Capítulo 6 - Propriedades Estatísticas de Estimadores

Eduardo Mendes

emmendes@cpdee.ufmg.br

Departamento de Engenharia Eletrônica
Universidade Federal de Minas Gerais
Av. Antônio Carlos 6627, Belo Horizonte, MG, Brasil

Introdução

Seja o sinal medido a superposição do sinal ideal com o ruído

$$y(k) = y^{\mathrm{i}}(k) + e(k).$$

Após vários testes, ainda que $y^{\mathrm{i}}(k)$ permaneça o mesmo, é de se esperar que parâmetros estimados usando-se $y(k)$ também variem. Se tais parâmetros forem considerados como variáveis aleatórias, quais seriam as suas características estatísticas?

Polarização de Estimadores — Conceitos

A polarização (*bias*) é definida como

$$\mathbf{b} = \mathrm{E}[\hat{\boldsymbol{\theta}}] - \boldsymbol{\theta}.$$

Polarização de um estimador de ganho

- Considere

$$y(k) = Ku(k) + e(k), \quad k = 1, 2, \ldots N,$$

sendo que $u(k)$ é a entrada do sistema e K é o ganho.

Polarização de um estimador de ganho

- Considere

$$y(k) = Ku(k) + e(k), \quad k = 1, 2, \dots N,$$

sendo que $u(k)$ é a entrada do sistema e K é o ganho.

- Supondo que o ruído não afete a entrada $u(k)$, uma forma de se estimar o ganho K a partir dos dados de entrada e saída do sistema é calcular

$$\hat{K} = \frac{1}{N}\left(\sum_{k=1}^{N} \frac{y(k)}{u(k)}\right).$$

- A polarização deste estimador pode ser determinada a partir da definição como se segue:

$$
\begin{aligned}
b &= \mathrm{E}[\hat{K}] - K = \frac{1}{N}\sum_{k=1}^{N}\left(\frac{\mathrm{E}[Ku(k)+e(k)]}{u(k)}\right) - K \\
&= \frac{1}{N}\sum_{k=1}^{N}\left(K + \frac{\mathrm{E}[e(k)]}{u(k)}\right) - K \\
&= \frac{1}{N}\sum_{k=1}^{N}\frac{\overline{e(k)}}{u(k)}.
\end{aligned}
$$

Polarização - modelo $y(k) = \boldsymbol{\psi}^{\mathrm{T}}(k-1)\boldsymbol{\theta} + e(k)$

- Será considerada a polarização do estimador do vetor de parâmetros de um modelo do tipo $y(k) = \boldsymbol{\psi}^{\mathrm{T}}(k-1)\boldsymbol{\theta} + e(k)$, que é linear em $\boldsymbol{\theta}$. Um modelo deste tipo gera um conjunto de equações do tipo $\mathbf{y} = \Psi\boldsymbol{\theta} + \mathbf{e}$, e

$$\hat{\boldsymbol{\theta}} \;=\; A\mathbf{y},$$

  sendo que Ψ é a chamada *matriz de regressores* e A é uma matriz cujos elementos dependem dos regressores.

Polarização - modelo $y(k) = \boldsymbol{\psi}^{\mathrm{T}}(k-1)\boldsymbol{\theta} + e(k)$

- Será considerada a polarização do estimador do vetor de parâmetros de um modelo do tipo $y(k) = \boldsymbol{\psi}^{\mathrm{T}}(k-1)\boldsymbol{\theta} + e(k)$, que é linear em $\boldsymbol{\theta}$. Um modelo deste tipo gera um conjunto de equações do tipo $\mathbf{y} = \Psi\boldsymbol{\theta} + \mathbf{e}$, e

$$\hat{\boldsymbol{\theta}} = A\mathbf{y},$$

sendo que Ψ é a chamada *matriz de regressores* e A é uma matriz cujos elementos dependem dos regressores.

- Então, tem-se

$$\begin{aligned}
\mathbf{b} &= \mathrm{E}[A\mathbf{y}] - \boldsymbol{\theta}, \\
&= \mathrm{E}[A(\Psi\boldsymbol{\theta} + \mathbf{e})] - \boldsymbol{\theta}, \\
&= \mathrm{E}[(A\Psi - I)\boldsymbol{\theta}] + \mathrm{E}[A\mathbf{e}], \\
&= \mathrm{E}[A\Psi - I]\boldsymbol{\theta} + \mathrm{E}[A\mathbf{e}],
\end{aligned}$$

sendo que $\boldsymbol{\theta}$ foi considerado determinístico.

Para que a polarização seja nula ($\mathbf{b} = 0$) as seguintes considerações precisam ser satisfeitas:

1) A matriz A deve ser tal que $A\Psi = I$.

Para que a polarização seja nula ($\mathbf{b} = 0$) as seguintes considerações precisam ser satisfeitas:

1) A matriz A deve ser tal que $A\Psi = I$.

2) Os elementos de A não estão correlacionados com o vetor de erro **e**, logo $\mathrm{E}[A\mathbf{e}] = \mathrm{E}[A]\mathrm{E}[\mathbf{e}]$.

Para que a polarização seja nula ($\mathbf{b} = 0$) as seguintes considerações precisam ser satisfeitas:

1) A matriz A deve ser tal que $A\Psi = I$.
2) Os elementos de A não estão correlacionados com o vetor de erro **e**, logo $\mathrm{E}[A\mathbf{e}] = \mathrm{E}[A]\mathrm{E}[\mathbf{e}]$.
3) O erro deve ter média zero, isto é, $\mathrm{E}[\mathbf{e}] = 0$.

Polarização do Estimador MQ

- O estimador de mínimos quadrados é um estimador linear do tipo $\hat{\boldsymbol{\theta}} = A\mathbf{y}$, sendo que

$$A = \left[\Psi^{\mathrm{T}}\Psi\right]^{-1}\Psi^{\mathrm{T}}.$$

Polarização do Estimador MQ

- O estimador de mínimos quadrados é um estimador linear do tipo $\hat{\boldsymbol{\theta}} = A\mathbf{y}$, sendo que

$$A = \left[\Psi^{\mathrm{T}}\Psi\right]^{-1}\Psi^{\mathrm{T}}.$$

- A condição mais relevante para o estimador de mínimos quadrados é a 7-a.

Polarização do Estimador MQ

- O estimador de mínimos quadrados é um estimador linear do tipo $\hat{\boldsymbol{\theta}} = A\mathbf{y}$, sendo que

$$A = \left[\Psi^{\mathrm{T}}\Psi\right]^{-1}\Psi^{\mathrm{T}}.$$

- A condição mais relevante para o estimador de mínimos quadrados é a 7-a.

Polarização do estimador de MQP

Usando a definição de polarização

$$\begin{aligned}
\mathbf{b} &= \mathrm{E}[\hat{\boldsymbol{\theta}}_{\mathrm{w}}] - \boldsymbol{\theta}, \\
&= \mathrm{E}[A_{\mathrm{w}}\mathbf{y}] - \boldsymbol{\theta}, \\
&= \mathrm{E}[A_{\mathrm{w}}(\Psi\boldsymbol{\theta} + \mathbf{e})] - \boldsymbol{\theta}, \\
&= \mathrm{E}[(A_{\mathrm{w}}\Psi - I)\boldsymbol{\theta}] + \mathrm{E}[A_{\mathrm{w}}\mathbf{e}], \\
&= \mathrm{E}[A_{\mathrm{w}}\Psi - I]\boldsymbol{\theta} + \mathrm{E}[A_{\mathrm{w}}\mathbf{e}],
\end{aligned}$$

sendo que $A_{\mathrm{w}} = [\Psi^{\mathrm{T}} W \Psi]^{-1}\, \Psi^{\mathrm{T}} W$ e W é uma matriz de pesos constantes. As mesmas considerações 7 a 7-a, mas com $A = A_{\mathrm{w}}$, devem ser satisfeitas para garantir que o vetor de parâmetros estimados, $\hat{\boldsymbol{\theta}}_{\mathrm{w}}$, seja não polarizado.

Uma interpretação de polarização

- Se o estimador MQ for usado, o vetor de resíduos será, $\boldsymbol{\xi} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \Psi\hat{\boldsymbol{\theta}}_{\mathrm{MQ}}$. Logo,

$$\begin{aligned}
\boldsymbol{\xi} &= \mathbf{y} - \Psi\hat{\boldsymbol{\theta}}_{\mathrm{MQ}} \\
&= \Psi\boldsymbol{\theta} + \mathbf{e} - \Psi\hat{\boldsymbol{\theta}}_{\mathrm{MQ}} \\
&= \Psi(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{\mathrm{MQ}}) + \mathbf{e}.
\end{aligned}$$

Uma interpretação de polarização

- Se o estimador MQ for usado, o vetor de resíduos será, $\boldsymbol{\xi} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \Psi\hat{\boldsymbol{\theta}}_{\mathrm{MQ}}$. Logo,

$$
\begin{aligned}
\boldsymbol{\xi} &= \mathbf{y} - \Psi\hat{\boldsymbol{\theta}}_{\mathrm{MQ}} \\
&= \Psi\boldsymbol{\theta} + \mathbf{e} - \Psi\hat{\boldsymbol{\theta}}_{\mathrm{MQ}} \\
&= \Psi(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{\mathrm{MQ}}) + \mathbf{e}.
\end{aligned}
$$

- Os resíduos são ortogonais aos regressores, portanto

$$
\begin{aligned}
\Psi^{\mathrm{T}}\boldsymbol{\xi} &= \mathbf{0} \\
&= \Psi^{\mathrm{T}}\Psi(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{\mathrm{MQ}}) + \Psi^{\mathrm{T}}\mathbf{e},
\end{aligned}
$$

e

$$
\begin{aligned}
\Psi^{\mathrm{T}}\mathbf{e} &= \Psi^{\mathrm{T}}\Psi(\hat{\boldsymbol{\theta}}_{\mathrm{MQ}} - \boldsymbol{\theta}) \\
\mathrm{E}[\Psi^{\mathrm{T}}\mathbf{e}] &= \mathrm{E}[\Psi^{\mathrm{T}}\Psi(\hat{\boldsymbol{\theta}}_{\mathrm{MQ}} - \boldsymbol{\theta})] \\
&= \Psi^{\mathrm{T}}\Psi\mathrm{E}[(\hat{\boldsymbol{\theta}}_{\mathrm{MQ}} - \boldsymbol{\theta})],
\end{aligned}
$$

- de onde vem que

$$\mathbf{b}_{\mathrm{MQ}} \quad = \quad [\Psi^{\mathrm{T}}\Psi]^{-1}\mathrm{E}[\Psi^{\mathrm{T}}\mathbf{e}].$$

- de onde vem que

$$\mathbf{b}_{\mathrm{MQ}} = [\Psi^{\mathrm{T}}\Psi]^{-1}\mathrm{E}[\Psi^{\mathrm{T}}\mathbf{e}].$$

- É necessário que o vetor de erro seja ortogonal à hipersuperfície gerada pelos regressores

$$\Psi^{\mathrm{T}}\mathbf{e} = \mathbf{0}.$$

Polarização devida à tendência

- Seja $y(k) = \boldsymbol{\psi}^{\mathrm{T}}(k-1)\boldsymbol{\theta} + e(k)$. Supondo-se que o erro seja

$$e(k) = \alpha k + \beta + \nu(k),$$

sendo que $\nu(k)$ é uma variável aleatória não autocorrelacionada (ruído branco) e α e β são constantes. Portanto, $e(k)$ não tem média nula.

Polarização devida à tendência

- Seja $y(k) = \boldsymbol{\psi}^{\mathrm{T}}(k-1)\boldsymbol{\theta} + e(k)$. Supondo-se que o erro seja

$$e(k) = \alpha k + \beta + \nu(k),$$

  sendo que $\nu(k)$ é uma variável aleatória não autocorrelacionada (ruído branco) e α e β são constantes. Portanto, $e(k)$ não tem média nula.

- O valor da saída medido no instante k é

$$y(k) = \boldsymbol{\psi}^{\mathrm{T}}(k-1)\boldsymbol{\theta} + \alpha k + \beta + \nu(k).$$

Polarização devida à tendência

- Seja $y(k) = \boldsymbol{\psi}^{\mathrm{T}}(k-1)\boldsymbol{\theta} + e(k)$. Supondo-se que o erro seja

$$e(k) = \alpha k + \beta + \nu(k),$$

  sendo que $\nu(k)$ é uma variável aleatória não autocorrelacionada (ruído branco) e α e β são constantes. Portanto, $e(k)$ não tem média nula.

- O valor da saída medido no instante k é

$$y(k) = \boldsymbol{\psi}^{\mathrm{T}}(k-1)\boldsymbol{\theta} + \alpha k + \beta + \nu(k).$$

- Suponha que um dos regressores seja $y(k-1)$. Por comparação com o caso anterior, tem-se

$$y(k-1) = \boldsymbol{\psi}^{\mathrm{T}}(k-2)\boldsymbol{\theta} + e(k-1),$$

- mas há uma relação clara entre $e(k)$ e $e(k-1)$ que é

$$
\begin{aligned}
e(k-1) &= \alpha[k-1] + \beta + \nu(k-1) \\
&= \alpha k + \beta + \nu(k) + \nu(k-1) - \alpha - \nu(k) \\
&= e(k) + \nu(k-1) - \alpha - \nu(k).
\end{aligned}
$$

- mas há uma relação clara entre $e(k)$ e $e(k-1)$ que é

$$\begin{aligned} e(k-1) &= \alpha[k-1] + \beta + \nu(k-1) \\ &= \alpha k + \beta + \nu(k) + \nu(k-1) - \alpha - \nu(k) \\ &= e(k) + \nu(k-1) - \alpha - \nu(k). \end{aligned}$$

- O regressor $y(k-1)$ está correlacionado com $e(k)$, logo, $\mathrm{E}[\Psi^{\mathrm{T}}\mathbf{e}] \neq 0$.

Sinais reais com tendência

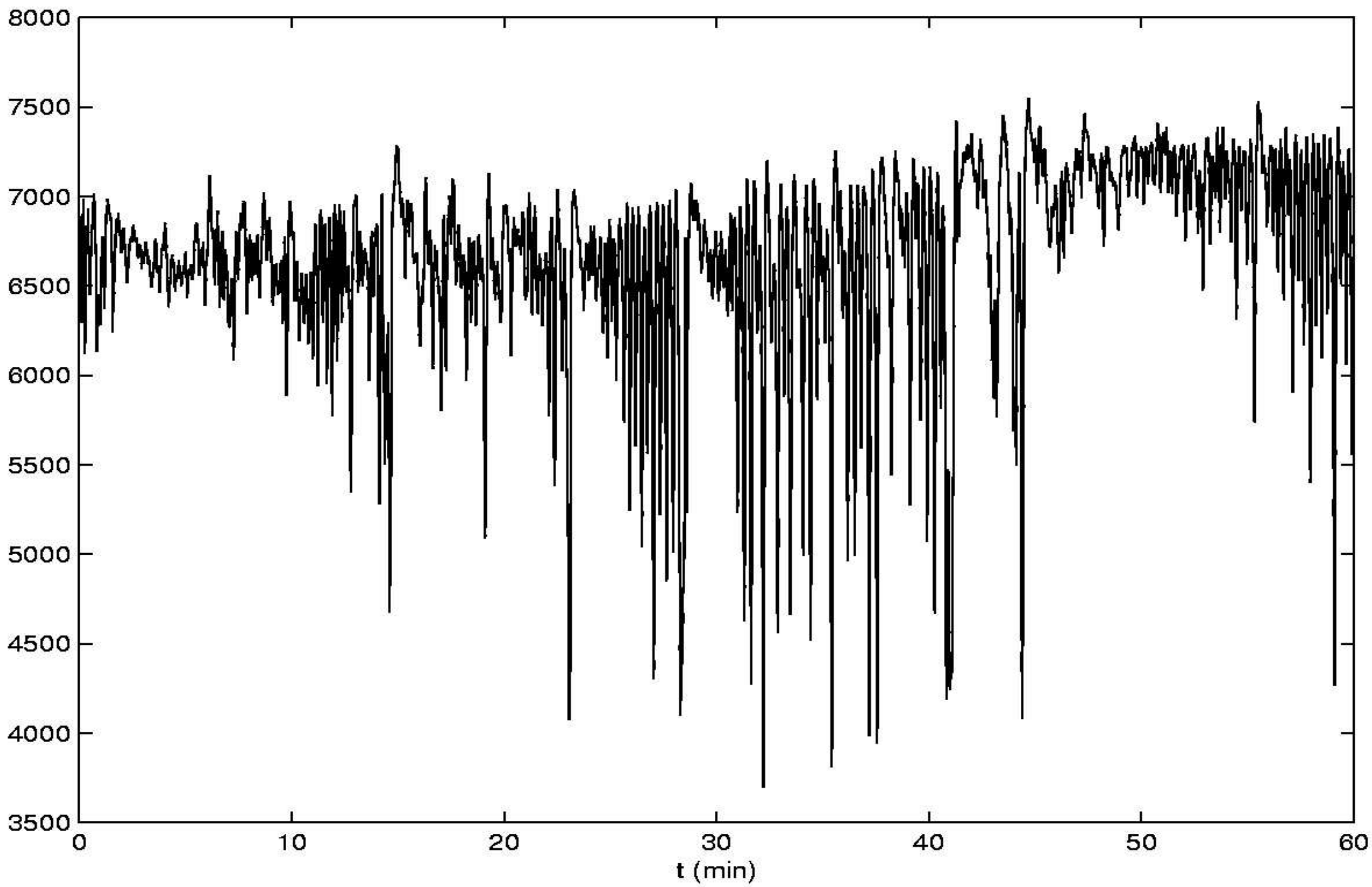

Figura 1: Sinais reais com tendência Série de saturação de oxigênio dissolvido no sangue de um humano. A escala vertical está em unidades arbitrárias

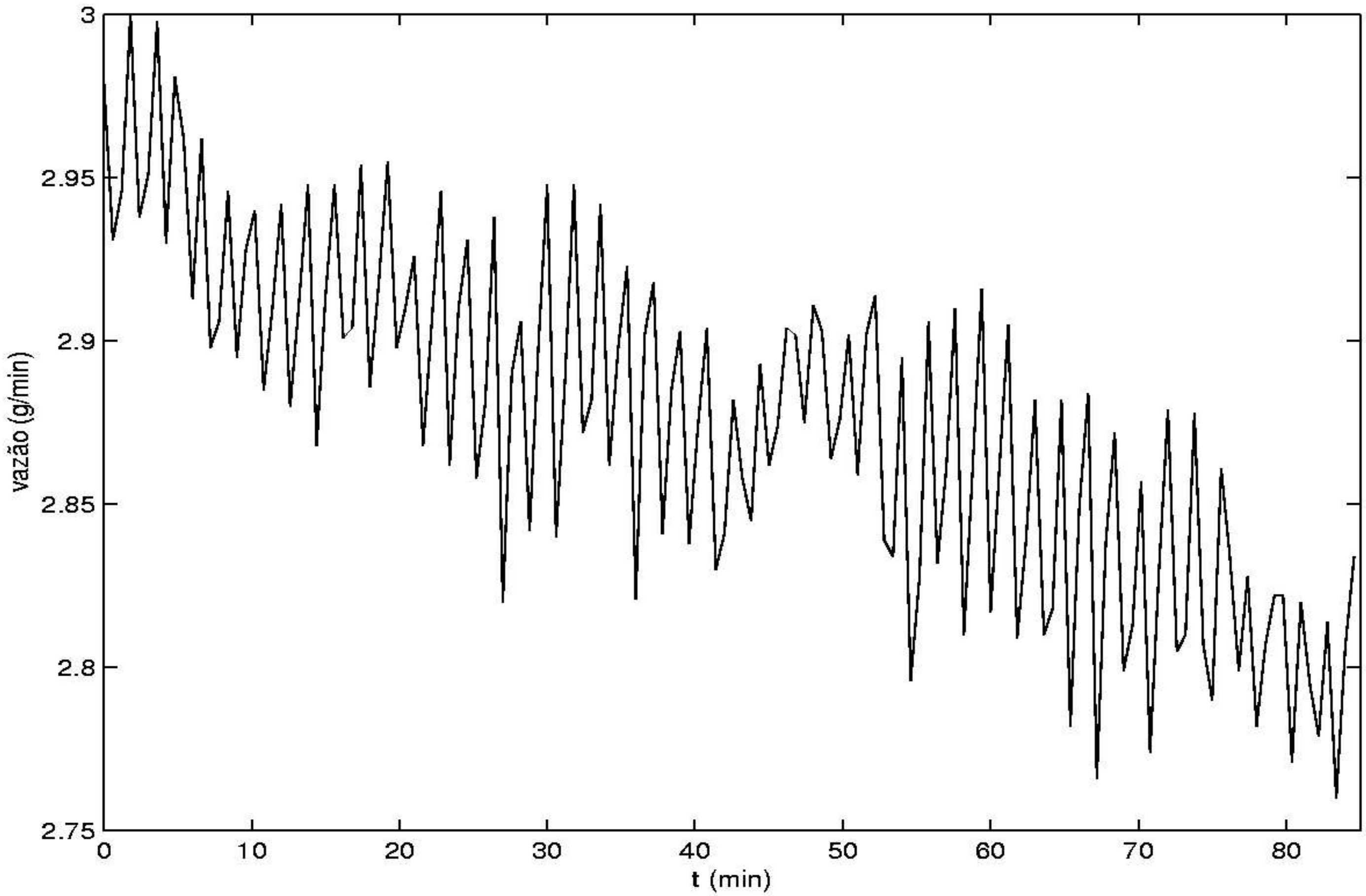

Figura 2: Sinais reais com tendência - Série de vazão de ar numa planta piloto de flotação em coluna.

Polarização em modelos ARX

- Seja um modelo ARX

$$
\begin{aligned}
\boldsymbol{\psi}^{\mathrm{T}}(k-1) &= [y(k-1)\ldots y(k-n_y)\ u(k-1)\ldots u(k-n_u)] \\
\boldsymbol{\theta} &= \left[a_1\ldots a_{n_y}\ b_1\ldots b_{n_u}\right]^{\mathrm{T}}.
\end{aligned}
$$

Polarização em modelos ARX

- Seja um modelo ARX

$$
\begin{aligned}
\boldsymbol{\psi}^{\mathrm{T}}(k-1) &= [y(k-1) \ldots y(k-n_y)\; u(k-1) \ldots u(k-n_u)] \\
\boldsymbol{\theta} &= \left[a_1 \ldots a_{n_y}\; b_1 \ldots b_{n_u}\right]^{\mathrm{T}}.
\end{aligned}
$$

- Aplicando-se o modelo acima ao longo dos dados, gera-se a equação matricial $\mathbf{y} = \Psi\boldsymbol{\theta} + \boldsymbol{\nu}$. A condição para que o estimador MQ seja não polarizado é que os regressores e $\boldsymbol{\nu}$ não sejam correlacionados.

É conveniente escrever o modelo na forma

$$F(q)y(k) = B(q)u(k) + e(k),$$

sendo que $e(k) = F(q)\nu(k)$ e $\nu(k)$ é branco.

- A menos que $F(q) = 1$, $e(k)$ não será ruído branco.

É conveniente escrever o modelo na forma

$$F(q)y(k) = B(q)u(k) + e(k),$$

sendo que $e(k) = F(q)\nu(k)$ e $\nu(k)$ é branco.

- A menos que $F(q) = 1$, $e(k)$ não será ruído branco.
- Em modelos ARX erros de regressão brancos não resultam em polarização do estimador MQ.

É conveniente escrever o modelo na forma

$$F(q)y(k) = B(q)u(k) + e(k),$$

sendo que $e(k) = F(q)\nu(k)$ e $\nu(k)$ é branco.

- A menos que $F(q) = 1$, $e(k)$ não será ruído branco.
- Em modelos ARX erros de regressão brancos não resultam em polarização do estimador MQ.
- Ao passo que a adição de ruído branco em modelos de erro na saída (ruído de medição) acarreta em polarização do estimador de MQ.

Polarização em modelos de erro na saída e ARX

- O modelo ARX simulado foi

$$y(k) = 1,5y(k-1) - 0,7y(k-2) + 0,5u(k-1) + \nu(k),$$

sendo que $u(k)$ e $\nu(k)$ são sinais aleatórios com distribuição normal, média nula e variâncias $\sigma_u^2 = 1$ e $\sigma_\nu^2 = 0,04$.

Polarização em modelos de erro na saída e ARX

- O modelo ARX simulado foi

$$y(k) = 1,5y(k-1) - 0,7y(k-2) + 0,5u(k-1) + \nu(k),$$

sendo que $u(k)$ e $\nu(k)$ são sinais aleatórios com distribuição normal, média nula e variâncias $\sigma_u^2 = 1$ e $\sigma_\nu^2 = 0,04$.

- O modelo de erro na saída usado foi

$$\begin{aligned} w(k) &= 1,5w(k-1) - 0,7w(k-2) + 0,5u(k-1) \\ y(k) &= w(k) + \nu(k), \end{aligned}$$

sendo $u(k)$ e $\nu(k)$ definidos da mesma forma que para o modelo ARX.

- Observar que a simulação do modelo de erro na saída corresponde a simular a função de transferência

$$G(q) = \frac{B(q)}{F(q)} = \frac{0,5q^{-1}}{1 - 1,5q^{-1} + 0,7q^{-2}},$$

excitada pela entrada $u(k)$, gerando assim $w(k)$ e posteriormente acrescentar o ruído $\nu(k)$, ponto a ponto.

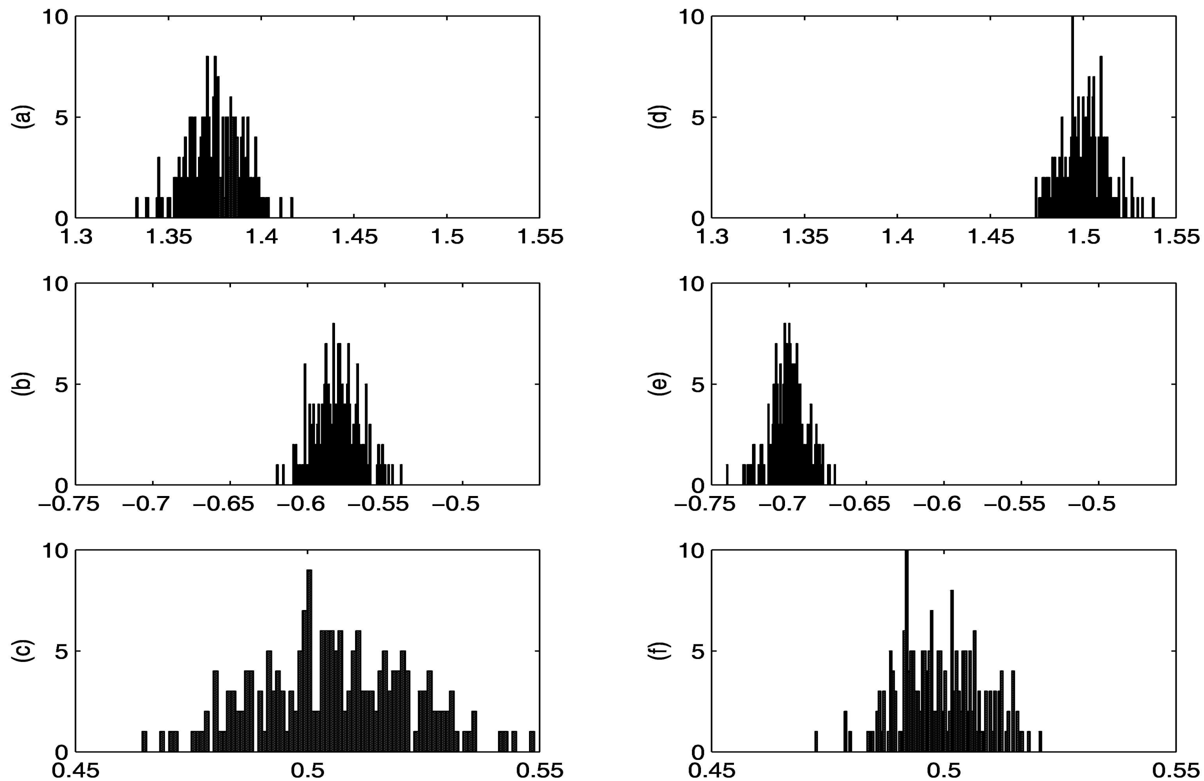

Figura 3: Polarização em modelos de erro na saída e ARX - Histogramas de parâmetros estimados usando-se o estimador MQ. Os gráficos (a), (b) e (c) se referem ao modelo de erro na saída e os gráficos (d), (e) e (f) se referem ao modelo ARX. Os valores reais dos parâmetros são: (a) e (d) $\theta_1 = 1,5$; (b) e (e) $\theta_2 = -0,7$ e (c) e (f) $\theta_3 = 0,5$.

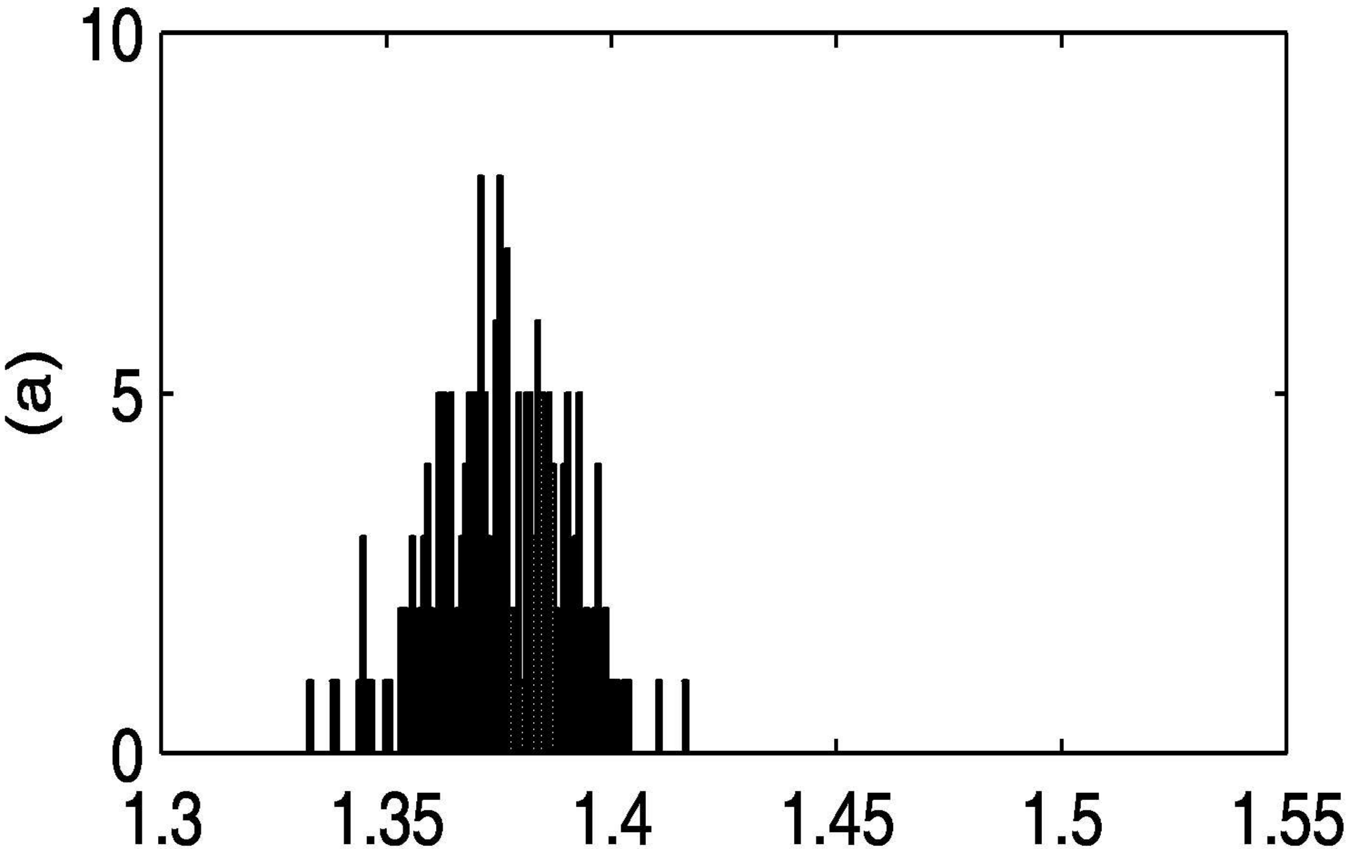

Figura 4: Polarização em modelo erro na saída - $\theta_1 = 1,5$

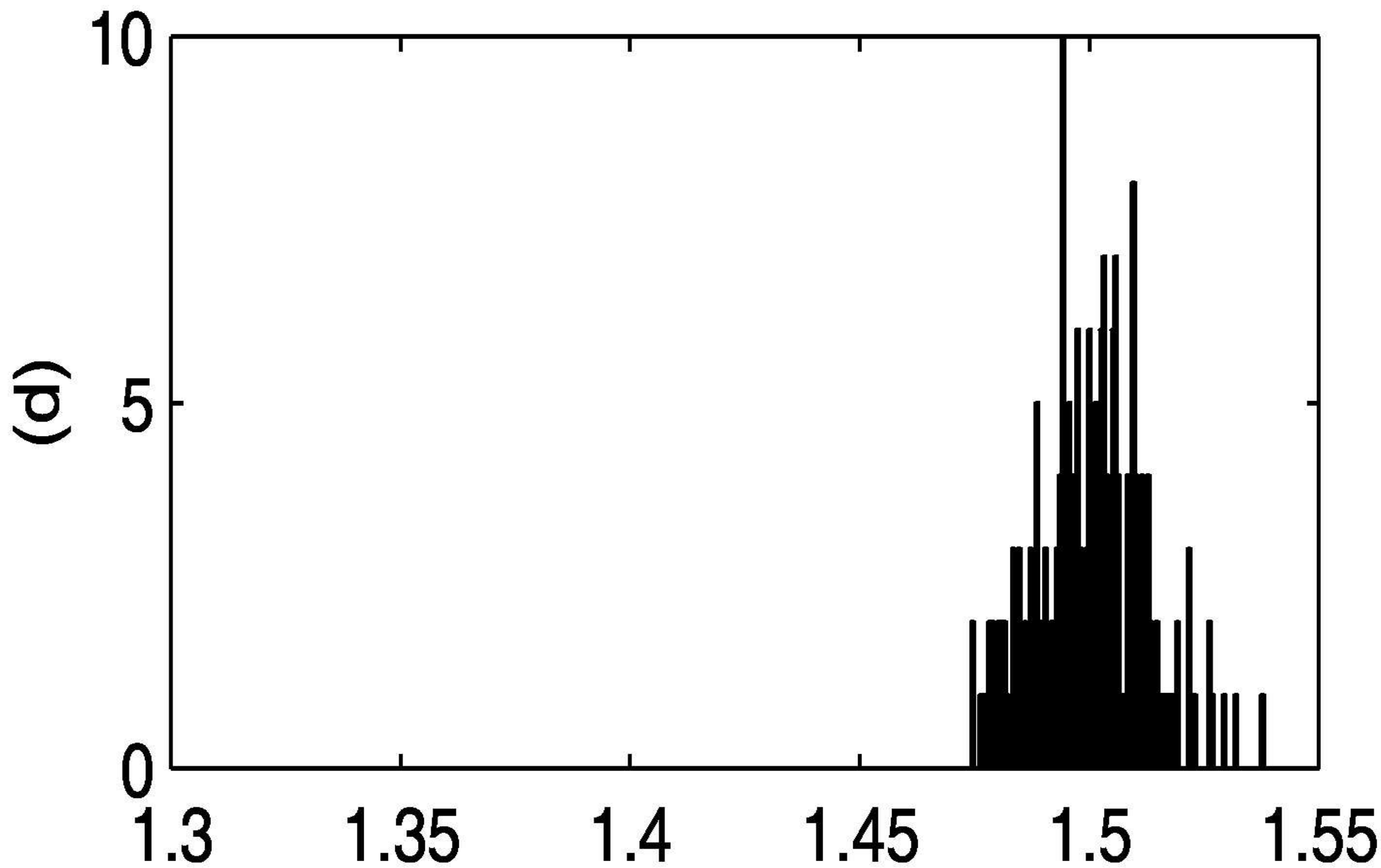

Figura 5: Polarização em modelo ARX - $\theta_1 = 1,5$

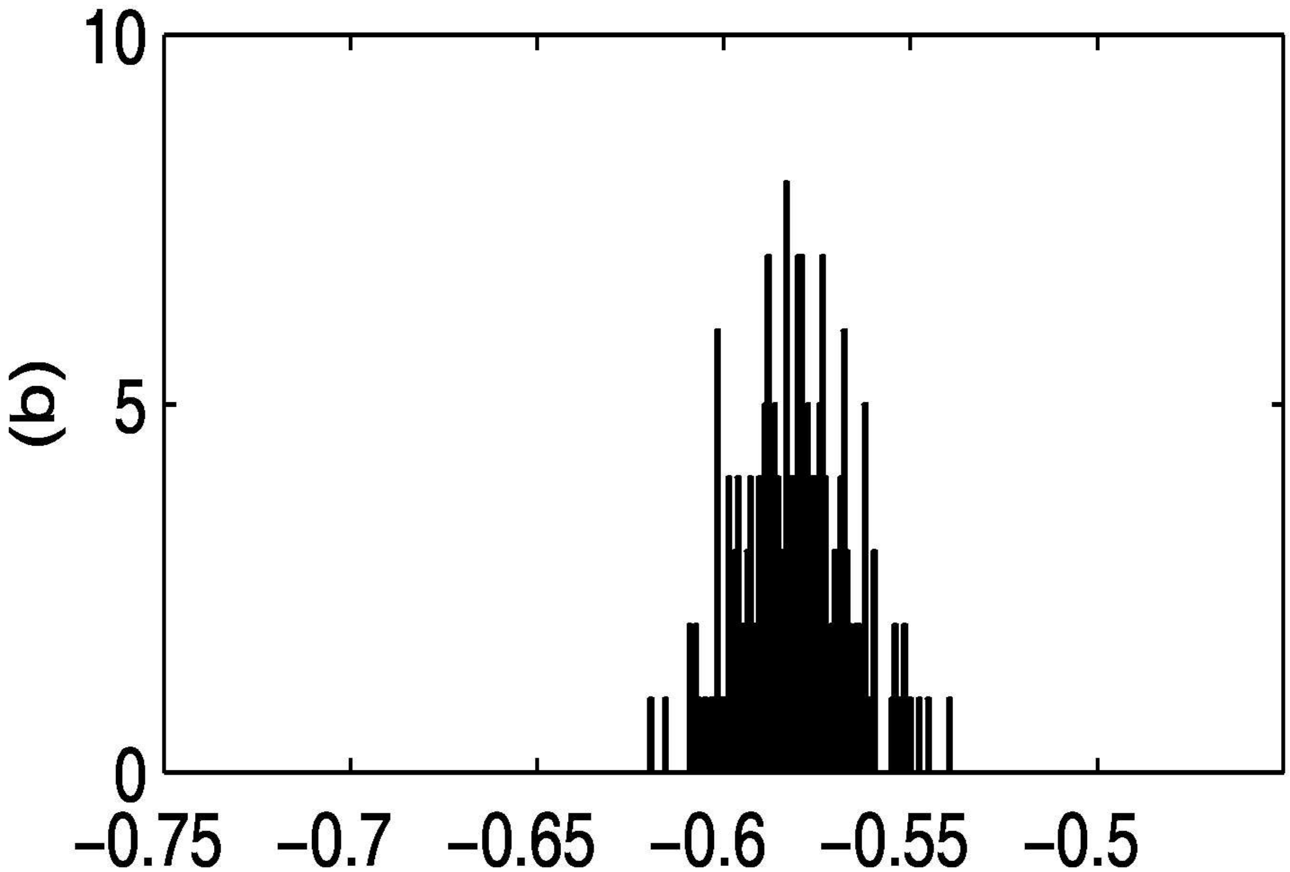

Figura 6: Polarização em modelo erro na saída - $\theta_2 = -0,7$

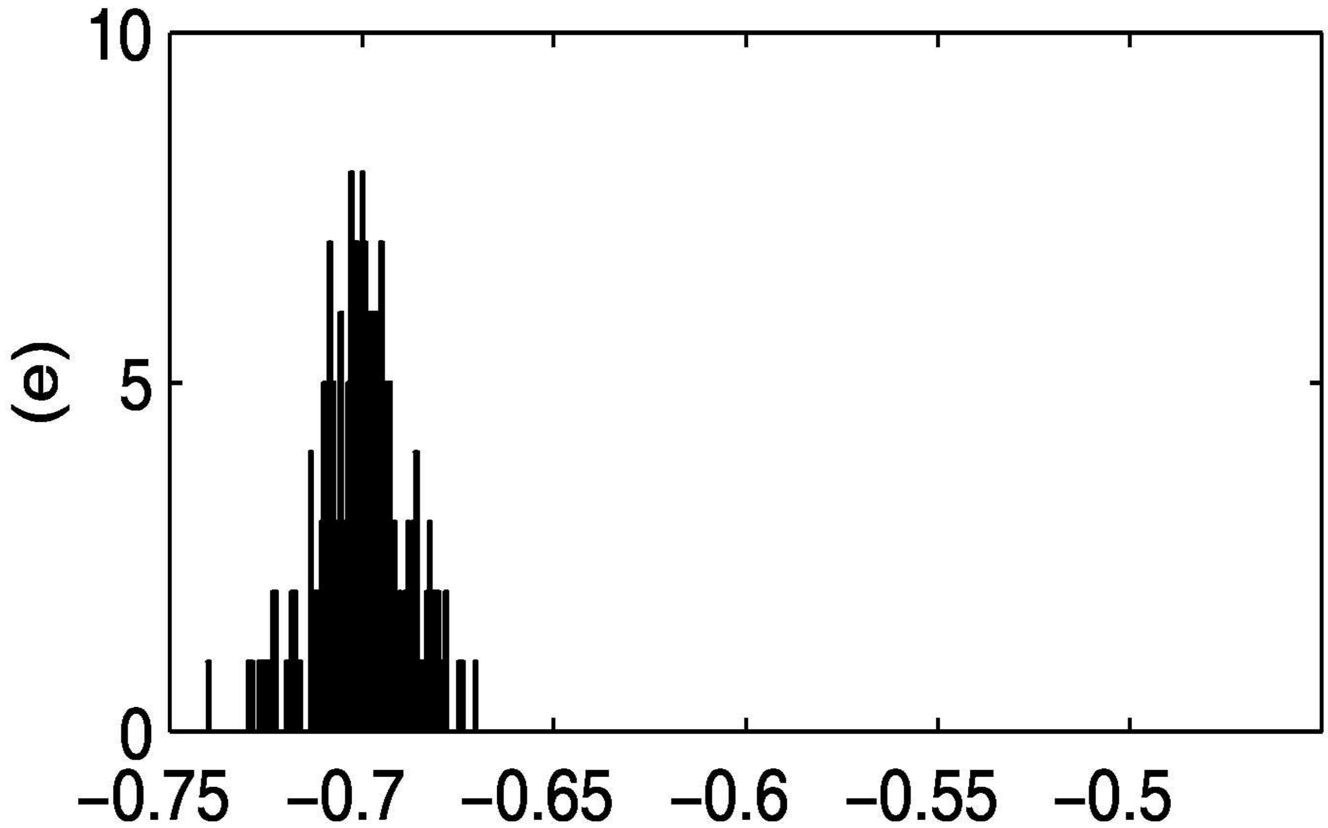

Figura 7: Polarização em modelo ARX - $\theta_2 = -0,7$

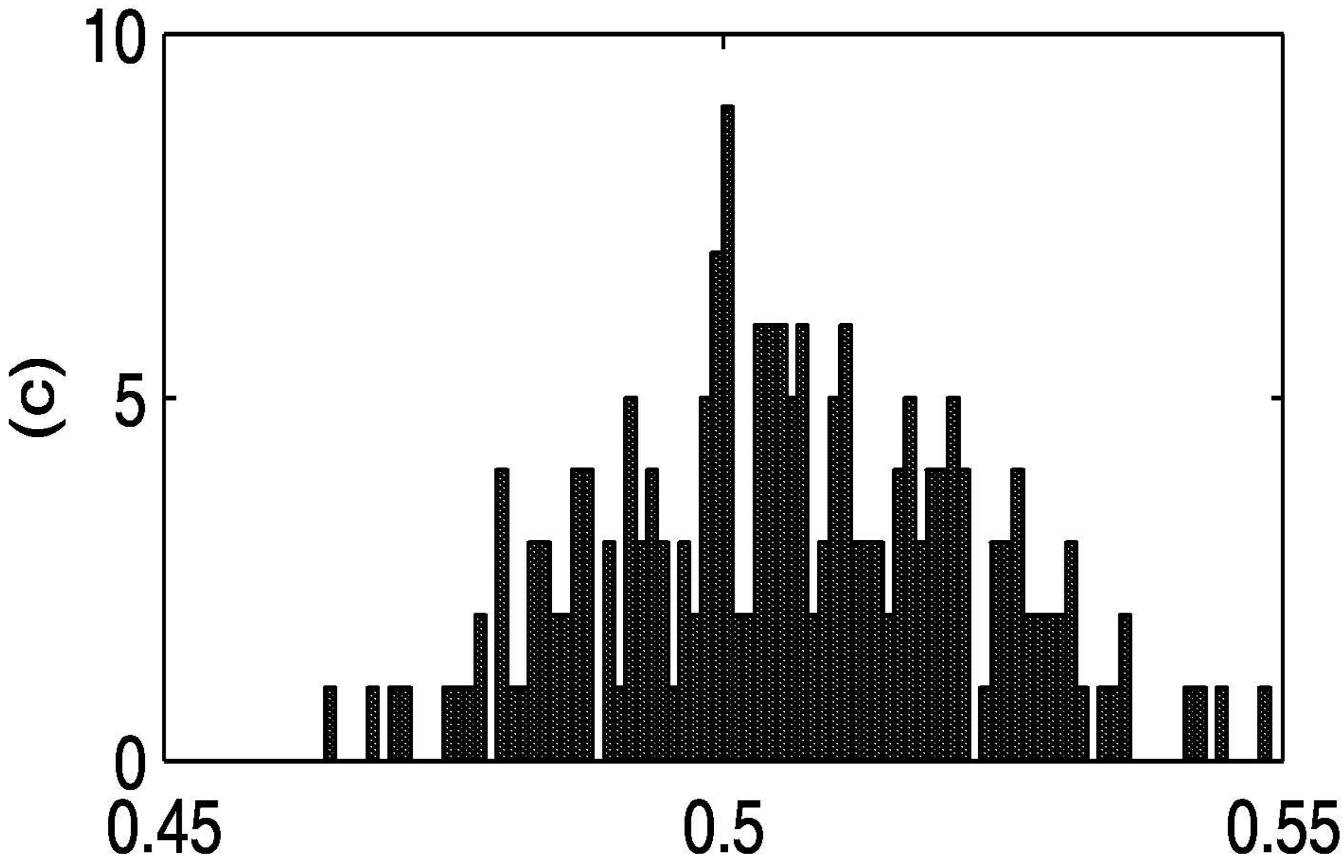

Figura 8: Polarização em modelo erro na saída - $\theta_3 = 0,5$

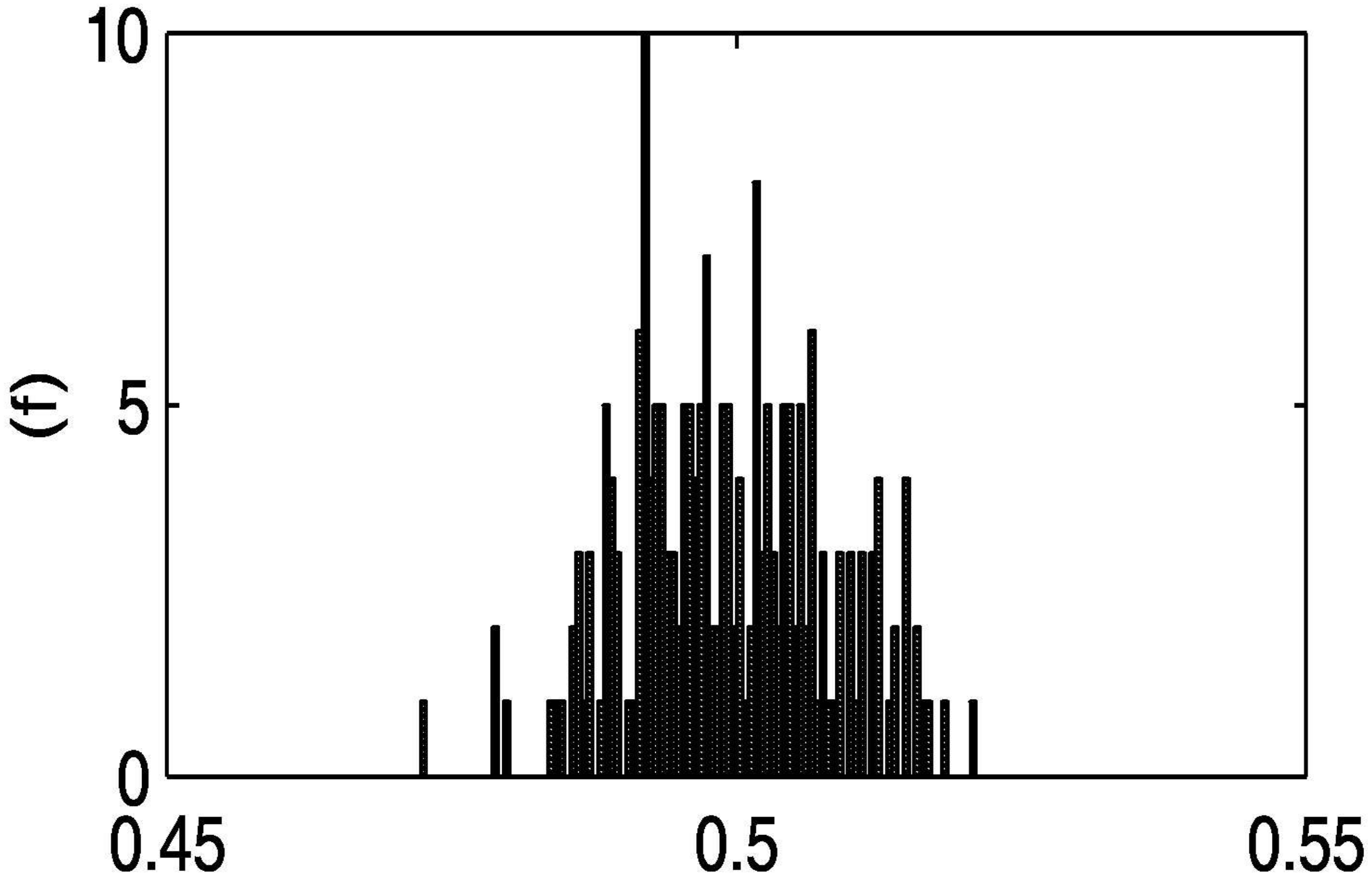

Figura 9: Polarização em modelo ARX - $\theta_3 = 0,5$

Polarização devida a ruído AR e regressores $y(k-i)$

Seja o modelo ARX

$$y(k) = a\,y(k-1) + b\,u(k-1) + e(k)$$

com
$$e(k) = c\,e(k-1) + \nu(k),$$

sendo que $r_{\nu\nu}(\tau) = 0,\ \forall \tau \neq 0.$

- Tem-se então

$$y(k) = a\,y(k-1) + b\,u(k-1) + c\,e(k-1) + \nu(k)$$

Polarização devida a ruído AR e regressores $y(k-i)$

Seja o modelo ARX

$$y(k) = a\,y(k-1) + b\,u(k-1) + e(k)$$

com
$$e(k) = c\,e(k-1) + \nu(k),$$

sendo que $r_{\nu\nu}(\tau) = 0,\ \forall \tau \neq 0.$

- Tem-se então

$$y(k) = a\,y(k-1) + b\,u(k-1) + c\,e(k-1) + \nu(k)$$

- e

$$y(k-1) = a\,y(k-2) + b\,u(k-2) + e(k-1).$$

- Fazendo-se substituições, obtém-se

$$
\begin{aligned}
y(k) &= \Big\{a[a\,y(k-2) + b\,u(k-2) + \underline{e(k-1)}] + b\,u(k-1)\Big\} \\
&\quad + c\,\underline{e(k-1)} + \nu(k).
\end{aligned}
$$

- Fazendo-se substituições, obtém-se

$$
\begin{aligned}
y(k) &= \Big\{a[a\,y(k-2)+b\,u(k-2)+\underline{e(k-1)}]+b\,u(k-1)\Big\} \\
&\quad +c\,\underline{e(k-1)}+\nu(k).
\end{aligned}
$$

- Se não houvesse regressores do tipo $y(k-i)$, ter-se-ia

$$
\begin{aligned}
y(k) &= b\,u(k-1)+c\,e(k-1)+\nu(k) \\
&= b\,u(k-1)+e(k).
\end{aligned}
$$

Nesse caso, o estimador MQ estimaria b sem polarização.

Polarização - Significado

A polarização pode ser vista como o resultado do estimador estar "forçando" $\Psi^{\mathrm{T}}\boldsymbol{\xi} = 0$ (propriedade de ortogonalidade entre regressores e resíduos), sendo que $\Psi^{\mathrm{T}}\mathbf{e} \neq 0$. Em outras palavras, a polarização surge em função do conflito que existe entre a propriedade do estimador de mínimos quadrados, $\Psi^{\mathrm{T}}\boldsymbol{\xi} = 0$, e a propriedade do erro de regressão, $\Psi^{\mathrm{T}}\mathbf{e} \neq 0$.

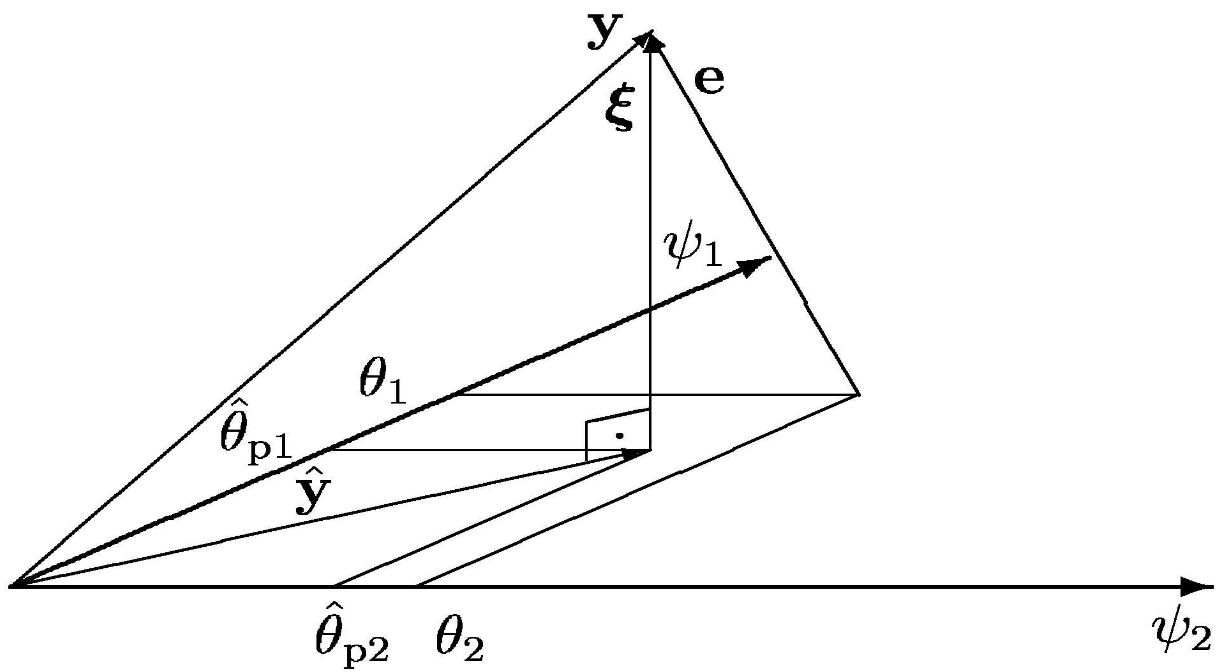

Figura 10: Polarização do estimador de MQ Polarização do estimador MQ quando o erro não é ortogonal aos regressores. Nesse caso os parâmetros corretos são $[\theta_1 \ \theta_2]^{\mathrm{T}}$, e o estimador MQ determina $[\hat{\theta}_{\mathrm{p}1} \ \hat{\theta}_{\mathrm{p}2}]^{\mathrm{T}}$.

O problema de “erros nas variáveis”

Considere

$$\begin{aligned} u(k) &= u^{\mathrm{i}}(k) + \nu_u(k), \\ y(k) &= y^{\mathrm{i}}(k) + e_y(k). \end{aligned}$$

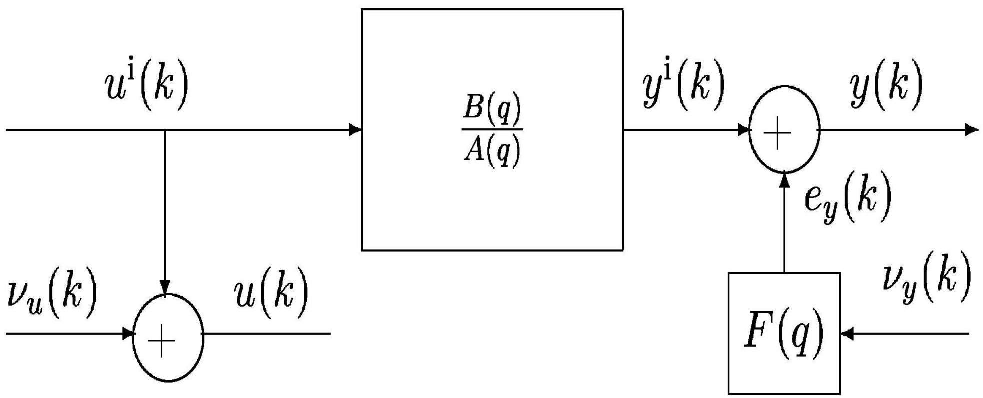

Figura 11: Ruído de observação na entrada e na saída. Tanto o sinal de entrada quanto o de saída estão contaminados por *ruído de observação*.

- Matematicamente, tem-se

$$
\begin{aligned}
A(q)y^{\mathrm{i}}(k) &= B(q)u^{\mathrm{i}}(k)\\
A(q)\left[y(k)-e_y(k)\right] &= B(q)\left[u(k)-\nu_u(k)\right]\\
A(q)y(k) &= B(q)u(k)+\left[A(q)F(q)\nu_y(k)-B(q)\nu_u(k)\right],
\end{aligned}
$$

sendo que $\nu_y(k)$ e $\nu_u(k)$ são ruídos brancos não correlacionados entre si.

- Matematicamente, tem-se

$$
\begin{aligned}
A(q)y^{\mathrm{i}}(k) &= B(q)u^{\mathrm{i}}(k) \\
A(q)\left[y(k)-e_y(k)\right] &= B(q)\left[u(k)-\nu_u(k)\right] \\
A(q)y(k) &= B(q)u(k)+\left[A(q)F(q)\nu_y(k)-B(q)\nu_u(k)\right],
\end{aligned}
$$

sendo que $\nu_y(k)$ e $\nu_u(k)$ são ruídos brancos não correlacionados entre si.

- Os parâmetros do modelo serão estimados usando-se

$$
A(q)y(k) = B(q)u(k) + e(k),
$$

O erro na equação de regressão não é branco, pois apesar de $\nu_y(k)$ e $\nu_u(k)$ serem brancos, assume-se que os filtros $A(q)F(q) \neq 1$ e $B(q) \neq 1$.

- Uma outra forma de entender a polarização neste caso é observar que tanto o erro de regressão

$$e(k) = A(q)F(q)\nu_y(k) - B(q)\nu_u(k)$$

quanto a própria entrada dependem de $\nu_u(k)$. Portanto $u(k)$ e $e(k)$ estão correlacionados, ou seja, $r_{ue}(k) \neq 0$.

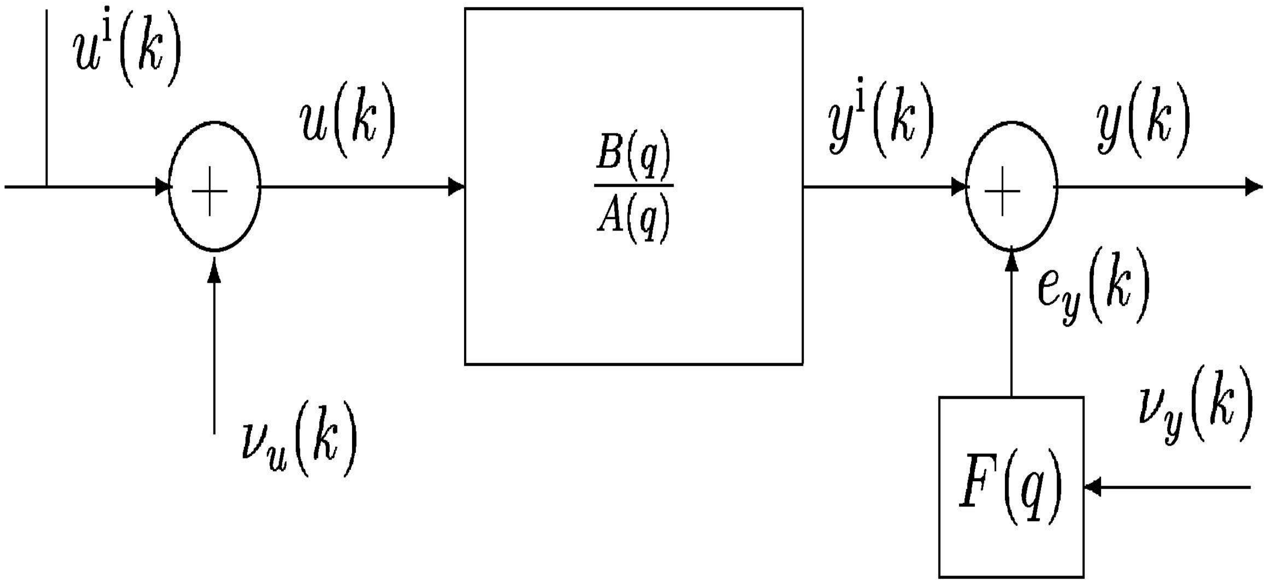

Figura 12: Entrada contaminada por ruído no atuador O sinal de entrada está contaminado por *ruído de atuador*. O algoritmo, entretanto, usa o sinal ideal $u^{\mathrm{i}}(k)$. O sinal de saída está contaminado por *ruído de medição*. Nesta situação, o estimador MQ não é polarizado.

- Matematicamente, tem-se

$$\begin{aligned} A(q)y^{\mathrm{i}}(k) &= B(q)u(k) \\ A(q)\left[y(k)-e_y(k)\right] &= B(q)\left[u^{\mathrm{i}}(k)+\nu_u(k)\right] \\ A(q)y(k) &= B(q)u^{\mathrm{i}}(k)+A(q)F(q)\nu_y(k)+B(q)\nu_u(k). \end{aligned}$$

- Matematicamente, tem-se

$$
\begin{aligned}
A(q)y^{\mathrm{i}}(k) &= B(q)u(k) \\
A(q)\left[y(k)-e_y(k)\right] &= B(q)\left[u^{\mathrm{i}}(k)+\nu_u(k)\right] \\
A(q)y(k) &= B(q)u^{\mathrm{i}}(k)+A(q)F(q)\nu_y(k)+B(q)\nu_u(k).
\end{aligned}
$$

- Portanto, o estimador MQ só será polarizado se $A(q)F(q)\nu_y(k)$ não for branco, uma vez que $u^{\mathrm{i}}(k)$ não depende de $\nu_u(k)$. Ou seja, o ruído de atuador aparece no erro da equação de regressão, mas não no sinal de entrada utilizado pelo estimador.

Covariância de Estimadores

- A *variância de uma variável escalar é*

$$\mathrm{var}[x] = \sigma_x^2 = \mathrm{E}\left[(x - \mathrm{E}[x])^2\right],$$

Covariância de Estimadores

- A *variância de uma variável escalar* é

$$\mathrm{var}[x] = \sigma_x^2 = \mathrm{E}\left[(x - \mathrm{E}[x])^2\right],$$

- Para um vetor $\mathbf{x} \in \mathbb{R}^n$, a *matriz de covariância* tem dimensão $n \times n$ e é

$$\begin{aligned} \mathrm{cov}[\mathbf{x}] &= \mathrm{E}\left[(\mathbf{x} - \mathrm{E}[\mathbf{x}])(\mathbf{x} - \mathrm{E}[\mathbf{x}])^{\mathrm{T}}\right] \\ &= \mathrm{E}[\mathbf{x}\mathbf{x}^{\mathrm{T}}] - \mathrm{E}[\mathbf{x}]\mathrm{E}[\mathbf{x}^{\mathrm{T}}]. \end{aligned}$$

  $\mathrm{cov}[\mathbf{x}]$ é simétrica e semidefinida positiva. Também,

$$\mathrm{cov}[C\mathbf{x}] = C\mathrm{cov}[\mathbf{x}]C^{\mathrm{T}}.$$

- O *valor quadrático médio* de um escalar x é

$$\mathrm{vqm}[x] = \mathrm{E}[x^2],$$

e a *matriz de valor quadrático médio* de um vetor **x** é

$$\mathrm{mms}[\mathbf{x}] = \mathrm{E}\left[\mathbf{x}\mathbf{x}^{\mathrm{T}}\right].$$

Variância de estimadores do tipo $\hat{\theta} = A\mathbf{y}$

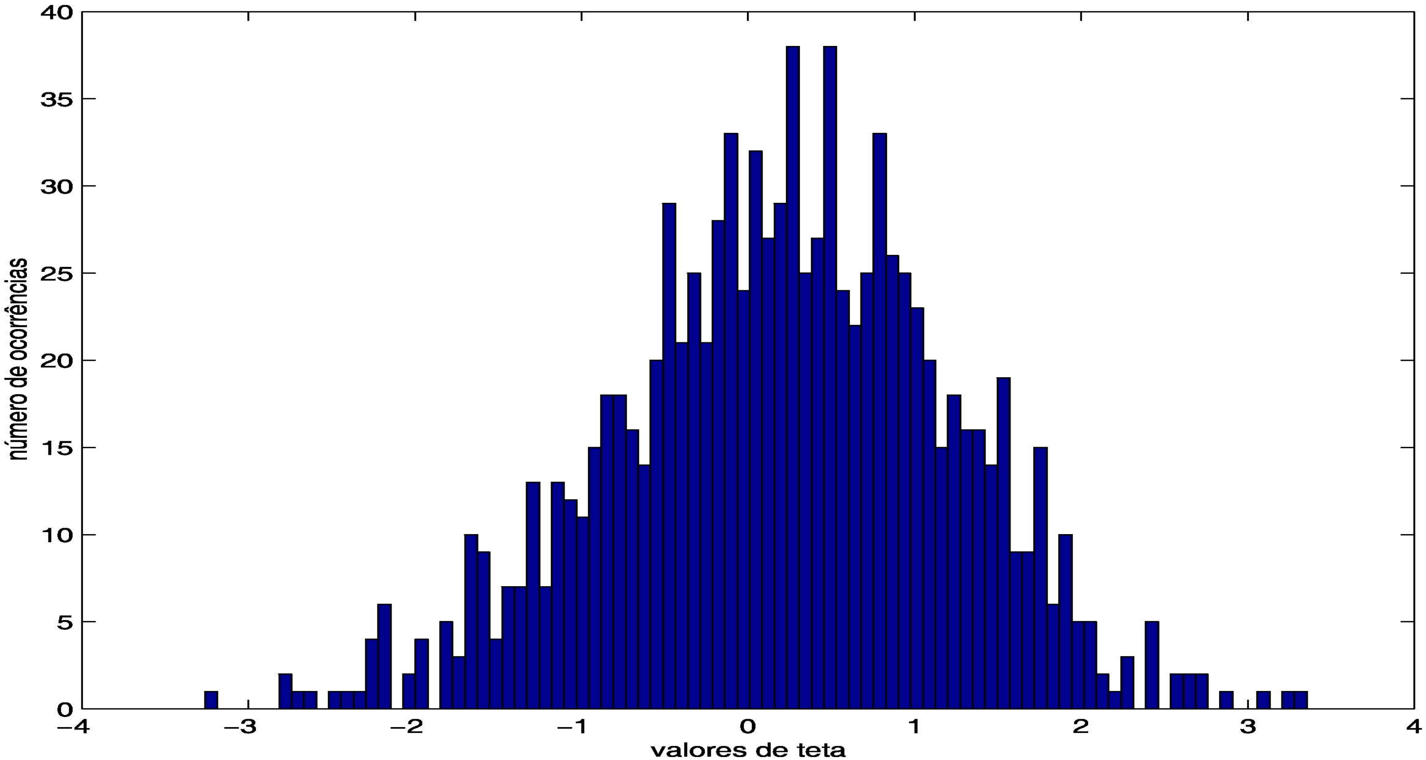

Figura 13: Histograma de um parâmetro estimado Histograma de valores estimados de um parâmetro com valor 0,2. Mesmo que o estimador utilizado não fosse polarizado (conforme a figura), percebe-se que valores relativamente afastados de 0,2 têm probabilidade alta de ocorrer.

- Dado o sistema $y(k) = \boldsymbol{\psi}^{\mathrm{T}}(k-1)\boldsymbol{\theta} + e(k)$, o vetor de parâmetros pode ser obtido por um estimador do tipo $\hat{\boldsymbol{\theta}} = A\mathbf{y}$. Se o estimador for não polarizado, a sua matriz de covariância será

$$
\begin{aligned}
\mathrm{cov}[\hat{\boldsymbol{\theta}}] &= \mathrm{E}\left[(\hat{\boldsymbol{\theta}} - \mathrm{E}[\hat{\boldsymbol{\theta}}])(\hat{\boldsymbol{\theta}} - \mathrm{E}[\hat{\boldsymbol{\theta}}])^{\mathrm{T}}\right] \\
&= \mathrm{E}\left[(A\mathbf{y} - \boldsymbol{\theta})(A\mathbf{y} - \boldsymbol{\theta})^{\mathrm{T}}\right] \\
&= \mathrm{E}\left[A\mathbf{y}\mathbf{y}^{\mathrm{T}}A^{\mathrm{T}} - A\mathbf{y}\boldsymbol{\theta}^{\mathrm{T}} - \boldsymbol{\theta}\mathbf{y}^{\mathrm{T}}A^{\mathrm{T}} + \boldsymbol{\theta}\boldsymbol{\theta}^{\mathrm{T}}\right] \\
&= \mathrm{E}\left[A(\Psi\boldsymbol{\theta} + \mathbf{e})(\Psi\boldsymbol{\theta} + \mathbf{e})^{\mathrm{T}}A^{\mathrm{T}} - A(\Psi\boldsymbol{\theta} + \mathbf{e})\boldsymbol{\theta}^{\mathrm{T}}\right. \\
&\quad \left. -\boldsymbol{\theta}(\Psi\boldsymbol{\theta} + \mathbf{e})^{\mathrm{T}}A^{\mathrm{T}} + \boldsymbol{\theta}\boldsymbol{\theta}^{\mathrm{T}}\right] \\
&= \mathrm{E}\left[(A\Psi\boldsymbol{\theta} + A\mathbf{e})(A\Psi\boldsymbol{\theta} + A\mathbf{e})^{\mathrm{T}} - (A\Psi\boldsymbol{\theta} + A\mathbf{e})\boldsymbol{\theta}^{\mathrm{T}}\right. \\
&\quad \left. -\boldsymbol{\theta}(A\Psi\boldsymbol{\theta} + A\mathbf{e})^{\mathrm{T}} + \boldsymbol{\theta}\boldsymbol{\theta}^{\mathrm{T}}\right].
\end{aligned}
$$

- Mas como $A\Psi = I$

$$\begin{aligned}\operatorname{cov}[\hat{\boldsymbol{\theta}}] &= \mathrm{E}\left[(\boldsymbol{\theta} + A\mathbf{e})(\boldsymbol{\theta} + A\mathbf{e})^{\mathrm{T}} - (\boldsymbol{\theta} + A\mathbf{e})\boldsymbol{\theta}^{\mathrm{T}}\right. \\ &\quad \left. -\boldsymbol{\theta}(\boldsymbol{\theta} + A\mathbf{e})^{\mathrm{T}} + \boldsymbol{\theta}\boldsymbol{\theta}^{\mathrm{T}}\right].\end{aligned}$$

- Mas como $A\Psi = I$

$$\begin{aligned}\mathrm{cov}[\hat{\boldsymbol{\theta}}] &= \mathrm{E}\left[(\boldsymbol{\theta} + A\mathbf{e})(\boldsymbol{\theta} + A\mathbf{e})^{\mathrm{T}} - (\boldsymbol{\theta} + A\mathbf{e})\boldsymbol{\theta}^{\mathrm{T}}\right. \\ &\quad \left. -\boldsymbol{\theta}(\boldsymbol{\theta} + A\mathbf{e})^{\mathrm{T}} + \boldsymbol{\theta}\boldsymbol{\theta}^{\mathrm{T}}\right].\end{aligned}$$

- Finalmente chega-se a

$$\mathrm{cov}[\hat{\boldsymbol{\theta}}] = \mathrm{E}\left[A\mathbf{e}\mathbf{e}^{\mathrm{T}}A^{\mathrm{T}}\right].$$

Variância de estimadores do tipo $\hat{\boldsymbol{\theta}} = A\mathbf{y}$

- Qual é a variância de um estimador do tipo $\hat{\boldsymbol{\theta}} = A\mathbf{y}$ no caso em que o erro de regressão é ruído branco?

Variância de estimadores do tipo $\hat{\boldsymbol{\theta}} = A\mathbf{y}$

- Qual é a variância de um estimador do tipo $\hat{\boldsymbol{\theta}} = A\mathbf{y}$ no caso em que o erro de regressão é ruído branco?
- Se **e** é ruído branco, ele pode ser expresso da seguinte forma: $\mathrm{cov}[\mathbf{e}] = \mathrm{E}[\mathbf{e}\mathbf{e}^{\mathrm{T}}] = \sigma_e^2 I$. Portanto, a variância é

$$\mathrm{cov}[\hat{\boldsymbol{\theta}}] = \mathrm{E}\left[AA^{\mathrm{T}}\right]\sigma_e^2,$$

sendo que σ_e^2 é a variância do ruído.

Variância de estimadores do tipo $\hat{\boldsymbol{\theta}} = A\mathbf{y}$

- Qual é a variância de um estimador do tipo $\hat{\boldsymbol{\theta}} = A\mathbf{y}$ no caso em que o erro de regressão é ruído branco?
- Se **e** é ruído branco, ele pode ser expresso da seguinte forma: $\mathrm{cov}[\mathbf{e}] = \mathrm{E}[\mathbf{e}\mathbf{e}^{\mathrm{T}}] = \sigma_e^2 I$. Portanto, a variância é

$$\mathrm{cov}[\hat{\boldsymbol{\theta}}] = \mathrm{E}\left[AA^{\mathrm{T}}\right]\sigma_e^2,$$

sendo que σ_e^2 é a variância do ruído.

- Para o estimador de mínimos quadrados, sabe-se que $A = [\Psi^{\mathrm{T}}\Psi]^{-1}\Psi^{\mathrm{T}}$ e, portanto, a última equação torna-se

$$\mathrm{cov}[\hat{\boldsymbol{\theta}}_{\mathrm{MQ}}] = \mathrm{E}\left[(\Psi^{\mathrm{T}}\Psi)^{-1}\right]\sigma_e^2.$$

Uma observação importante a ser feita é que, quando o vetor de erros **e** é branco, a variância do estimador de mínimos quadrados é a menor de todos os estimadores do tipo $\hat{\boldsymbol{\theta}} = A\mathbf{y}$ que sejam *não polarizados*. Ou seja, $\mathrm{cov}[\hat{\boldsymbol{\theta}}_{\mathrm{MQ}}] \leq \mathrm{cov}[\hat{\boldsymbol{\theta}}]$.

Eficiência de Estimadores

Seria interessante comparar a variância do estimador a um valor de variância "ótimo". O padrão comumente usado para estimadores não polarizados é a *norma de Cramér-Rao*. A *eficiência de um estimador* é definida como

$$\mathrm{E_f} = \frac{F^{-1}}{\mathrm{cov}[\hat{\boldsymbol{\theta}}]},$$

sendo $\mathrm{E_f}$ a eficiência do estimador de $\boldsymbol{\theta}$ e F^{-1}, a norma de Cramér-Rao.

- Para efeito de discussão, um estimador qualquer será representado por $\hat{\boldsymbol{\theta}} = E(\mathbf{y}, \mathbf{u})$. Note que o estimador $E(\cdot)$ não deve ser confundido com o operador de esperança matemática.

- Para efeito de discussão, um estimador qualquer será representado por $\hat{\boldsymbol{\theta}} = E(\mathbf{y}, \mathbf{u})$. Note que o estimador $E(\cdot)$ não deve ser confundido com o operador de esperança matemática.
- Portanto, estimadores lineares podem ser representados da seguinte forma conhecida $\hat{\boldsymbol{\theta}} = A\mathbf{y}$, sendo que A normalmente depende dos vetores de dados de entrada e saída **u** e **y**, de dimensão N.

- Para efeito de discussão, um estimador qualquer será representado por $\hat{\boldsymbol{\theta}} = E(\mathbf{y}, \mathbf{u})$. Note que o estimador $E(\cdot)$ não deve ser confundido com o operador de esperança matemática.
- Portanto, estimadores lineares podem ser representados da seguinte forma conhecida $\hat{\boldsymbol{\theta}} = A\mathbf{y}$, sendo que A normalmente depende dos vetores de dados de entrada e saída **u** e **y**, de dimensão N.
- Assim, um estimador não polarizado $\hat{\boldsymbol{\theta}} = E(\mathbf{y}, \mathbf{u})$ é *mais eficiente* do que um outro estimador $\hat{\hat{\boldsymbol{\theta}}} = E(\mathbf{y}, \mathbf{u})$ se $\mathrm{cov}[\hat{\boldsymbol{\theta}}] \leq \mathrm{cov}[\hat{\hat{\boldsymbol{\theta}}}]$, para todo valor de N.

- Para efeito de discussão, um estimador qualquer será representado por $\hat{\boldsymbol{\theta}} = E(\mathbf{y}, \mathbf{u})$. Note que o estimador $E(\cdot)$ não deve ser confundido com o operador de esperança matemática.
- Portanto, estimadores lineares podem ser representados da seguinte forma conhecida $\hat{\boldsymbol{\theta}} = A\mathbf{y}$, sendo que A normalmente depende dos vetores de dados de entrada e saída **u** e **y**, de dimensão N.
- Assim, um estimador não polarizado $\hat{\boldsymbol{\theta}} = E(\mathbf{y}, \mathbf{u})$ é *mais eficiente* do que um outro estimador $\hat{\hat{\boldsymbol{\theta}}} = E(\mathbf{y}, \mathbf{u})$ se $\mathrm{cov}[\hat{\boldsymbol{\theta}}] \leq \mathrm{cov}[\hat{\hat{\boldsymbol{\theta}}}]$, para todo valor de N.
- Além disso, diz-se que um estimador é *eficiente* se sua variância atinge a norma de Cramér-Rao, que é o mínimo atingível por todos os estimadores (lineares e não-lineares) não polarizados.

A Norma de Cramér-Rao

- A *norma de Cramér-Rao* de um estimador de $\boldsymbol{\theta}$ é F^{-1}, onde F, a *matriz de informação de Fisher*, é definida como

$$
\begin{aligned}
F &= \mathrm{E}_{y|\boldsymbol{\theta}}\left[\left(\frac{\partial}{\partial\boldsymbol{\theta}}\ln p(y \mid \boldsymbol{\theta})\right)\left(\frac{\partial}{\partial\boldsymbol{\theta}}\ln p(y \mid \boldsymbol{\theta})\right)^{\mathrm{T}}\right] \\
&= -\left(\mathrm{E}_{y|\boldsymbol{\theta}}\left[\frac{\partial^2 \ln p(y \mid \boldsymbol{\theta})}{\partial\boldsymbol{\theta}^2}\right]\right),
\end{aligned}
$$

sendo $p(y \mid \boldsymbol{\theta})$ a função de densidade de probabilidade das medições y dado o vetor de parâmetros $\boldsymbol{\theta}$. Nota-se que tal matriz quantifica como variações de $\boldsymbol{\theta}$ afetam $p(y \mid \boldsymbol{\theta})$.

A Norma de Cramér-Rao

- A *norma de Cramér-Rao* de um estimador de $\boldsymbol{\theta}$ é F^{-1}, onde F, a *matriz de informação de Fisher*, é definida como

$$
\begin{aligned}
F &= \mathrm{E}_{y|\boldsymbol{\theta}}\left[\left(\frac{\partial}{\partial\boldsymbol{\theta}}\ln p(y \mid \boldsymbol{\theta})\right)\left(\frac{\partial}{\partial\boldsymbol{\theta}}\ln p(y \mid \boldsymbol{\theta})\right)^{\mathrm{T}}\right] \\
&= -\left(\mathrm{E}_{y|\boldsymbol{\theta}}\left[\frac{\partial^2 \ln p(y \mid \boldsymbol{\theta})}{\partial\boldsymbol{\theta}^2}\right]\right),
\end{aligned}
$$

  sendo $p(y \mid \boldsymbol{\theta})$ a função de densidade de probabilidade das medições y dado o vetor de parâmetros $\boldsymbol{\theta}$. Nota-se que tal matriz quantifica como variações de $\boldsymbol{\theta}$ afetam $p(y \mid \boldsymbol{\theta})$.

- É possível provar que para estimadores não polarizados

$$
\mathrm{cov}[\hat{\boldsymbol{\theta}}(y)] \geq F^{-1},
$$

  sendo que o argumento no vetor estimado simplesmente evidencia que $\hat{\boldsymbol{\theta}}(y)$ é estimado a partir de medições $y(k)$.

Eficiência de um estimador não polarizado

Deseja-se estimar o parâmetro α da função de densidade de probabilidade $p(y \mid \alpha) = (1/\alpha)\exp[-y/\alpha],\ y \geq 0$, a partir de N medições independentes de y.

- Primeiramente, será determinada a norma mínima de Cramér-Rao de um estimador não polarizado qualquer do parâmetro α.

Eficiência de um estimador não polarizado

Deseja-se estimar o parâmetro α da função de densidade de probabilidade $p(y \mid \alpha) = (1/\alpha)\exp[-y/\alpha],\ y \geq 0$, a partir de N medições independentes de y.

- Primeiramente, será determinada a norma mínima de Cramér-Rao de um estimador não polarizado qualquer do parâmetro α.
- Em segundo lugar, será determinada a variância do estimador em questão, $\hat{\alpha} = \sum_{k=1}^{N} y(k)/N$, e, finalmente, será verificado se ela atinge ou não a norma de Cramér-Rao.

Eficiência de um estimador não polarizado

Deseja-se estimar o parâmetro α da função de densidade de probabilidade $p(y \mid \alpha) = (1/\alpha)\exp[-y/\alpha],\ y \geq 0$, a partir de N medições independentes de y.

- Primeiramente, será determinada a norma mínima de Cramér-Rao de um estimador não polarizado qualquer do parâmetro α.
- Em segundo lugar, será determinada a variância do estimador em questão, $\hat{\alpha} = \sum_{k=1}^{N} y(k)/N$, e, finalmente, será verificado se ela atinge ou não a norma de Cramér-Rao.
- A função densidade de probabilidade conjunta das medições $y(k)$ para $k = 1, \ldots, N$ é

$$p(y(1), \ldots, y(N) \mid \alpha) \quad = \quad \prod_{k=1}^{N} p(y(k) \mid \alpha) = \alpha^{-N} \exp\left[-\sum_{k=1}^{N} \frac{y(k)}{\alpha}\right].$$

- As seguintes entidades podem agora ser determinadas:

$$
\begin{aligned}
\ln p(y \mid \alpha) &= -N \ln \alpha - \sum_{k=1}^{N} y(k)/\alpha \\
\frac{\partial}{\partial \alpha} \ln p(y \mid \alpha) &= -N/\alpha + \sum_{k=1}^{N} y(k)/\alpha^2 \\
\frac{\partial^2}{\partial \alpha^2} \ln p(y \mid \alpha) &= N/\alpha^2 - 2\sum_{k=1}^{N} y(k)/\alpha^3 .
\end{aligned}
$$

- As seguintes entidades podem agora ser determinadas:

$$
\begin{aligned}
\ln p(y \mid \alpha) &= -N \ln \alpha - \sum_{k=1}^{N} y(k)/\alpha \\
\frac{\partial}{\partial \alpha} \ln p(y \mid \alpha) &= -N/\alpha + \sum_{k=1}^{N} y(k)/\alpha^2 \\
\frac{\partial^2}{\partial \alpha^2} \ln p(y \mid \alpha) &= N/\alpha^2 - 2 \sum_{k=1}^{N} y(k)/\alpha^3 .
\end{aligned}
$$

- Substituindo-se, tem-se

$$
\begin{aligned}
F &= -\mathrm{E}\left[N/\alpha^2 - 2 \sum_{k=1}^{N} y(k)/\alpha^3 \right] \\
&= -N/\alpha^2 + 2 \frac{\mathrm{E}[\sum_{k=1}^{N} y(k)]}{\alpha^3},
\end{aligned}
$$

mas como $\mathrm{E}[y(k)] = \alpha$, então $\mathrm{E}[\sum_{k=1}^{N} y(k)] = N\alpha$.

- Então

$$
\begin{aligned}
F &= -N/\alpha^2 + 2\frac{\mathrm{E}[N\alpha]}{\alpha^3} \\
&= N/\alpha^2.
\end{aligned}
$$

- Então

$$
\begin{aligned}
F &= -N/\alpha^2 + 2\frac{\mathrm{E}[N\alpha]}{\alpha^3} \\
&= N/\alpha^2.
\end{aligned}
$$

- Logo, $F^{-1} = \alpha^2/N$ é a norma mínima de Cramér-Rao para este exemplo. Portanto, $\mathrm{var}[\hat{\alpha}] \geq \alpha^2/N$. Resta agora determinar a variância do estimador:

$$
\begin{aligned}
\mathrm{var}[\hat{\alpha}] &= \mathrm{E}[(\alpha - \hat{\alpha})^2] \\
&= \mathrm{E}\left[\frac{(\sum y(k))^2}{N^2} - 2\alpha\frac{\sum y(k)}{N} + \alpha^2\right] \\
&= \alpha^2/N.
\end{aligned}
$$

Portanto, o estimador deste exemplo é eficiente.

Polarização do estimador de MQ usando correlação — Um exemplo

Seja um sistema representado pelo seguinte modelo ARMAX:

$$y(k) + a_1 y(k-1) = b_1 u(k-1) + c_1 \nu(k-1) + \nu(k), \tag{-47}$$

sendo que $u(k)$ e $\nu(k)$ são seqüências aleatórias de média nula e variâncias σ_u^2 e σ_ν^2, respectivamente. Os valores das funções de autocorrelação e correlação cruzada da saída e entrada para os atrasos 0 e 1 podem ser determinados analiticamente, para $N \to \infty$, como

$$\begin{aligned}
r_{yy}(0) &= \frac{(1 + c_1^2 - 2a_1c_1)\sigma_\nu^2 + b_1^2\sigma_u^2}{1 - a_1^2} \qquad (-46) \\
r_{yy}(-1) &= \frac{(c_1 - a_1 - a_1c_1^2 - a_1^2c_1)\sigma_\nu^2 + a_1b_1^2\sigma_u^2}{1 - a_1^2} \\
r_{yu}(0) &= 0 \\
r_{yu}(-1) &= b_1\sigma_u^2.
\end{aligned}$$

Usando-se a seguinte equação de regressão:

$$y(k) + a_1 y(k-1) = b_1 u(k-1) + e(k), \qquad (-45)$$

é facil mostrar que a equação normal pode ser expressa em termos das funções de autocorrelação e correlação cruzada da saída e entrada da seguinte forma (note que $r_{uy}(1) = r_{yu}(-1)$):

$$\begin{bmatrix} r_{yu}(-1) \\ -r_{yy}(-1) \end{bmatrix} = \begin{bmatrix} -r_{yu}(0) & r_{uu}(0) \\ r_{yy}(0) & -r_{yu}(0) \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{b}_1 \end{bmatrix}, \qquad (-44)$$

sendo que o que premultiplica o vetor de parâmetros é a matriz de covariância R_{yu}.

A solução de equação anterior é é (usando-se os valores analiticamente calculados para as funções de correlação)

$$\begin{aligned} \hat{a}_1 &= a_1 - \frac{c_1 \sigma_\nu^2 (1 - a_1^2)}{(1 + c_1^2 - 2a_1 c_1)\sigma_\nu^2 + b_1^2 \sigma_u^2} \\ \hat{b}_1 &= b_1. \end{aligned}$$

O resultado acima tem diversos aspectos interessantes, dentre os quais ressaltam-se os seguintes:

- $\hat{a}_1$ estará normalmente polarizado;

O resultado acima tem diversos aspectos interessantes, dentre os quais ressaltam-se os seguintes:

- $\hat{a}_1$ estará normalmente polarizado;
- $\hat{b}_1$ nunca estará polarizado. Portanto, é possível que num mesmo vetor de parâmetros alguns revelem polarização e outros não;

O resultado acima tem diversos aspectos interessantes, dentre os quais ressaltam-se os seguintes:

- $\hat{a}_1$ estará normalmente polarizado;
- $\hat{b}_1$ nunca estará polarizado. Portanto, é possível que num mesmo vetor de parâmetros alguns revelem polarização e outros não;
- para evitar polarização na estimativa de a_1, usando-se o modelo de regressão (-45) e o estimador MQ, é necessário eliminar o erro em (-47) de forma a garantir $\sigma_\nu^2 = 0$ ou $c_1 = 0$. A última condição é equivalente a garantir que o erro na equação de regressão seja branco.

Variância mínima do estimador MQ

Neste complemento será mostrado que o estimador de mínimos quadrados tem a menor variância dentre todos os estimadores lineares não polarizados quando o vetor de erro de regressão é ruído branco. Como indicado anteriormente, isso corresponde a provar que $\sigma_e^2 \left(AA^{\mathrm{T}} - [\Psi^{\mathrm{T}}\Psi]^{-1}\right)$ é semi-definida positiva.

A fim de provar esta propriedade, deve ser notado que a matriz esperada do produto de uma matriz real pela sua transposta é semidefinida positiva, ou seja $\mathrm{E}[XX^{\mathrm{T}}]$ é semidefinida positiva se X for uma matriz real (ver propriedade **??**). Se $X = A - [\Psi^{\mathrm{T}}\Psi]^{-1}\Psi^{\mathrm{T}}$, tem-se

$$\begin{aligned} XX^{\mathrm{T}} &= \left(A - [\Psi^{\mathrm{T}}\Psi]^{-1}\Psi^{\mathrm{T}}\right)\left(A - [\Psi^{\mathrm{T}}\Psi]^{-1}\Psi^{\mathrm{T}}\right)^{\mathrm{T}} \\ &= AA^{\mathrm{T}} - A\Psi[\Psi^{\mathrm{T}}\Psi]^{-1} - [\Psi^{\mathrm{T}}\Psi]^{-1}\Psi^{\mathrm{T}}A^{\mathrm{T}} + [\Psi^{\mathrm{T}}\Psi]^{-1} \end{aligned} \tag{-.43}$$

- Escolhendo $A = [C\Psi]^{-1}C$ (note que esta escolha satisfaz a condição $A\Psi = I$), tem-se

$$
\begin{aligned}
A\Psi[\Psi^{\mathrm{T}}\Psi]^{-1} &= [C\Psi]^{-1}C\Psi[\Psi^{\mathrm{T}}\Psi]^{-1} \\
&= \Psi^{-1}C^{-1}C\Psi[\Psi^{\mathrm{T}}\Psi]^{-1} \\
&= [\Psi^{\mathrm{T}}\Psi]^{-1}, \qquad \text{(-42)}
\end{aligned}
$$

sendo que a propriedade $[XY]^{-1} = Y^{-1}X^{-1}$ foi utilizada.

- Escolhendo $A = [C\Psi]^{-1}C$ (note que esta escolha satisfaz a condição $A\Psi = I$), tem-se

$$\begin{aligned} A\Psi[\Psi^{\mathrm{T}}\Psi]^{-1} &= [C\Psi]^{-1}C\Psi[\Psi^{\mathrm{T}}\Psi]^{-1} \\ &= \Psi^{-1}C^{-1}C\Psi[\Psi^{\mathrm{T}}\Psi]^{-1} \\ &= [\Psi^{\mathrm{T}}\Psi]^{-1}, \end{aligned} \tag{-40}$$

  sendo que a propriedade $[XY]^{-1} = Y^{-1}X^{-1}$ foi utilizada.
- Analogamente, é possível verificar que

$$[\Psi^{\mathrm{T}}\Psi]^{-1}\Psi^{\mathrm{T}}A^{\mathrm{T}} = [\Psi^{\mathrm{T}}\Psi]^{-1}. \tag{-41}$$

- Escolhendo $A = [C\Psi]^{-1}C$ (note que esta escolha satisfaz a condição $A\Psi = I$), tem-se

$$\begin{aligned} A\Psi[\Psi^{\mathrm{T}}\Psi]^{-1} &= [C\Psi]^{-1}C\Psi[\Psi^{\mathrm{T}}\Psi]^{-1} \\ &= \Psi^{-1}C^{-1}C\Psi[\Psi^{\mathrm{T}}\Psi]^{-1} \\ &= [\Psi^{\mathrm{T}}\Psi]^{-1}, \end{aligned} \quad (-42)$$

sendo que a propriedade $[XY]^{-1} = Y^{-1}X^{-1}$ foi utilizada.

- Analogamente, é possível verificar que

$$[\Psi^{\mathrm{T}}\Psi]^{-1}\Psi^{\mathrm{T}}A^{\mathrm{T}} = [\Psi^{\mathrm{T}}\Psi]^{-1}. \quad (-43)$$

- Portanto, substituindo-se (-42) e (-41) em (-43), tem-se

$$XX^{\mathrm{T}} = AA^{\mathrm{T}} - [\Psi^{\mathrm{T}}\Psi]^{-1}. \quad (-44)$$

- Finalmente, aplicando-se a esperança matemática à última equação, obtém-se

$$\begin{aligned} \mathrm{E}[XX^{\mathrm{T}}] &= \mathrm{E}[AA^{\mathrm{T}}] - \mathrm{E}[(\Psi^{\mathrm{T}}\Psi)^{-1}] \\ &= \frac{\mathrm{cov}[\hat{\boldsymbol{\theta}}]}{\sigma_e^2} - \frac{\mathrm{cov}[\hat{\boldsymbol{\theta}}_{\mathrm{MQ}}]}{\sigma_e^2}, \end{aligned} \tag{-43}$$

sendo que a matriz do lado esquerdo da igualdade é semidefinida positiva por construção. Portanto, a variância do estimador de mínimos quadrados é a menor dentre todos os estimadores lineares não polarizados no caso do vetor de erro ser ruído branco.

Eficiência do estimador MQ

Considere a saída $y(k)$ de um processo parametrizado da seguinte forma $y(k) = \boldsymbol{\psi}^{\mathrm{T}}\boldsymbol{\theta} + e(k)$, sendo que o erro $e(k)$ é normalmente distribuído com média zero e desvio padrão igual a σ_e. Além disso, a função de densidade de probabilidade das medições é dada por

$$p(y \mid \boldsymbol{\theta}) = \frac{\exp\left[-0,5(\mathbf{y} - \Psi\boldsymbol{\theta})^{\mathrm{T}}(\sigma_e^2 I_N)^{-1}(\mathbf{y} - \Psi\boldsymbol{\theta})\right]}{(2\pi)^{N/2}\det[\sigma_e^2 I_N]^{0,5}}, \qquad \text{(-42)}$$

sendo que N é o número de observações e I_N é a matriz identidade de dimensão N. Agora será determinada a norma de Cramér-Rao de um estimador não polarizado de $\boldsymbol{\theta}$.

Tem-se

$$\begin{aligned}\ln p(y \mid \boldsymbol{\theta}) &= \ln(2\pi)^{-N/2} + \ln(\sigma_e^2)^{-N/2} - \frac{(\mathbf{y}-\Psi\boldsymbol{\theta})^{\mathrm{T}}(\mathbf{y}-\Psi\boldsymbol{\theta})}{2\sigma_e^2} \\ &= -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma_e^2) - \frac{(\mathbf{y}-\Psi\boldsymbol{\theta})^{\mathrm{T}}(\mathbf{y}-\Psi\boldsymbol{\theta})}{2\sigma_e^2} \qquad (\text{-}41)\end{aligned}$$

sendo que foi usado $\det[kA] = k^n \det[A]$, $n = \dim[A]$. Além disso,

$$\begin{aligned}\frac{\partial \ln p(y \mid \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= -\frac{\partial}{2\sigma_e^2\,\partial\boldsymbol{\theta}}\left[\mathbf{y}^{\mathrm{T}}\mathbf{y} - \boldsymbol{\theta}^{\mathrm{T}}\Psi^{\mathrm{T}}\mathbf{y} - \mathbf{y}^{\mathrm{T}}\Psi\boldsymbol{\theta} + \boldsymbol{\theta}^{\mathrm{T}}\Psi^{\mathrm{T}}\Psi\boldsymbol{\theta}\right] \\ &= -\frac{1}{2\sigma_e^2}\left[-\Psi^{\mathrm{T}}\mathbf{y} - \Psi^{\mathrm{T}}\mathbf{y} + \left(\Psi^{\mathrm{T}}\Psi + \Psi^{\mathrm{T}}\Psi\right)\boldsymbol{\theta}\right] \\ &= -\frac{1}{2\sigma_e^2}\left[-2\Psi^{\mathrm{T}}(\mathbf{y}-\Psi\boldsymbol{\theta})\right]. \qquad (\text{-}40)\end{aligned}$$

Finalmente,

$$\frac{\partial^2 \ln p(y \mid \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} = \frac{1}{\sigma_e^2}[\Psi^{\mathrm{T}}\Psi].$$

Logo, a desigualdade de Cramér-Rao se torna

$$\mathrm{cov}[\hat{\boldsymbol{\theta}}] \geq -\left\{\mathrm{E}\left[\frac{\Psi^{\mathrm{T}}\Psi}{\sigma_e^2}\right]\right\}^{-1} = \mathrm{E}\left[(\Psi^{\mathrm{T}}\Psi)^{-1}\right]\sigma_e^2, \tag{-39}$$

que é precisamente a covariância do estimador de mínimos quadrados (veja a equação (-35)). Logo, o estimador MQ atinge a norma mínima de Cramér-Rao e portanto é eficiente. Este resultado mostra que o estimador MQ apresenta a menor covariância dentre todos os estimadores (lineares e não-lineares) não polarizados, *quando o erro for gaussiano e a função densidade de probabilidade for como mostrada na equação (-42)*.

Convergência e consistência do estimador MQ

Suponha que os dados observados tenham sido gerados pelo sistema

$$y(k) = \boldsymbol{\psi}^{\mathrm{T}}(k-1)\boldsymbol{\theta} + e(k), \qquad \text{(-38)}$$

sendo $e(k)$ uma seqüência de erros qualquer. Usando-se (-38) como equação de regressão, resulta na equação normal $\Psi^{\mathrm{T}}\mathbf{y} = \Psi^{\mathrm{T}}\Psi\boldsymbol{\theta}$. Portanto, o estimador MQ pode ser expresso como

$$\hat{\boldsymbol{\theta}}_{\mathrm{MQ}} = [\Psi^{\mathrm{T}}\Psi]^{-1}\Psi^{\mathrm{T}}\mathbf{y}.$$

Lembrando que a equação (-38) pode ser usada para gerar a equação matricial $\mathbf{y} = \Psi\boldsymbol{\theta} + \mathbf{e}$, e substituindo-se tal equação na do MQ, resulta em

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_{\mathrm{MQ}} &= \boldsymbol{\theta} + [\Psi^{\mathrm{T}}\Psi]^{-1}\Psi^{\mathrm{T}}\mathbf{e} \\
&= \boldsymbol{\theta} + \left[\frac{1}{N}\sum_{k=1}^{N}\boldsymbol{\psi}(k-1)\boldsymbol{\psi}^{\mathrm{T}}(k-1)\right]^{-1}\left[\frac{1}{N}\sum_{k=1}^{N}\boldsymbol{\psi}(k-1)e(k)\right] \\
&= \boldsymbol{\theta} + [R_{\psi}]^{-1}\frac{1}{N}\sum_{k=1}^{N}\boldsymbol{\psi}(k-1)e(k), \qquad \text{(-37)}
\end{aligned}$$

sendo R_{ψ} uma matriz de covariância (ver equações (**??**) e (-44), que contêm matrizes análogas). Seria desejável que $\hat{\boldsymbol{\theta}}_{\mathrm{MQ}} \approx \boldsymbol{\theta}$, o que requer que a segunda parcela do lado direito de (-37) seja pequena. Além disso, também seria desejável que $\hat{\boldsymbol{\theta}}_{\mathrm{MQ}} \to \boldsymbol{\theta}$, para $N \to \infty$.

A fim de investigar o comportamento de $\hat{\boldsymbol{\theta}}_{\mathrm{MQ}}$ quando $N \to \infty$, é conveniente supor que $e(k)$ e $u(k)$ são processos estacionários de maneira que matrizes de covariância do tipo

$$R_{uu}(N,\tau) = \frac{1}{N}\sum_{k=1}^{N} u(k)u(k-\tau) \to R_{uu}(\tau), \qquad \text{(-36)}$$

para $N \to \infty$. Neste caso, $R_\psi \to \overline{R}_\psi$ convergirá em probabilidade, uma vez que é composta de somatórios do tipo (-36). Semelhantemente, a segunda parcela do lado direito de (-37) convergirá em probabilidade para $\overline{R}_{\psi e}$. Portanto, desde que R_ψ seja não singular,

$$\hat{\boldsymbol{\theta}}_{\mathrm{MQ}} = \boldsymbol{\theta} + [\overline{R}_\psi]^{-1}\overline{R}_{\psi e}, \quad N \to \infty. \qquad \text{(-35)}$$

Finalmente, para que o estimador seja *consistente*, deve-se ter $\hat{\boldsymbol{\theta}}_{\mathrm{MQ}} \to \boldsymbol{\theta}$ à medida que $N \to \infty$. As condições necessárias para isso são:

- $\overline{R}_{\psi}$ seja não singular. Para isso $u(k)$ e $e(k)$ devem ser independentes e $u(k)$ deve ser persistentemente excitante de ordem suficientemente elevada, conforme discutido no capítulo 4.

Finalmente, para que o estimador seja *consistente*, deve-se ter $\hat{\boldsymbol{\theta}}_{\rm MQ} \to \boldsymbol{\theta}$ à medida que $N \to \infty$. As condições necessárias para isso são:

- $\overline{R}_{\psi}$ seja não singular. Para isso $u(k)$ e $e(k)$ devem ser independentes e $u(k)$ deve ser persistentemente excitante de ordem suficientemente elevada, conforme discutido no capítulo 4.
- $e(k)$ é ruído branco. Neste caso $\overline{R}_{\psi e} = 0$, pois $e(k)$ não dependerá do que tiver acontecido até o instante $k-1$ (note que o vetor de regressores, $\boldsymbol{\psi}(k-1)$ em (-38), só tem informação até o instante $k-1$); ou

Finalmente, para que o estimador seja *consistente*, deve-se ter $\hat{\boldsymbol{\theta}}_{\mathrm{MQ}} \rightarrow \boldsymbol{\theta}$ à medida que $N \rightarrow \infty$. As condições necessárias para isso são:

- $\overline{R}_{\psi}$ seja não singular. Para isso $u(k)$ e $e(k)$ devem ser independentes e $u(k)$ deve ser persistentemente excitante de ordem suficientemente elevada, conforme discutido no capítulo 4.
- $e(k)$ é ruído branco. Neste caso $\overline{R}_{\psi e} = 0$, pois $e(k)$ não dependerá do que tiver acontecido até o instante $k-1$ (note que o vetor de regressores, $\boldsymbol{\psi}(k-1)$ em (-38), só tem informação até o instante $k-1$); ou
- $u(k)$ e $e(k)$ sejam independentes e não haja regressores da saída em $\boldsymbol{\psi}(k-1)$, ou seja, $n_y = 0$ em (**??**), assim $\boldsymbol{\psi}(k-1)$ é independente de $e(k)$.

Finalmente, para que o estimador seja *consistente*, deve-se ter $\hat{\boldsymbol{\theta}}_{\mathrm{MQ}} \to \boldsymbol{\theta}$ à medida que $N \to \infty$. As condições necessárias para isso são:

- $\overline{R}_{\psi}$ seja não singular. Para isso $u(k)$ e $e(k)$ devem ser independentes e $u(k)$ deve ser persistentemente excitante de ordem suficientemente elevada, conforme discutido no capítulo 4.
- $e(k)$ é ruído branco. Neste caso $\overline{R}_{\psi e} = 0$, pois $e(k)$ não dependerá do que tiver acontecido até o instante $k-1$ (note que o vetor de regressores, $\boldsymbol{\psi}(k-1)$ em (-38), só tem informação até o instante $k-1$); ou
- $u(k)$ e $e(k)$ sejam independentes e não haja regressores da saída em $\boldsymbol{\psi}(k-1)$, ou seja, $n_y = 0$ em (**??**), assim $\boldsymbol{\psi}(k-1)$ é independente de $e(k)$.

É interessante notar que algumas das condições listadas acima