Aula 6: Modelos de espaço de estados

Marcus L. Nascimento

1 de dezembro de 2025

Introdução

Introdução

- Modelos de espaço de estados podem ser aplicados tanto quando a variável aleatória Y_t é unidimensional quanto quando Y_t é um vetor de dimensão k.
  - Tempo (t) discreto, $t = 1, 2, \ldots, n$
- Modelos de espaço de estados são determinados por duas equações:

$$
\begin{aligned}
Y_t &= F_t(X_t, W_t) \text{ (Equação de observação)} \\
X_t &= G_t(X_{t-1}, V_t) \text{ (Equação de evolução),}
\end{aligned}
$$

onde X_t é um vetor de dimensão d e W_t e V_t são ruídos aleatórios.

Introdução

- **Equação de evolução** ($X_t = G_t(X_{t-1}, V_t)$):
  - Descreve a evolução temporal do estado X_t no sistema.
- **Equação de observação** ($Y_t = F_t(X_t, W_t)$):
  - A dimensão k de Y_t é, tipicamente, menor que a dimensão d de X_t.
- **Inferência sequencial**: Estima o estado presente X_t do sistema dinâmico utilizando o histórico $\mathcal{J}_t = (y_1, y_2, \ldots, y_t)$.

Introdução

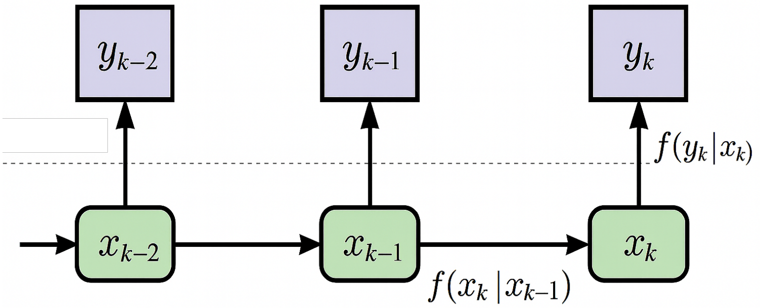

Figura: Estrutura de independência condicional.

Introdução

- Para qualquer tempo t e quaisquer valores $y_1, y_2, \ldots, y_t$, estamos interessados em encontrar estimativas para

  - O vetor de estados X_t no mesmo instante de tempo - **Problema de Filtragem**;

  - Os estados futuros X_{t+l} do sistema para $l > 0$ - Problema de Previsão;

  - Uma ocorrência passada do vetor de estados X_m, $m < t$ - Problema de Suavização.

Filtro de Kalman

Introdução

- Estimadores “ótimos”:
  - Perda absoluta: mediana;
  - Perda quadrática: média;
  - Perda 0-1: moda.
- O **filtro de Kalman** aplica a média condicional como estimador para o estado:

$$\hat{x}_t^+ = \mathrm{E}(X_t|\mathcal{J}_t) = \int x_t \ f(x_t|\mathcal{J}_t) \ dx_t.$$

Introdução

- Para uma melhor compreensão do estimador, exploramos a densidade $f(x_t|\mathcal{J}_t)$:

$$
\begin{aligned}
f(x_t|\mathcal{J}_t) &= \frac{f(\mathcal{J}_t|x_t)f(x_t)}{f(\mathcal{J}_t)} \text{ (Regra de Bayes)} \\
&= \frac{f(y_t,\mathcal{J}_{t-1}|x_t)f(x_t)}{f(y_t,\mathcal{J}_{t-1})} \; (\mathcal{J}_t = (y_t,\mathcal{J}_{t-1})) \\
&= \frac{f(y_t|\mathcal{J}_{t-1},x_t)f(\mathcal{J}_{t-1}|x_t)f(x_t)}{f(y_t|\mathcal{J}_{t-1})f(\mathcal{J}_{t-1})} \; (f(a,b) = f(a|b)f(b)) \\
&= \frac{f(y_t|\mathcal{J}_{t-1},x_t)[f(x_t|\mathcal{J}_{t-1})f(\mathcal{J}_{t-1})]f(x_t)}{f(y_t|\mathcal{J}_{t-1})[f(x_t)]f(\mathcal{J}_{t-1})} \text{ (Regra de Bayes)} \\
&= \frac{f(y_t|\mathcal{J}_{t-1},x_t)f(x_t|\mathcal{J}_{t-1})}{f(y_t|\mathcal{J}_{t-1})} \\
&= \frac{f(y_t|x_t)f(x_t|\mathcal{J}_{t-1})}{f(y_t|\mathcal{J}_{t-1})} \text{ (Independência condicional).} \qquad (1)
\end{aligned}
$$

Introdução

- A Equação (1) mostra que podemos computar a densidade desejada em dois passos:

  1. Computar a densidade para previsão de X_t condicional nas observações passadas:

  $$f(x_t|\mathcal{J}_{t-1}) = \int f(x_t|x_{t-1})f(x_{t-1}|\mathcal{J}_{t-1})dx_{t-1}.$$

  2. Atualizar a previsão via

  $$f(x_t|\mathcal{J}_t) = \frac{f(y_t|x_t)f(x_t|\mathcal{J}_{t-1})}{f(y_t|\mathcal{J}_{t-1})},$$

  onde $f(y_t|\mathcal{J}_{t-1}) = \int f(y_t|x_{t-1})f(x_{t-1}|\mathcal{J}_{t-1})dx_{t-1}$.

- A solução para o procedimento de inferência sequencial consiste em prever/atualizar.

Hipótese de Normalidade

- Soluções em forma fechada para as integrais anteriores não podem ser encontradas em muitos problemas reais.
  - Soluções simplificadas podem ser obtidas supondo que as densidades sejam gaussianas.
- Defina:
  - Erro de previsão: $\tilde{x}_t^- = x_t - \hat{x}_t^-$, onde $\hat{x}_t^- = \mathrm{E}(X_t|\mathcal{J}_{t-1})$.

$$\begin{aligned}
\mathrm{E}(\tilde{x}_t^-) &= \mathrm{E}(x_t) - \mathrm{E}(\hat{x}_t^-) = \mathrm{E}(x_t) - \mathrm{E}(\mathrm{E}(x_t|\mathcal{J}_{t-1})) = 0 \\
\mathrm{E}(\tilde{x}_t^-|\mathcal{J}_{t-1}) &= \mathrm{E}(x_t - \mathrm{E}(x_t|\mathcal{J}_{t-1})|\mathcal{J}_{t-1}) = \mathrm{E}(x_t|\mathcal{J}_{t-1}) - \mathrm{E}(x_t|\mathcal{J}_{t-1}) = 0 = \mathrm{E}(\tilde{x}_t^-)
\end{aligned}$$

  - Inovação: $\tilde{y}_t = y_t - \hat{y}_t$, onde $\hat{y}_t = \mathrm{E}(Y_t|\mathcal{J}_{t-1})$.

$$\mathrm{E}(\tilde{y}_t) = \mathrm{E}(y_t) - \mathrm{E}(\hat{y}_t) = \mathrm{E}(y_t) - \mathrm{E}(\mathrm{E}(y_t|\mathcal{J}_{t-1})) = 0$$

Hipótese de Normalidade

- Objeto de interesse: $\hat{x}_t^+ = \mathrm{E}(X_t|\mathcal{J}_t)$.

  A partir do erro de previsão, temos:

$$\mathrm{E}(\tilde{x}_t^-|\mathcal{J}_t) = \underbrace{\mathrm{E}(x_t|\mathcal{J}_t)}_{\hat{x}_t^+} - \underbrace{\mathrm{E}(\hat{x}_t^-|\mathcal{J}_t)}_{\hat{x}_t^-}.$$

  Em contrapartida, podemos escrever:

$$\mathrm{E}(\tilde{x}_t^-|\mathcal{J}_t) = \mathrm{E}(\tilde{x}_t^-|y_t, \mathcal{J}_{t-1}) = \mathrm{E}(\tilde{x}_t^-|y_t).$$

  Logo,

$$\hat{x}_t^+ = \hat{x}_t^- + \mathrm{E}(\tilde{x}_t^-|y_t), \tag{2}$$

  onde o procedimento de inferência sequencial consiste novamente em prever/atualizar.

Hipótese de Normalidade

- Na Equação (2), precisamos caracterizar $\mathrm{E}(\tilde{x}_t^- | y_t)$.
  - Assumiremos que a distribuição conjunta é normal.

Tomando

$$\begin{bmatrix} X \\ Y \end{bmatrix} \text{ com média } \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} \text{ e covariância } \begin{bmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_y \end{bmatrix},$$

temos:

$$f(x|y) \sim N(\mu_x + \Sigma_{xy}\Sigma_y^{-1}(y - \mu_y), \Sigma_x - \Sigma_{xy}\Sigma_y^{-1}\Sigma_{yx}).$$

Logo,

$$\mathrm{E}(X|Y) = \mu_x + \Sigma_{xy}\Sigma_y^{-1}(y - \mu_y).$$

Hipótese de Normalidade

- Retornando para $\mathrm{E}(\tilde{x}_t^-|y_t)$ e observando que $y_t = \tilde{y}_t + \hat{y}_t$, temos:

$$\begin{aligned}
\mathrm{E}(\tilde{x}_t^-|y_t) &= \mathrm{E}(\tilde{x}_t^-|\tilde{y}_t + \hat{y}_t) \\
&= \mathrm{E}(\tilde{x}_t^-) + \Sigma_{\tilde{x}\tilde{y},t}^{-}\Sigma_{\tilde{y},t}^{-1}(\tilde{y}_t + \hat{y}_t - \mathrm{E}(\tilde{y}_t + \hat{y}_t)) \\
&= 0 + \Sigma_{\tilde{x}\tilde{y},t}^{-}\Sigma_{\tilde{y},t}^{-1}(\tilde{y}_t + \hat{y}_t - (0 + \hat{y}_t)) \\
&= \underbrace{\Sigma_{\tilde{x}\tilde{y},t}^{-}\Sigma_{\tilde{y},t}^{-1}}_{L_t}\tilde{y}_t.
\end{aligned}$$

- Reescrevendo a Equação (2), temos a seguinte equação de atualização:

$$\hat{x}_t^+ = \hat{x}_t^- + L_t\tilde{y}_t.$$

Hipótese de Normalidade

- A incerteza associada a $\hat{x}_t^+$ pode ser mensurada através da matriz de covariância $\Sigma_{\tilde{x},t}^+$:

$$
\begin{aligned}
\Sigma_{\tilde{x},t}^+ &= \mathrm{E}(\tilde{x}_t^+ \tilde{x}_t^{+'}) \\
&= \mathrm{E}((x_t - \hat{x}_t^+)(x_t - \hat{x}_t^+)') \\
&= \mathrm{E}([(x_t - \hat{x}_t^-) - L_t \tilde{y}_t][(x_t - \hat{x}_t^-) - L_t \tilde{y}_t]') \\
&= \mathrm{E}((\tilde{x}_t^- - L_t \tilde{y}_t)(\tilde{x}_t^- - L_t \tilde{y}_t)') \\
&= \Sigma_{\tilde{x},t}^- - L_t \underbrace{\mathrm{E}(\tilde{y}_t \tilde{x}_t^{-'})}_{\Sigma_{\tilde{y},t} L_t'} - \underbrace{\mathrm{E}(\tilde{x}_t^- \tilde{y}_t')}_{L_t \Sigma_{\tilde{y},t}} L_t' + L_t \Sigma_{\tilde{y},t} L_t' \\
&= \Sigma_{\tilde{x},t}^- - L_t \Sigma_{\tilde{y},t} L_t'.
\end{aligned}
$$

Hipótese de Normalidade

- O Filtro de Kalman pode, portanto, ser descrito através dos seguintes passos:
    - **Definições**: $\tilde{x}_t^- = x_t - \hat{x}_t^-$ e $\tilde{y}_t = y_t - \hat{y}_t$, onde $\hat{x}_t^- = \mathrm{E}(x_t|\mathcal{J}_{t-1})$ e $\hat{y}_t = \mathrm{E}(y_t|\mathcal{J}_{t-1})$.
    - **Inicialização**: Para $t = 0$, $\hat{x}_0^+ = \mathrm{E}(x_0)$ e $\Sigma_{\tilde{x},0}^+ = \mathrm{E}((x_0 - \hat{x}_0^+)(x_0 - \hat{x}_0^+)')$.
    - **Inferência sequencial**: Para $t = 1, 2, \ldots, n$, calcule

1a. $\hat{x}_t^- = \mathrm{E}(x_t|\mathcal{J}_{t-1}) = \mathrm{E}(G_t(x_{t-1}, v_t)|\mathcal{J}_{t-1})$;

1b. $\Sigma_{\tilde{x},t}^- = \mathrm{E}(\tilde{x}_t^- \tilde{x}_t^{-'}) = \mathrm{E}((x_t - \hat{x}_t^-)(x_t - \hat{x}_t^-)')$;

1c. $\hat{y}_t = \mathrm{E}(y_t|\mathcal{J}_{t-1}) = \mathrm{E}(F_t(x_t, w_t)|\mathcal{J}_{t-1})$;

2a. $L_t = \mathrm{E}(\tilde{x}_t^- \tilde{y}_t')[\mathrm{E}(\tilde{y}_t \tilde{y}_t')]^{-1}$;

2b. $\hat{x}_t^+ = \hat{x}_t^- + L_t(y_t - \hat{y}_t)$

2c. $\Sigma_{\tilde{x},t}^+ = \mathrm{E}(\tilde{x}_t^+ \tilde{x}_t^{+'}) = \mathrm{E}((x_t - \hat{x}_t^+)(x_t - \hat{x}_t^+)') = \Sigma_{\tilde{x},t}^- - L_t \Sigma_{\tilde{y},t} L_t'$.

Hipótese de Linearidade

- **Equação de evolução** ($X_t = G_t(X_{t-1}, V_t)$):
  - $X_t = G_t(X_{t-1}, V_t) = G_t X_{t-1} + V_t$, onde G_t é uma matriz de dimensão $d \times d$.
  - V_t denota o erro de evolução:
    - Média 0 e covariância Σ_v;
    - Assume-se que V_t é independente de X_m para todo $m < t$.
  - Na maioria dos casos, a matriz G_t será independente de t.

Hipótese de Linearidade

- **Equação de observação** $(Y_t = F_t(X_t, W_t))$:
  - $Y_t = F_t(X_t, W_t) = F_t X_t + W_t$, onde F_t é uma matriz de dimensão $k \times k$.
  - W_t denota o erro observacional:
    - Média 0 e variância Σ_w;
    - Supõe-se que os erros observacionais e de evolução sejam independentes.
  - Na maioria dos casos, o vetor F_t será independente de t.

Hipótese de Linearidade

- No caso linear, pode ser provado que o Filtro de Kalman é ótimo (menor erro quadrático médio) e o estimador de máxima verossimilhança.
- V_t e W_t são ruídos brancos mutuamente independentes:

$$\mathrm{E}(W_t W_s') = \begin{cases} \Sigma_w, t = s \\ 0, t \neq s \end{cases} \qquad \mathrm{E}(V_t V_s') = \begin{cases} \Sigma_v, t = s \\ 0, t \neq s \end{cases}$$

.

- Apesar de as hipóteses de linearidade e independência não serem normalmente observadas na prática, há consenso, tanto na prática quanto na literatura, de que o método ainda assim funciona bastante bem.
- A seguir, traduziremos os passos $1a$, $1b$, $1c$, $2a$, $2b$ e $2c$ do Filtro de Kalman para o caso linear.

Hipótese de Linearidade

Filtro de Kalman: 1a.

$$\begin{aligned}
\hat{x}_t^- &= \mathrm{E}(x_t | \mathcal{J}_{t-1}) \\
&= \mathrm{E}(G_t x_{t-1} + v_t | \mathcal{J}_{t-1}) \\
&= \mathrm{E}(G_t x_{t-1} | \mathcal{J}_{t-1}) + \mathrm{E}(v_t | \mathcal{J}_{t-1}) \\
&= G_t \hat{x}_{t-1}^+
\end{aligned}$$

Filtro de Kalman: 1b.

$$\begin{aligned}
\tilde{x}_t^- &= x_t - \hat{x}_t^- \\
&= G_t x_{t-1} + v_t - G_t \hat{x}_{t-1}^+ \\
&= G_t \tilde{x}_{t-1}^+ + v_t \\
\Sigma_{\tilde{x},t}^- &= \mathrm{E}((G_t \tilde{x}_{t-1}^+ + v_t)(G_t \tilde{x}_{t-1}^+ + v_t)') \\
&= \mathrm{E}(G_t \tilde{x}_{t-1}^+ \tilde{x}_{t-1}^{+'} G_t' + G_t \tilde{x}_{t-1}^+ v_t' + v_t \tilde{x}_{t-1}^{+'} G_t' + v_t v_t') \\
&= G_t \Sigma_{\tilde{x},t-1}^+ G_t' + \Sigma_v
\end{aligned}$$

Hipótese de Linearidade

Filtro de Kalman: $1c$.

$$
\begin{aligned}
\hat{y}_t &= \mathrm{E}(F_t x_t + w_t | \mathcal{J}_{t-1}) \\
&= \mathrm{E}(F_t x_t | \mathcal{J}_{t-1}) + \mathrm{E}(w_t | \mathcal{J}_{t-1}) \\
&= F_t \hat{x}_t^-
\end{aligned}
$$

Filtro de Kalman: $2a$. ($L_t = \Sigma^-_{\tilde{x}\tilde{y},t} \Sigma^{-1}_{\tilde{y},t}$).
Primeiro, focaremos em $\Sigma_{\tilde{y},t}$:

$$
\begin{aligned}
\tilde{y}_t &= y_t - \hat{y}_t \\
&= F_t x_t + w_t - F_t \hat{x}_t^- \\
&= F_t \tilde{x}_t^- + w_t \\
\Sigma_{\tilde{y},t} &= \mathrm{E}((F_t \tilde{x}_t^- + w_t)(F_t \tilde{x}_t^- + w_t)') \\
&= \mathrm{E}(F_t \tilde{x}_t^- \tilde{x}_t^{-\prime} F_t' + F_t \tilde{x}_t^- w_t' + w_t \tilde{x}_t^{-\prime} F_t' + w_t w_t') \\
&= F_t \Sigma^-_{\tilde{x},t} F_t' + \Sigma_w
\end{aligned}
$$

Hipótese de Linearidade

Para $\Sigma^{-}_{\tilde{x}\tilde{y},t}$, temos:

$$\begin{aligned}
\Sigma^{-}_{\tilde{x}\tilde{y},t} &= \mathrm{E}(\tilde{x}^{-}_{t}\tilde{y}'_{t}) \\
&= \mathrm{E}(\tilde{x}^{-}_{t}(F_t\tilde{x}^{-}_{t} + w_t)') \\
&= \mathrm{E}(\tilde{x}^{-}_{t}\tilde{x}^{-\prime}_{t}F'_t + \tilde{x}^{-}_{t}w'_t) \\
&= \Sigma^{-}_{\tilde{x},t}F'_t
\end{aligned}$$

Por fim,

$$L_t = \Sigma^{-}_{\tilde{x},t}F'_t[F_t\Sigma^{-}_{\tilde{x},t}F'_t + \Sigma_w]^{-1}$$

Filtro de Kalman: $2b$.

$$\hat{x}^{+}_{t} = \hat{x}^{-}_{t} + L_t(y_t - \hat{y}_t)$$

Hipótese de Linearidade

Filtro de Kalman: $2c$.

$$
\begin{aligned}
\Sigma^{+}_{\tilde{x},t} &= \Sigma^{-}_{\tilde{x},t} - L_t \Sigma_{\tilde{y},t} L'_t \\
&= \Sigma^{-}_{\tilde{x},t} - L_t \Sigma_{\tilde{y},t} \Sigma^{-1}_{\tilde{y},t} \Sigma^{-\prime}_{\tilde{x}\tilde{y},t} \\
&= \Sigma^{-}_{\tilde{x},t} - L_t F_t \Sigma^{-}_{\tilde{x},t} \\
&= (I - L_t F_t) \Sigma^{-}_{\tilde{x},t}
\end{aligned}
$$

Hipótese de Linearidade

- No caso linear, o Filtro de Kalman pode, portanto, ser descrito através dos seguintes passos:
  - **Definições**: $\tilde{x}_t^- = x_t - \hat{x}_t^-$ e $\tilde{y}_t = y_t - \hat{y}_t$, onde $\hat{x}_t^- = \mathrm{E}(x_t|\mathcal{J}_{t-1})$ e $\hat{y}_t = \mathrm{E}(y_t|\mathcal{J}_{t-1})$.
  - **Inicialização**: Para $t = 0$, $\hat{x}_0^+ = \mathrm{E}(x_0)$ e $\Sigma_{\tilde{x},0}^+ = \mathrm{E}((x_0 - \hat{x}_0^+)(x_0 - \hat{x}_0^+)')$.
  - **Inferência sequencial**: Para $t = 1, 2, \ldots, n$, calcule

    1a. $\hat{x}_t^- = G_t \hat{x}_{t-1}^+$;

    1b. $\Sigma_{\tilde{x},t}^- = G_t \Sigma_{\tilde{x},t-1}^+ G_t' + \Sigma_v$;

    1c. $\hat{y}_t = F_t \hat{x}_t^-$;

    2a. $L_t = \Sigma_{\tilde{x},t}^- F_t' [F_t \Sigma_{\tilde{x},t}^- F_t' + \Sigma_w]^{-1}$;

    2b. $\hat{x}_t^+ = \hat{x}_t^- + L_t(y_t - \hat{y}_t)$

    2c. $\Sigma_{\tilde{x},t}^+ = (I - L_t F_t)\Sigma_{\tilde{x},t}^-$.

Modelos relacionados

Regressão Linear

- Considere a notação a seguir:
  - Y_t: série temporal univariada;
  - $X_{t1}, X_{t2}, \ldots, X_{td}$: vetor de variáveis independentes (observável);
  - $X_{t1} = 1$: termo constante para todo t;
  - $\beta_t = (\beta_{t1}, \beta_{t2}, \ldots, \beta_{td})'$: vetor de coeficientes da regressão (não observável).

Regressão Linear

- Modelo de regressão linear com coeficientes variando no tempo:

$$\begin{aligned} Y_t &= F_t'\beta_t + W_t \\ \beta_t &= G\beta_{t-1} + V_t, \end{aligned}$$

onde

- $F_t = (X_{t1}, X_{t2}, \ldots, X_{td})'$ e $G = I_d$;
- V_t é um escalar;
- W_t é uma matriz de dimensão $d \times s$.

Capital Asset Pricing Model (CAPM)

- Considere a notação a seguir:
  - $R_t^{(i)}$: retornos do ativo i no tempo t;
  - $R_t^{(m)}$: retorno do mercado no tempo t - vetor com d ações.
- O modelo CAPM assume:

$$R_t^{(i)} = F_t'\beta + W_t,$$

onde $F_t = R_t^{(m)}$.

- Modelo CAPM com coeficientes variando no tempo:

$$\begin{aligned} Y_t &= F_t'\beta_t + W_t \\ \beta_t &= G\beta_{t-1} + V_t, \end{aligned}$$

onde $G = I_d$.

Modelo Autorregressivo de ordem p - AR(p)

Seja Y_t uma série temporal univariada, temos:

$$
\begin{aligned}
Y_t &= F'X_t + W_t \\
X_t &= GX_{t-1} + V_t
\end{aligned}
$$

onde

- $X_t = (X_{t-1}, X_{t-2}, \ldots, X_{t-p})'$;
- $F = (1, \underbrace{0, \ldots, 0}_{p-1})'$ e $G = \begin{bmatrix} \phi_1 & \phi_2 & \phi_3 & \ldots & \phi_{p-1} & \phi_p \\ 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & 0 & \ldots & 1 & 0 \end{bmatrix}$.

Modelo Autorregressivo de ordem p - AR(p)

Modelo com nível variando no tempo:

- $X_t = (\mu_t, X_{t-1}, X_{t-2}, \ldots, X_{t-p})'$;
- $F = (1, 1, \underbrace{0, \ldots, 0}_{p-1})'$ e $G = \begin{bmatrix} 1 & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & \phi_1 & \phi_2 & \phi_3 & \ldots & \phi_{p-1} & \phi_p \\ 0 & 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & \ldots & 1 & 0 \end{bmatrix}$.

Modelo Autorregressivo de ordem p - AR(p)

Modelo com nível e tendência variando no tempo:

- $X_t = (\mu_t, \delta_t, X_{t-1}, X_{t-2}, \ldots, X_{t-p})'$;
- $F = (1, 0, 1, \underbrace{0, \ldots, 0}_{p-1})'$ e $G = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & \phi_1 & \phi_2 & \phi_3 & \ldots & \phi_{p-1} & \phi_p \\ 0 & 0 & 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \ldots & 1 & 0 \end{bmatrix}$.