin{bmatrix} \alpha_t \\ \beta_t \end{bmatrix}}_{\boldsymbol{\theta}_t} = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\boldsymbol{G}_t = \boldsymbol{I}} \times \underbrace{\begin{bmatrix} \alpha_{t-1} \\ \beta_{t-1} \end{bmatrix}}_{\boldsymbol{\theta}_{t-1}} + \boldsymbol{\omega}_t
\end{aligned} \tag{8}$$

Onde:
x_t: Variável explicativa observada (ex: Preço do produto no dia t);
β_t: Elasticidade-preço.

Modelo de Função de Transferência (Efeito Memória)

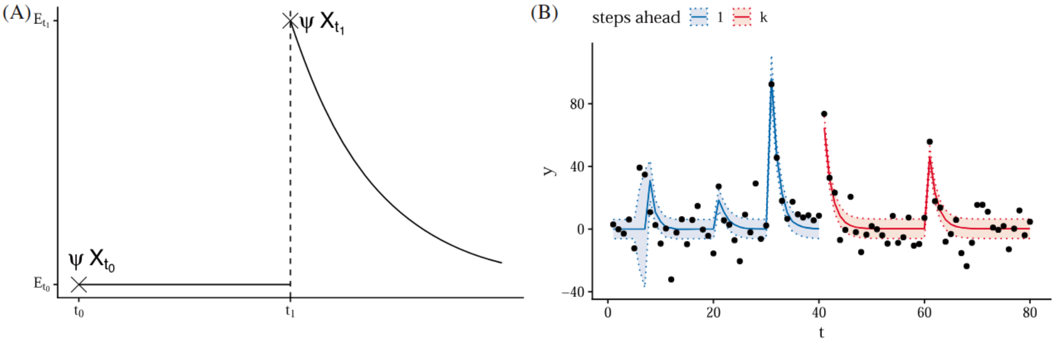
- Modela o efeito defasado de uma intervenção x_t (ex: Propaganda) sobre y_t.
- Seja E_t o estado latente do "efeito acumulado":

$$y_t = \mu_t + E_t + \nu_t \Longrightarrow y_t = \underbrace{\begin{bmatrix} 1 & 1 \end{bmatrix}}_{\boldsymbol{F}_t^\top} \times \underbrace{\begin{bmatrix} \mu_t \\ E_t \end{bmatrix}}_{\boldsymbol{\theta}_t} + \nu_t, \quad E_t = \underbrace{\lambda}_{\text{Decaimento}} E_{t-1} + \underbrace{\psi}_{\text{Impacto}} x_t \tag{9}$$

Onde:
ψx_t: Impacto imediato do investimento em mídia x_t;
$\lambda \in (0, 1)$: Fator de retenção (memória).
μ_t: Vendas sem efeito da campanha.
Ex: Propaganda de TV. Você gasta x_t hoje. O efeito nas vendas $(E_t) \uparrow$ instantaneamente (ψ) e $\downarrow$ suavemente em x_{t+h} a uma taxa λ, mesmo se $x_{t+1} = 0$.

Exemplo: O Modelo de Função de Transferência

Figure: **(A)** Função de transferência de 1ª ordem: Ganho estocástico ($\psi \sim N$) define o salto inicial, e fator de decaimento ($0 < \lambda < 1$) define a memória. **(B)** Ajuste do Modelo: A linha azul ($h = 1$) segue os dados observados; a linha vermelha ($h = k$) projeta a curva teórica de esquecimento a longo prazo.



Fonte: Migon, Alves, Menezes e Pinheiro (2023)

Lab250

Modelo de Superposição de Blocos (Tend. + Saz. + Regressão)

- $y_t = \text{Tendência}(\mu_t) + \text{Sazonal}(S_t) + \text{Regressão}(\beta_t x_t) + \nu_t$

$$
\begin{aligned}
y_t &= \underbrace{\begin{bmatrix} \overbrace{1 \quad 0}^{\text{Tendência}} & \overbrace{1 \quad 0}^{\text{Sazonal}} & \overbrace{x_t}^{\text{Reg}} \end{bmatrix}}_{\boldsymbol{F}_t^\top} \times \underbrace{\begin{bmatrix} \mu_t \\ \beta_t^{tend} \\ S_{1,t} \\ S_{2,t} \\ \beta_t^{reg} \end{bmatrix}}_{\boldsymbol{\theta}_{global,t}} + \nu_t \\
&\vdots \\
\underbrace{\begin{bmatrix} \mu_t \\ \beta_t^{tend} \\ S_{1,t} \\ S_{2,t} \\ \beta_t^{reg} \end{bmatrix}}_{\boldsymbol{\theta}_t} &= \underbrace{\left[\begin{array}{cc|cc|c} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & \cos(\omega) & \sin(\omega) & 0 \\ 0 & 0 & -\sin(\omega) & \cos(\omega) & 0 \\ \hline 0 & 0 & 0 & 0 & 1 \end{array}\right]}_{\boldsymbol{G}=\text{diag}(\boldsymbol{G}_{tend}, \boldsymbol{G}_{saz}, \boldsymbol{G}_{reg})} \times \underbrace{\begin{bmatrix} \mu_{t-1} \\ \beta_{tend-1} \\ S_{1,t-1} \\ S_{2,t-1} \\ \beta_{reg-1} \end{bmatrix}}_{\boldsymbol{\theta}_{t-1}} + \boldsymbol{\omega}_t
\end{aligned}
\tag{10}
$$

$\Longrightarrow$ **Modelos estáticos**

$$p(\boldsymbol{\theta}|y,\boldsymbol{\psi}) \propto \ell(\boldsymbol{\theta}|y)p(\boldsymbol{\theta}|\boldsymbol{\psi}), \quad \ell(\boldsymbol{\theta}|y) = p(y|\boldsymbol{\theta}), \;\; \boldsymbol{\theta} \in \boldsymbol{\Theta}$$

$$p(y_f|y) = \int_{\Theta} p(y_f|\boldsymbol{\theta})p(\boldsymbol{\theta}|y)d\boldsymbol{\theta} = E_{\boldsymbol{\theta}|y}[p(y_f|\boldsymbol{\theta})]$$

Inferência Bayesiana em Modelos Lineares Dinâmicos

$\Longrightarrow$ Modelos estáticos

$$p(\boldsymbol{\theta}|y,\psi) \propto \ell(\boldsymbol{\theta}|y)p(\boldsymbol{\theta}|\psi), \quad \ell(\boldsymbol{\theta}|y) = p(y|\boldsymbol{\theta}), \;\; \boldsymbol{\theta} \in \boldsymbol{\Theta}$$

$$p(y_f|y) = \int_{\Theta} p(y_f|\boldsymbol{\theta})p(\boldsymbol{\theta}|y)d\boldsymbol{\theta} = E_{\boldsymbol{\theta}|y}[p(y_f|\boldsymbol{\theta})]$$

onde:

- ψ é Hiperparâmetro, i.e., representa os parâmetros da distribuição dos parâmetros de cada modelo assumido.
- $p(\boldsymbol{\theta}|\psi)$ é Priori, i.e., crença subjetiva sobre os parâmetros antes de ver os dados.
- $p(\boldsymbol{\theta}|y)$ é Distribuição a Posteriori, i.e., crença atualizada (Priori + Dados) após a observação.
- $p(y_f|y)$ é Distribuição Preditiva, i.e., distr. dos dados futuros y_f.

Lab250

Inferência Baysiana em Modelos Lineares Dinâmicos

$\Longrightarrow$ **Modelos Dinâmicos**

Passo 1: Definir prior ($\theta_{t=0}$) dado todo conhecimento prévio (D_0)

$$(\boldsymbol{\theta}_0|D_0) \sim N[\mathbf{m}_0, \mathbf{C}_0], \ \ \mathbf{m}_0 = \mathbb{E}[\boldsymbol{\theta}_0|D_0], \ \ \mathbf{C}_0 = \text{Var}[\boldsymbol{\theta}_0|D_0]$$

$$\Downarrow$$

$$\therefore, (\boldsymbol{\theta}_{t-1}|D_{t-1}) \sim N[\mathbf{m}_{t-1}, \mathbf{C}_{t-1}]$$

Inferência Baysiana em Modelos Lineares Dinâmicos

$\Longrightarrow$ Modelos Dinâmicos

Passo 1: Definir prior ($\theta_{t=0}$) dado todo conhecimento prévio (D_0)

$$(\boldsymbol{\theta}_0|D_0) \sim N[\mathbf{m}_0, \mathbf{C}_0], \quad \mathbf{m}_0 = \mathbb{E}[\boldsymbol{\theta}_0|D_0], \quad \mathbf{C}_0 = \text{Var}[\boldsymbol{\theta}_0|D_0]$$

$$\Downarrow$$

$$\therefore, (\boldsymbol{\theta}_{t-1}|D_{t-1}) \sim N[\mathbf{m}_{t-1}, \mathbf{C}_{t-1}]$$

Passo 2: Encontrar a dist. de θ_t, note que Posteriori $t-1 \implies$ Priori t)

$$\boldsymbol{\theta}_t = \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N[\mathbf{0}, \mathbf{W}_t] \quad \text{(Eq. do sistema))}$$

$$\mathbf{a}_t = \mathbb{E}[\boldsymbol{\theta}_t|D_{t-1}] = \mathbb{E}[\mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t|D_{t-1}] = \mathbf{G}_t\underbrace{\mathbb{E}[\boldsymbol{\theta}_{t-1}|D_{t-1}]}_{\mathbf{m}_{t-1}} + \underbrace{\mathbb{E}[\boldsymbol{\omega}_t]}_{\mathbf{0}} = \mathbf{G}_t\mathbf{m}_{t-1}$$

$$\mathbf{R}_t = \text{Var}[\boldsymbol{\theta}_t|D_{t-1}] = \text{Var}[\mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t|D_{t-1}] = \mathbf{G}_t\underbrace{\text{Var}[\boldsymbol{\theta}_{t-1}|D_{t-1}]}_{\mathbf{C}_{t-1}}\mathbf{G}_t^\top + \underbrace{\text{Var}[\boldsymbol{\omega}_t]}_{\mathbf{W}_t} + \underbrace{2\text{Cov}(\dots)}_{\mathbf{0}\ \text{(indep.)}}$$

$$= \mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}_t^\top + \mathbf{W}_t$$

$$\therefore, (\boldsymbol{\theta}_t|D_{t-1}) \sim N[\mathbf{a}_t, \mathbf{R}_t]$$

Lab250

Inferência Baysiana em Modelos Lineares Dinâmicos

s Passo 3: Fazer a Predição de y_t (Priori $t \rightarrow$ Observável y_t)

$$y_t = \mathbf{F}_t^\top \boldsymbol{\theta}_t + \nu_t, \quad \nu_t \sim N[0, V_t]$$

$\Longrightarrow$ Média Preditiva (f_t)

$$f_t = \mathbb{E}[y_t|D_{t-1}] = \mathbb{E}[\mathbf{F}_t^\top \boldsymbol{\theta}_t + \nu_t|D_{t-1}] = \mathbf{F}_t^\top \underbrace{\mathbb{E}[\boldsymbol{\theta}_t|D_{t-1}]}_{\mathbf{a}_t} + \underbrace{\mathbb{E}[\nu_t]}_{0} = \mathbf{F}_t^\top \mathbf{a}_t$$

$\Longrightarrow$ Variância Preditiva (Q_t)

$$Q_t = \text{Var}[y_t|D_{t-1}] = \text{Var}[\mathbf{F}_t^\top \boldsymbol{\theta}_t + \nu_t|D_{t-1}] = \mathbf{F}_t^\top \underbrace{\text{Var}[\boldsymbol{\theta}_t|D_{t-1}]}_{\mathbf{R}_t} \mathbf{F}_t + \underbrace{\text{Var}[\nu_t]}_{V_t} = \mathbf{F}_t^\top \mathbf{R}_t \mathbf{F}_t + V_t$$

$\therefore$, $(y_t|D_{t-1}) \sim N[f_t, Q_t]$

Lab250

Inferência Baysiana em Modelos Lineares Dinâmicos

Passo 4: Atualizar a distribuição θ_t , i.e, encontrar posteriori de θ_t $(\boldsymbol{\theta}_t|y_t, D_{t-1})$
$\Longrightarrow$ Para obter $(\boldsymbol{\theta}_t|y_t, D_{t-1})$, construímos a distribuição conjunta de $\boldsymbol{\theta}_t$ e y_t.
1: Sabendo que $y_t = \mathbf{F}_t^\top \boldsymbol{\theta}_t + \nu_t$ e que $\nu_t \perp \boldsymbol{\theta}_t$:

$$\text{Cov}[\boldsymbol{\theta}_t, y_t|D_{t-1}] = \text{Cov}[\boldsymbol{\theta}_t, \mathbf{F}_t^\top \boldsymbol{\theta}_t + \nu_t|D_{t-1}] = \underbrace{\text{Var}[\boldsymbol{\theta}_t|D_{t-1}]}_{\mathbf{R}_t} \mathbf{F}_t + \underbrace{\text{Cov}[\boldsymbol{\theta}_t, \nu_t]}_{0} = \mathbf{R}_t\mathbf{F}_t \tag{11}$$

$$\Longrightarrow \left(\begin{bmatrix} \boldsymbol{\theta}_t \\ y_t \end{bmatrix} \middle| D_{t-1} \right) \sim N\left(\begin{bmatrix} \mathbf{a}_t \\ f_t \end{bmatrix}, \begin{bmatrix} \mathbf{R}_t & \mathbf{R}_t\mathbf{F}_t \\ \mathbf{F}_t^\top \mathbf{R}_t & Q_t \end{bmatrix} \right) \tag{12}$$

Inferência Baysiana em Modelos Lineares Dinâmicos

Passo 4: Atualizar a distribuição θ_t , i.e, encontrar posteriori de θ_t $(\boldsymbol{\theta}_t|y_t, D_{t-1})$
$\Longrightarrow$ Para obter $(\boldsymbol{\theta}_t|y_t, D_{t-1})$, construímos a distribuição conjunta de $\boldsymbol{\theta}_t$ e y_t.
1: Sabendo que $y_t = \mathbf{F}_t^\top \boldsymbol{\theta}_t + \nu_t$ e que $\nu_t \perp \boldsymbol{\theta}_t$:

$$\text{Cov}[\boldsymbol{\theta}_t, y_t|D_{t-1}] = \text{Cov}[\boldsymbol{\theta}_t, \mathbf{F}_t^\top \boldsymbol{\theta}_t + \nu_t|D_{t-1}] = \underbrace{\text{Var}[\boldsymbol{\theta}_t|D_{t-1}]}_{\mathbf{R}_t} \mathbf{F}_t + \underbrace{\text{Cov}[\boldsymbol{\theta}_t, \nu_t]}_{0} = \mathbf{R}_t \mathbf{F}_t \tag{11}$$

$$\Longrightarrow \left(\begin{bmatrix} \boldsymbol{\theta}_t \\ y_t \end{bmatrix} \middle| D_{t-1} \right) \sim N\left(\begin{bmatrix} \mathbf{a}_t \\ f_t \end{bmatrix}, \begin{bmatrix} \mathbf{R}_t & \mathbf{R}_t \mathbf{F}_t \\ \mathbf{F}_t^\top \mathbf{R}_t & Q_t \end{bmatrix} \right) \tag{12}$$

2: Prop: Se $\binom{X}{Y} \sim N\left(\binom{\mu_x}{\mu_y}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix} \right)$, então $(X|Y = y) \sim N(\mathbb{E}[X|Y], \text{Var}[X|Y])$:
$\mathbb{E}[X|Y] = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y)$, $\text{Var}[X|Y] = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}$

Inferência Baysiana em Modelos Lineares Dinâmicos

Passo 4: Atualizar a distribuição θ_t , i.e, encontrar posteriori de θ_t $(\boldsymbol{\theta}_t|y_t, D_{t-1})$
$\Longrightarrow$ Para obter $(\boldsymbol{\theta}_t|y_t, D_{t-1})$, construímos a distribuição conjunta de $\boldsymbol{\theta}_t$ e y_t.
1: Sabendo que $y_t = \mathbf{F}_t^\top \boldsymbol{\theta}_t + \nu_t$ e que $\nu_t \perp \boldsymbol{\theta}_t$:

$$\text{Cov}[\boldsymbol{\theta}_t, y_t|D_{t-1}] = \text{Cov}[\boldsymbol{\theta}_t, \mathbf{F}_t^\top \boldsymbol{\theta}_t + \nu_t|D_{t-1}] = \underbrace{\text{Var}[\boldsymbol{\theta}_t|D_{t-1}]}_{\mathbf{R}_t} \mathbf{F}_t + \underbrace{\text{Cov}[\boldsymbol{\theta}_t, \nu_t]}_{0} = \mathbf{R}_t\mathbf{F}_t \tag{11}$$

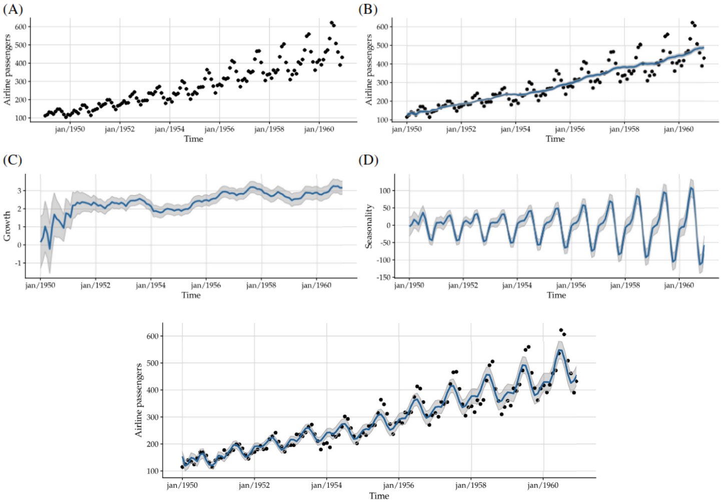
$$\Longrightarrow \left(\begin{bmatrix} \boldsymbol{\theta}_t \\ y_t \end{bmatrix} \middle| D_{t-1} \right) \sim N\left(\begin{bmatrix} \mathbf{a}_t \\ f_t \end{bmatrix}, \begin{bmatrix} \mathbf{R}_t & \mathbf{R}_t\mathbf{F}_t \\ \mathbf{F}_t^\top \mathbf{R}_t & Q_t \end{bmatrix} \right) \tag{12}$$

2: Prop: Se $\binom{X}{Y} \sim N\left(\binom{\mu_x}{\mu_y}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix} \right)$, então $(X|Y=y) \sim N(\mathbb{E}[X|Y], \text{Var}[X|Y])$:
$\mathbb{E}[X|Y] = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y)$, $\text{Var}[X|Y] = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}$
3: Aplicação ao DLM, temos que $\Sigma_{xy} = \mathbf{R}_t\mathbf{F}_t$ e $\Sigma_{yy}^{-1} = Q_t^{-1}$:

$$\begin{aligned} \mathbf{m}_t &= \mathbf{a}_t + (\mathbf{R}_t\mathbf{F}_t)Q_t^{-1}(y_t - f_t) \xrightarrow{\text{Def: } \mathbf{A}_t = \mathbf{R}_t\mathbf{F}_t Q_t^{-1}} \mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t e_t \\ \mathbf{C}_t &= \mathbf{R}_t - (\mathbf{R}_t\mathbf{F}_t)Q_t^{-1}(\mathbf{F}_t^\top \mathbf{R}_t) \xrightarrow{\text{Def: } \mathbf{A}_t = \mathbf{R}_t\mathbf{F}_t Q_t^{-1}} \mathbf{C}_t = \mathbf{R}_t - \mathbf{A}_t Q_t \mathbf{A}_t^\top \end{aligned} \tag{13}$$

$\therefore$, $(\boldsymbol{\theta}_t|y_t, D_t) \sim N[\mathbf{m}_t, \mathbf{C}_t]$ (Posteriori Atualizada).

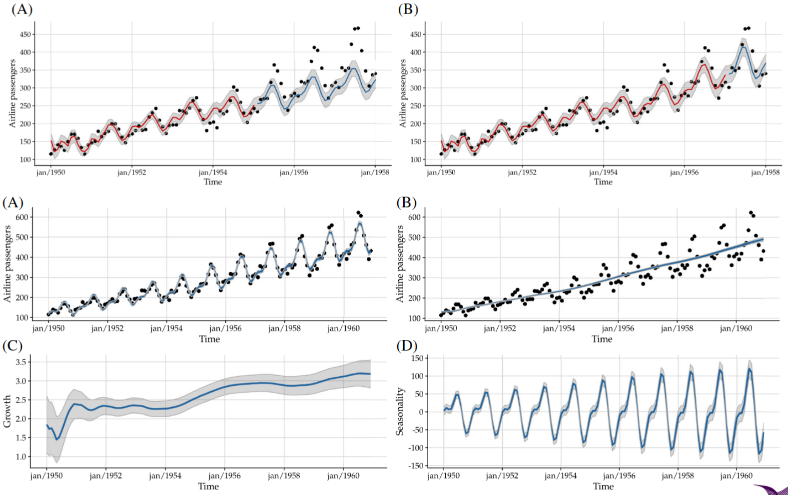
Figure: Número mensal de passageiros transportados no Reino Unido de 1949 a 1960: **(A)** Série Observada (y_t); **(B)** Nível Local (μ_t) filtrado; **(C)** Fator de Crescimento/Inclinação (β_t); **(D)** Componente Sazonal (S_t) com amplitude variante. **Abaixo:** Ajuste final do modelo.



Fonte: Migon, Alves, Menezes e Pinheiro (2023)

Lab250

Figure: **(A)** Previsão de longo prazo feita em 1955 (linha azul). **(B)** Previsão atualizada em 1957 após observar novos dados (linha vermelha). Suavização/Smoothing ($E[\boldsymbol{\theta}_t|D_T]$): **(A)** Ajuste global; **(B)** Nível; **(C)** Crescimento; **(D)** Sazonalidade.



Fonte: Migon, Alves, Menezes e Pinheiro (2023)

Lab250

Modelos Dinâmicos Generalizados (DGLM)

$\Longrightarrow$ Os modelos discutidos até aqui, impunham dist. normal, o que nem sempre é possível.

Modelos Dinâmicos Generalizados (DGLM)

$\Longrightarrow$ Os modelos discutidos até aqui, impunham dist. normal, o que nem sempre é possível.
$\Longrightarrow$ **Solução:** Generalizar a distribuição de y_t (Família Exponencial) e conectar aos parâmetros de regressão $\boldsymbol{\theta}_t$ via função de ligação $g(\cdot)$.

$$p(y_t|\eta_t, \phi_t) = b(y_t, \phi_t) \exp\left\{\frac{y_t\eta_t - a(\eta_t)}{\phi_t}\right\} \quad \text{(Eq. das observações - EF)}$$

$$\Downarrow \quad \mu_t = a'(\eta_t),\; V_t = \phi_t a''(\eta_t),\; g(\mu_t) = \lambda_t$$

$$\lambda_t = \mathbf{F}_t^\top \boldsymbol{\theta}_t \quad \text{(Preditor Linear)}$$

$$\boldsymbol{\theta}_t = \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim [\mathbf{0}, \mathbf{W}_t], \quad \text{(Eq. dos estados)}$$

DGLM $\Longrightarrow \{\text{EF}, g(\cdot), \mathbf{F}_t, \mathbf{G}_t, \mathbf{W}_t\}$

Lab250

Modelos Dinâmicos Generalizados (DGLM)

$\Longrightarrow$ Os modelos discutidos até aqui, impunham dist. normal, o que nem sempre é possível.
$\Longrightarrow$ **Solução:** Generalizar a distribuição de y_t (Família Exponencial) e conectar aos parâmetros de regressão $\boldsymbol{\theta}_t$ via função de ligação $g(\cdot)$.

$$p(y_t|\eta_t, \phi_t) = b(y_t, \phi_t) \exp\left\{\frac{y_t\eta_t - a(\eta_t)}{\phi_t}\right\} \quad \text{(Eq. das observações - EF)}$$

$$\Downarrow \quad \mu_t = a'(\eta_t), \;\; V_t = \phi_t a''(\eta_t), \;\; g(\mu_t) = \lambda_t$$

$$\lambda_t = \mathbf{F}_t^\top \boldsymbol{\theta}_t \quad \text{(Preditor Linear)}$$

$$\boldsymbol{\theta}_t = \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim [\mathbf{0}, \mathbf{W}_t], \quad \text{(Eq. dos estados)}$$

DGLM $\Longrightarrow \{\text{EF}, g(\cdot), \mathbf{F}_t, \mathbf{G}_t, \mathbf{W}_t\}$

- η_t: Parâmetro natural (variante no tempo); ϕ_t: Parâmetro de escala/dispersão.
- $a(\cdot)$: Função que define os momentos (Média μ_t e Variância V_t).
- $g(\cdot)$: Função de ligação monótona e diferenciável (conecta μ_t ao preditor λ_t).
- $\boldsymbol{\omega}_t \sim [\mathbf{0}, \mathbf{W}_t]$: Indica especificação apenas de 1° e 2° momentos (não necessariamente Normal).

Avanços Recentes: Alta Dimensionalidade

1. Modelos Matriciais (Matrix-Variate DLM): Para modelar m séries temporais simultaneamente (ex: Vendas de m produtos).

$$\begin{aligned} \mathbf{Y}_t &= \mathbf{F}_t^\top \mathbf{\Theta}_t + \mathbf{E}_t, \quad \mathbf{E}_t \sim N(0, V_t\mathbf{\Sigma}) \\ \mathbf{\Theta}_t &= \mathbf{G}\mathbf{\Theta}_{t-1} + \mathbf{\Omega}_t \end{aligned} \tag{14}$$

- $\mathbf{\Theta}_t$: Matriz de estados ($q \times m$).
- Permite compartilhar estrutura ($\mathbf{F}, \mathbf{G}$) mas ter parâmetros distintos para cada série. Ex: Intensidade do vento e Direção do vento, ambas são afetadas pelas mesmas variáveis climáticas (F_t), mas de formas diferentes (os θ são diferentes para direção e intensidade).

Lab250

Avanços Recentes: Alta Dimensionalidade

1. Modelos Matriciais (Matrix-Variate DLM): Para modelar m séries temporais simultaneamente (ex: Vendas de m produtos).

$$\begin{aligned} \mathbf{Y}_t &= \mathbf{F}_t^\top \mathbf{\Theta}_t + \mathbf{E}_t, \quad \mathbf{E}_t \sim N(0, V_t \mathbf{\Sigma}) \\ \mathbf{\Theta}_t &= \mathbf{G}\mathbf{\Theta}_{t-1} + \mathbf{\Omega}_t \end{aligned} \tag{14}$$

- $\mathbf{\Theta}_t$: Matriz de estados ($q \times m$).
- Permite compartilhar estrutura ($\mathbf{F}, \mathbf{G}$) mas ter parâmetros distintos para cada série. Ex: Intensidade do vento e Direção do vento, ambas são afetadas pelas mesmas variáveis climáticas (F_t), mas de formas diferentes (os θ são diferentes para direção e intensidade).

2. Modelos Hierárquicos Dinâmicos: Estrutura multinível para capturar dependência entre grupos (são a versão dinâmica dos Modelos Lineares Mistos.).

$$\begin{aligned} y_t &= \mathbf{F}_{1,t}\boldsymbol{\theta}_{1,t} + \epsilon_{1,t} \quad \text{(Nível Observacional)} \\ \boldsymbol{\theta}_{1,t} &= \mathbf{F}_{2,t}\boldsymbol{\theta}_t + \epsilon_{2,t} \quad \text{(Nível Estrutural/Grupos)} \\ \boldsymbol{\theta}_t &= \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \omega_t \quad \text{(Evolução Latente)} \end{aligned}$$

Lab250

Conclusão

- **Flexibilidade:** DLMs acomodam tendências, sazonalidade e regressores dinâmicos (como preço/mídia) naturalmente.
- **Adaptação:** Reagem rapidamente a mudanças estruturais (via Intervenção ou Fatores de Desconto), cruciais em ambientes voláteis (Marketing).
- **Interpretabilidade:** A decomposição em componentes latentes (nível, crescimento, sazonalidade) permite analisar separadamente cada efeito (Decouple/Recouple).
- **Generalização:** A extensão para DGLM e Modelos Hierárquicos permite aplicação em contagens (vendas discretas) e Big Data (múltiplas lojas).

Lab250

Referência

- Migon, H. S., Alves, M. B., Menezes, A. F., & Pinheiro, E. G. (2023). **A review of Bayesian dynamic forecasting models: Applications in marketing**. Applied Stochastic Models in Business and Industry, 39(3), 471-493.

Obrigado!

Siga LAB250 clicando Aqui

Lab250