

PREVISÕES BAYESIANAS E MODELOS GRÁFICOS

Ralph dos Santos Silva

<http://www.ralphsilva.org/previsoesbayesianasemodelosgraficos.html>

Instituto de Matemática
Centro de Ciências Matemáticas e da Natureza
Universidade Federal do Rio de Janeiro

Modelos em Espaço de Estados

Outra abordagem para modelagem de séries temporais.

As séries temporais não precisam ser estacionárias!

Modelos em espaço de estados

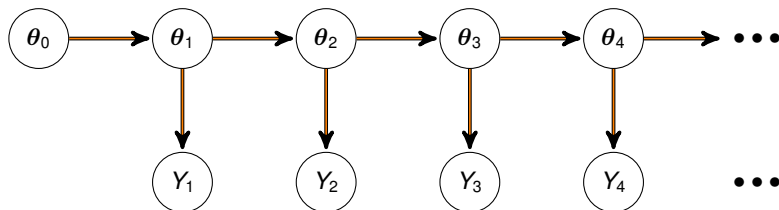


Figura 1: Um DAG de um modelo em espaço de estados.

Defina $\theta_{1:t} = (\theta_1, \theta_2, \dots, \theta_t)^\top$ como uma coleção de variáveis aleatórias.

Considera-se que

- $Y_{1:t}$ são variáveis observáveis; e
- $\theta_{1:t}$ são variáveis não observáveis (latentes).

Nestes modelos, temos as seguintes distribuições:

- $f(\theta_0 | \Psi)$, priori para o instante inicial;
- $f(\theta_t | \theta_{t-1}, \Psi)$, $t = 1, 2, \dots$, para a evolução dos estados; e
- $f(y_t | \theta_t, \Psi)$, $t = 1, 2, \dots$, para as observações.

A distribuição conjunta é dada por

$$f(\theta_0, \theta_{1:T}, \mathbf{y}_{1:T} | \Psi) = \left[\prod_{t=1}^T f(y_t | \theta_t, \Psi) \right] \left[\prod_{t=1}^T f(\theta_t | \theta_{t-1}, \Psi) \right] f(\theta_0 | \Psi)$$

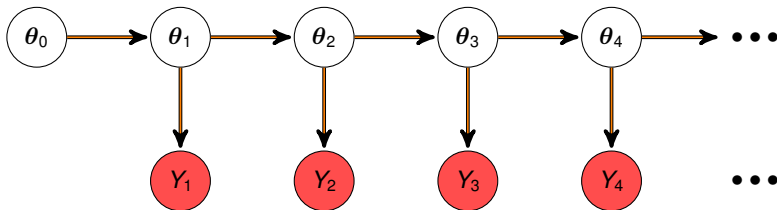


Figura 2: Um DAG de um modelo em espaço de estados.

Quando os vértices $\mathbf{Y}_{1:T}$ são observados, temos que

$$f(\theta_0, \theta_{1:T}, \mathbf{y}_{1:T} | \Psi) = \left[\prod_{t=1}^T f(y_t | \theta_t, \Psi) \right] \left[\prod_{t=1}^T f(\theta_t | \theta_{t-1}, \Psi) \right] f(\theta_0 | \Psi)$$

Modelos estáticos

Os **modelos estáticos** têm uma descrição fixa (através de parâmetros fixos) ao longo das unidades de observação.

- Análise de regressão,

$$Y_t = \beta_0 + \beta_1 z_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2).$$

- Modelos ARMA,

$$Y_t = \alpha + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2).$$

Em séries temporais, essa hipótese muitas vezes é violada:

as estruturas mudam com a passagem do tempo.

Nesta aula, faz-se uma introdução aos **modelos em espaço de estados**.

- **A descrição (os parâmetros) muda(m) com a passagem do tempo.**
- Os modelos em espaço de estados nos permitem a modelagem da tendência (média) e variância (volatilidade).
- Os modelos em espaço de estados incluem como casos particulares os modelos estáticos.

A passagem do tempo traz observações e aumenta o nosso conhecimento.

Em modelos em espaço de estados temos também perda de informação devido à passagem do tempo.

Por exemplo, o nível de vendas mês passado é mais relevante hoje que o nível de vendas em janeiro passado.

O modelo de regressão (linear múltiplo)

Uma variável resposta Y é função de um conjunto de variáveis explicativas $z_1, \dots, z_d$ através da relação

$$Y = \theta_0 + \theta_1 z_1 + \dots + \theta_d z_d + \varepsilon.$$

Em muitos casos, assume-se que $\varepsilon \sim \mathcal{N}(0, \sigma^2)$.

A equação acima pode ser compactamente escrita como:

$$Y = \mathbf{z}^\top \boldsymbol{\theta} + \varepsilon,$$

em que

$$\mathbf{z} = \begin{pmatrix} 1 \\ z_1 \\ \vdots \\ z_d \end{pmatrix} \quad \text{e} \quad \boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_d \end{pmatrix}.$$

A natureza das variáveis explicativas ou regressoras é bastante ampla.

Pode-se utilizar qualquer variável quantificável.

Os coeficientes de regressão $\theta_1, \dots, \theta_d$ informam sobre a influência que os regressores têm sobre a resposta Y .

Na prática, seus valores são desconhecidos e estimados a partir de uma coleção de observações feitas sobre o modelo.

Observamos respostas $y_1, \dots, y_T$ com seus respectivos regressores $\mathbf{z}_1, \dots, \mathbf{z}_T$. Simbolicamente, temos que

$$y_t = \mathbf{z}_t^\top \boldsymbol{\theta} + \varepsilon_t, \quad t = 1, \dots, T,$$

em que $\mathbf{z}_t = (z_{t1}, z_{t2}, \dots, z_{td})^\top$.

Modelo em espaço de estados

Em modelos em espaço de estados os parâmetros mudam com o passar do tempo.

O modelo de regressão é estendido para

$$Y_t = \mathbf{z}_t^\top \boldsymbol{\theta}_t + \varepsilon_t, \quad t = 1, \dots, T.$$

Note a **indexação de $\boldsymbol{\theta}$** .

Esta formulação cria uma profusão de parâmetros a serem estimados.

O modelo necessita de mais informação. Essa informação vem do fato que os parâmetros sucessivos estão intimamente relacionados.

Por exemplo, um parâmetro é igual ao seu antecessor mais uma pequena perturbação causada pelas mudanças às quais o sistema está sujeito.

Se o sistema é estático, como no modelo de regressão, temos que

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} = \boldsymbol{\theta}.$$

Em modelos em espaço de estados, admitimos a forma mais geral dada por

$$\theta_t = \mathbf{G}_t \theta_{t-1} + \mathbf{w}_t,$$

em que

- $\mathbf{G}_t$ contém valores conhecidos (ou parâmetros fixos); e
- $\mathbf{w}_t$ é uma perturbação aleatória.

A equação acima é conhecida como **equação do sistema**.

A **matriz de evolução** $\mathbf{G}_t$ controla a parte determinística da evolução do sistema e estabelece a propagação do sistema ao longo do tempo.

A **perturbação** $\mathbf{w}_t$ é responsável pela introdução de incertezas devidas à passagem do tempo e consequente perda de informação.

Note que se $\mathbf{G}_t = \mathbf{I}$ e $\mathbf{w}_t = \mathbf{0}$, o modelo se reduz ao caso estático.

Exemplo

Suponha que uma série de **vendas** possa ser explicada pela respectiva série de **preços** através da seguinte relação:

$$vendas_t = \mu_t + \beta_t \times preço_t + v_t$$

$$\mu_t = \mu_{t-1} + w_{1t}$$

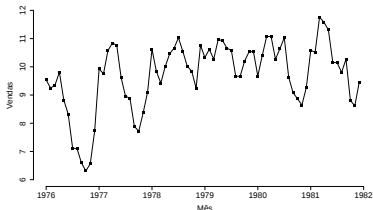
$$\beta_t = \beta_{t-1} + w_{2t}.$$

As **vendas** são explicadas pelo **preço** em uma regressão dinâmica.

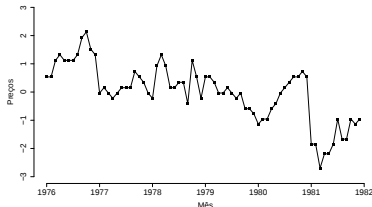
A maior (ou menor) estabilidade dessa relação será controlada pela magnitude dos incrementos (erros) w_{1t} e w_{2t} .

Exemplo

Analisaremos a série de vendas mensais de um certo tipo doce no Reino Unido de janeiro de 1976 a dezembro de 1981. Também utilizaremos a série de preços, na escala do logaritmo, deste doce como variável explicativa.



(a) Vendas



(b) Preços

Figura 3: Vendas e logaritmo dos preços mensais de um certo tipo doce no Reino Unido de janeiro de 1976 a dezembro de 1981.

Ver arquivo exemplo_1.r

Análise sequencial

A natureza sequencial de séries temporais é devida à obtenção sequencial de informação.

A **distribuição do estado** nos informa como a partir da **posteriori de ontem** podemos chegar à **priori de hoje**.

O **método bayesiano** nos ensina como combinar a **priori de hoje** com a informação que acabamos de obter para chegar à **posteriori de hoje**.

Para amanhã e dias futuros, o ciclo se repete.

Em termos probabilísticos, temos que

$$\cdots \rightarrow f(\theta_{t-1} \mid y_{1:t-1}, \Psi) \rightarrow f(\theta_t \mid y_{1:t-1}, \Psi) \rightarrow f(\theta_t \mid y_{1:t}, \Psi) \rightarrow \cdots$$

e também $f(y_t \mid y_{1:t-1}, \Psi)$ como a distribuição preditiva.

Suponha que a posteriori, $f(\theta_{t-1} \mid y_{1:t-1}, \Psi)$, no instante $t - 1$ seja conhecida. Então, a priori para o instante t é dada por

$$f(\theta_t \mid y_{1:t-1}, \Psi) = \int_{\Theta_{t-1}} f(\theta_t \mid \theta_{t-1}, \Psi) f(\theta_{t-1} \mid y_{1:t-1}, \Psi) d\theta_{t-1}.$$

Note que utilizamos a **distribuição do estado**.

Agora, a posteriori para o instante t é dada via **teorema de Bayes** por

$$f(\theta_t \mid y_{1:t}, \Psi) = \frac{f(y_t \mid \theta_t, \Psi) f(\theta_t \mid y_{1:t-1}, \Psi)}{f(y_t \mid y_{1:t-1}, \Psi)},$$

em que

$$f(y_t \mid y_{1:t-1}, \Psi) = \int_{\Theta_t} f(y_t \mid \theta_t, \Psi) f(\theta_t \mid y_{1:t-1}, \Psi) d\theta_t.$$

É possível descrever a previsão k passos à frente via $f(y_{t+k} \mid y_{1:t-1}, \Psi)$.

Se o modelo em espaço de estado for linear e gaussiano, então é possível obter uma solução de forma analítica (condicional a Ψ) da sequência acima.

Modelo em espaço de estados linear e gaussiano

O **modelo linear** pode ser definido como

- **Equação das observações:**

$$Y_t = \mathbf{F}_t^\top \boldsymbol{\theta}_t + v_t, \quad v_t \sim \mathcal{N}(0, V_t)$$

- **Equação do sistema:**

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{W}_t)$$

- **Priori:**

$$\boldsymbol{\theta}_0 \sim \mathcal{N}(\mathbf{m}_0, \mathbf{C}_0).$$

Em geral, supomos que $\{\mathbf{F}_t, V_t, \mathbf{G}_t, \mathbf{W}_t\}$ são conhecidos para todo instante de tempo t e podem depender de Ψ . Obviamente, precisamos que $\mathbf{m}_0$ e $\mathbf{C}_0$ sejam conhecidos também.

No exemplo anterior, $Y_t = \text{venda}_t$, $\mathbf{F}_t = (1, \text{preço}_t)^\top$.

$$\boldsymbol{\theta}_t = \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix}, \quad \mathbf{G}_t = \mathbf{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{e} \quad \mathbf{w}_t = \begin{pmatrix} w_{1t} \\ w_{2t} \end{pmatrix}$$

Filtro de Kalman (condicional a Ψ)

Priori:	$\boldsymbol{\theta}_0 \sim \mathcal{N}(\mathbf{m}_0, \mathbf{C}_0)$	
Observação:	$Y_t = \mathbf{F}_t^\top \boldsymbol{\theta}_t + v_t$	$v_t \sim \mathcal{N}(0, V_t)$
Sistema:	$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t$	$\mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{W}_t)$
Informação:	$(\boldsymbol{\theta}_{t-1} \mid y_{1:t-1}) \sim \mathcal{N}(\mathbf{m}_{t-1}, \mathbf{C}_{t-1})$ $(\boldsymbol{\theta}_t \mid y_{1:t-1}) \sim \mathcal{N}(\mathbf{a}_t, \mathbf{R}_t)$	$\mathbf{a}_t = \mathbf{G}_t \mathbf{m}_{t-1}$ $\mathbf{R}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t^\top + \mathbf{W}_t$
Previsão:	$(y_t \mid y_{1:t-1}) \sim \mathcal{N}(f_t, Q_t)$	$f_t = \mathbf{F}_t^\top \mathbf{a}_t$ $Q_t = \mathbf{F}_t^\top \mathbf{R}_t \mathbf{F}_t + V_t$
Posteriori:	$(\boldsymbol{\theta}_t \mid y_{1:t}) \sim \mathcal{N}(\mathbf{m}_t, \mathbf{C}_t)$	$\mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t \mathbf{e}_t$ $\mathbf{C}_t = \mathbf{R}_t - \mathbf{A}_t \mathbf{A}_t^\top Q_t$ $\mathbf{e}_t = y_t - f_t$ $\mathbf{A}_t = \mathbf{R}_t \mathbf{F}_t / Q_t$

Modelo de tendência estável

É composto apenas de um nível que varia segundo um passeio aleatório.

O modelo de tendência estável é dado por

$$\begin{aligned}Y_t &= \mu_t + v_t, & v_t &\sim \mathcal{N}(0, V_t) \\ \mu_t &= \mu_{t-1} + w_t, & w_t &\sim \mathcal{N}(0, W_t).\end{aligned}$$

O nível permanece localmente constante, mas varia quando se considera longos períodos de tempo.

Usualmente, a variação das observações em torno dos níveis (medida por V_t) é bem maior que as variações temporais do nível ao longo do tempo (medidas por W_t).

O modelo de tendência estável é obtido ao particularizar o modelo dinâmico com $\mathbf{F}_t = 1$ e $\mathbf{G}_t = 1$.

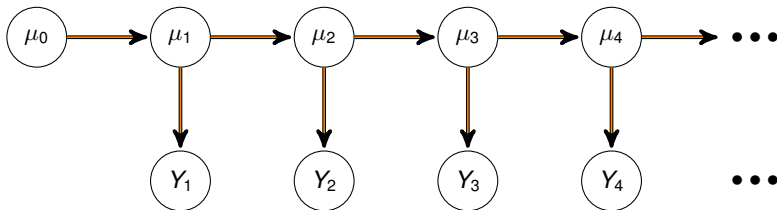


Figura 4: Um DAG do modelo de tendência estável (condicional a $\Psi = (V, W)$).

Aplicando a série de vendas, vendas_t e $\{\mathbf{F}_t = 1, V_t = V, \mathbf{G}_t = 1, \mathbf{W}_t = W\}$, temos que

$$\begin{aligned}\text{vendas}_t &= \mu_t + v_t, & v_t &\sim \mathcal{N}(0, V) \\ \mu_t &= \mu_{t-1} + w_t, & w_t &\sim \mathcal{N}(0, W)\end{aligned}$$

Para a série temporal de vendas de doce faremos $V_t = V$ e $W_t = W$ para todo t (homocedasticidade; variâncias constantes ao longo do tempo).

Modelo de tendência estável

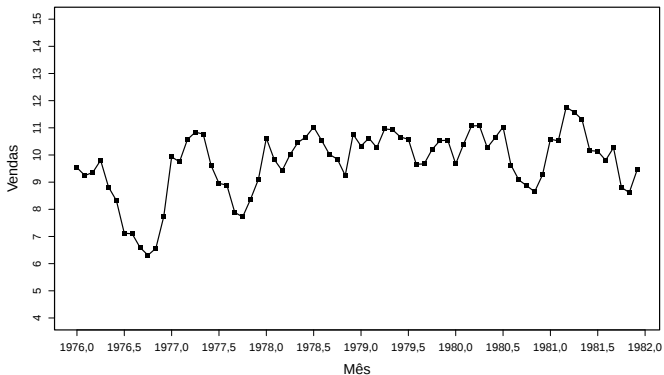


Figura 5: Vendas mensais de um tipo de doces de janeiro de 1976 a dezembro de 1981.

Ver arquivo exemplo_3.r

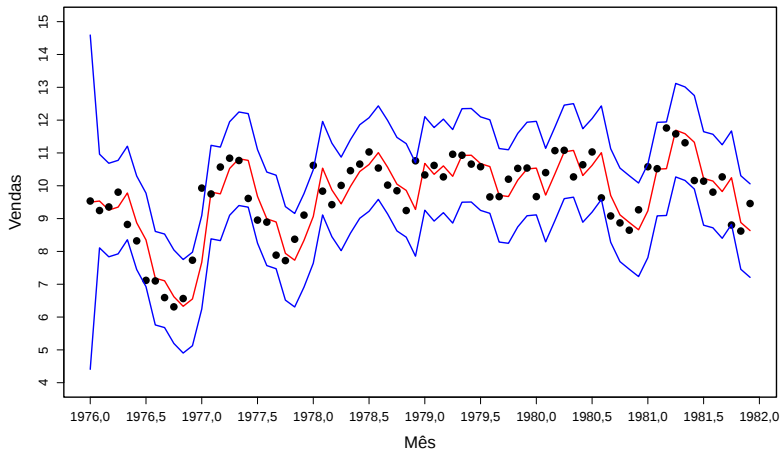


Figura 6: Previsão da média de vendas, $(\mu_t|y_{1:t-1})$. A distribuição a priori em t .

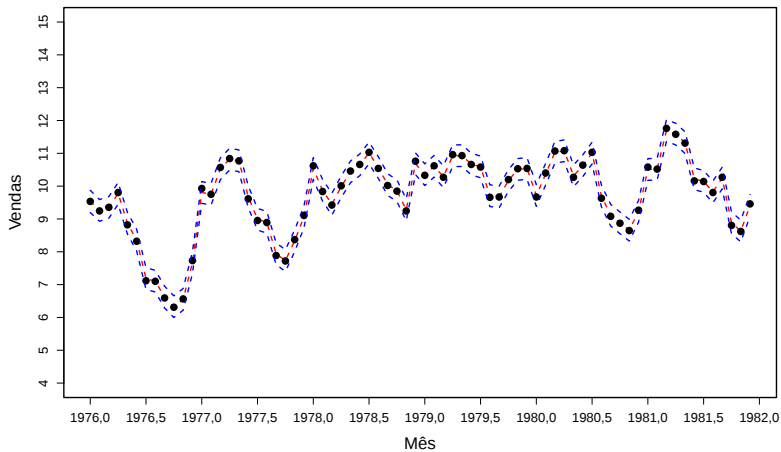


Figura 7: Estimativa da média de vendas, $(\mu_t | y_{1:t})$. A distribuição a posteriori em t .

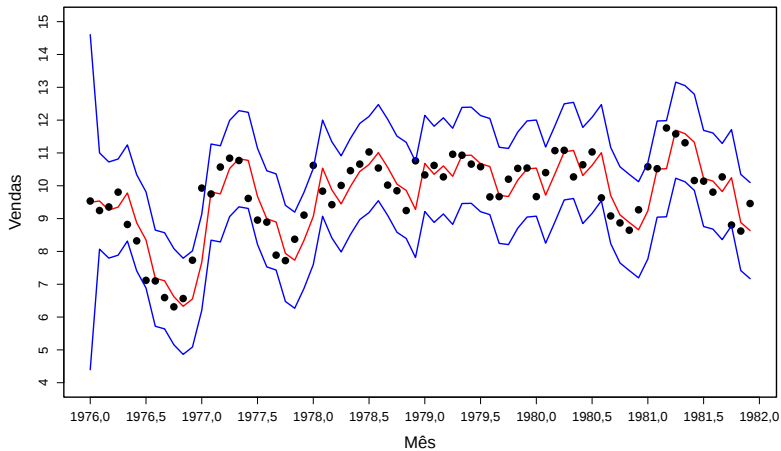


Figura 8: Previsão (um passo à frente) das vendas, $(y_t|y_{1:t-1})$.

Modelo de tendência linear

A tendência linear é quantificada através de um parâmetro adicional.

O modelo de tendência linear é dado por

$$\begin{aligned}Y_t &= \mu_t + v_t, & v_t &\sim \mathcal{N}(0, V_t) \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + w_{1t} \\ \beta_t &= \beta_{t-1} + w_{2t},\end{aligned}$$

em que $\mathbf{w}_t = (w_{1t}, w_{2t})^\top \sim \mathcal{N}(\mathbf{0}, \mathbf{W}_t)$.

Esse modelo é obtido com $\mathbf{F}_t = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ e $\mathbf{G}_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

O nível permanece localmente linear, mas a forma da reta pode variar com o tempo.

Para a série de vendas de doce faremos $V_t = V$ e $\mathbf{W}_t = \mathbf{W} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$ para todo t (independência e variâncias constantes ao longo do tempo).

Modelo de tendência linear

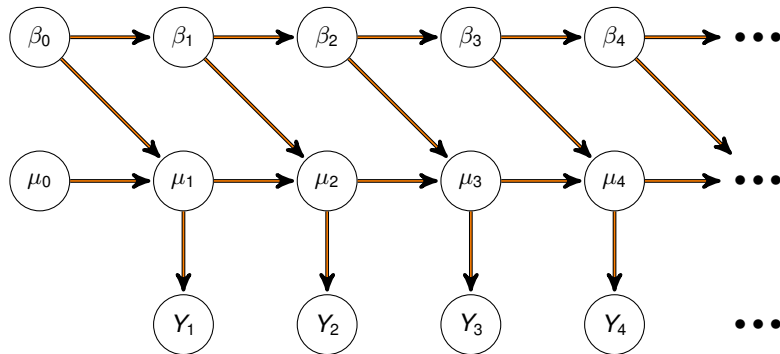


Figura 9: Um DAG do modelo de tendência linear (condicional a $\Psi = (V, \sigma_1^2, \sigma_2^2)$).

Ver arquivo exemplo_4.r

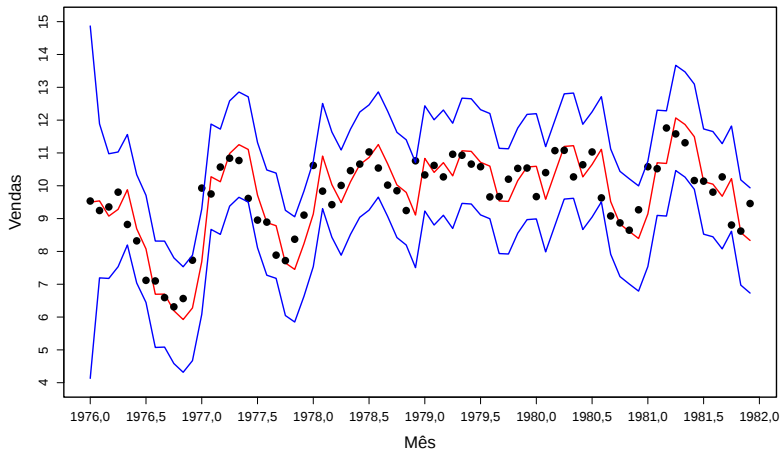


Figura 10: Previsão da média de vendas, $(\mu_t|y_{1:t-1})$. A distribuição a priori em t .

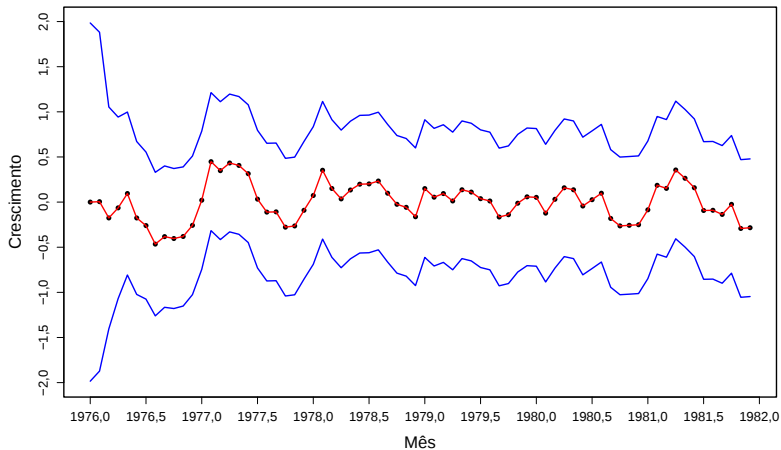


Figura 11: Previsão do crescimento das vendas, $(\beta_t|y_{1:t-1})$. A distribuição a priori em t .

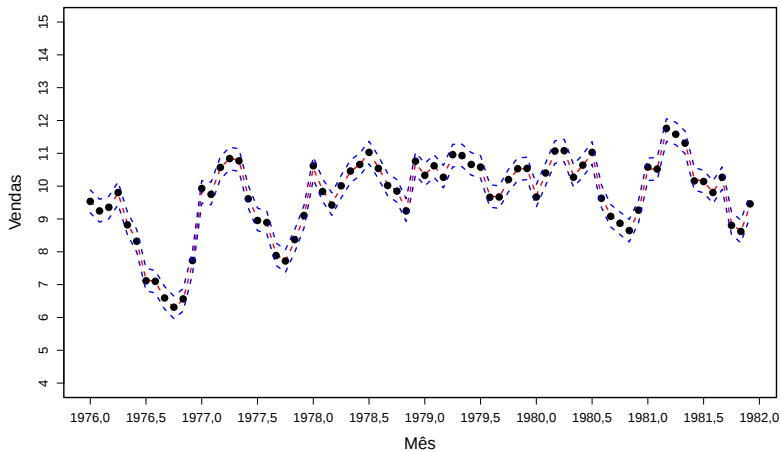


Figura 12: Estimativa da média de vendas, $(\mu_t|y_{1:t})$. A distribuição a posteriori em t .

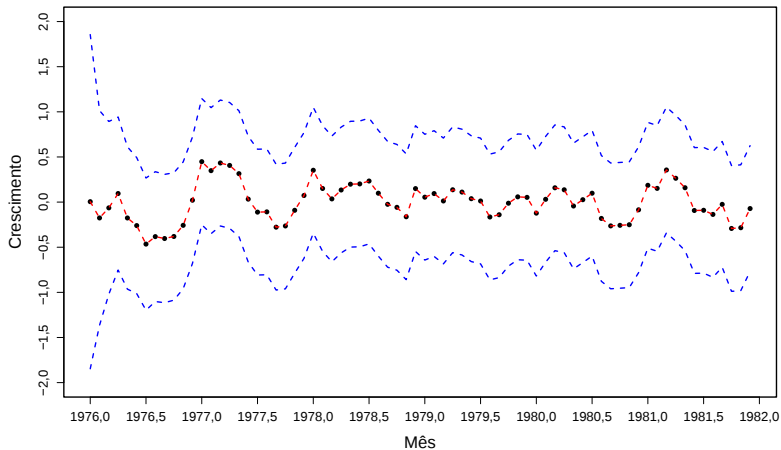


Figura 13: Estimativa do crescimento das vendas, $(\beta_t|y_{1:t})$. A distribuição a posteriori em t .

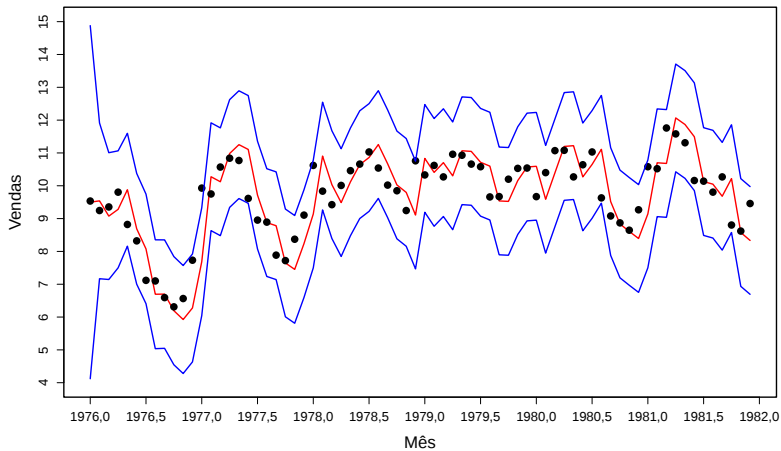


Figura 14: Previsão (um passo à frente) das vendas, $(y_t|y_{1:t-1})$.