Relatório de Atividade 07 - Métodos de estimação paramétricos

Débora Nunes Pinto de Oliveira

Prof. Antonio Marcus, Estimação e Identificação de Sistemas 21.2

21 de setembro de 2021

Esse documento tem por objetivo descrever estimação dos parâmetros de um modelo identificado conforme um modelo paramétrico.

Table of Contents

Introdução
Estimando os parâmetros do modelo por mínimos quadrados
Regressão linear com mínimos quadrados
Análise da polarização da estimativa
Análise probabilística
Análise de correlação do erro do preditor com dados anteriores
Desenvolvimento
Resultados e discussão
Referências bibliográficas

Introdução

A partir da metodologia reportada nas últimas atividades, temos disponível uma família de modelos para parametrizar uma planta desconhecida. Nesse estudo, será considerada o processo genérico modelado por:

$$y(t) = G(q,\theta)u(t) + H(q,\theta)e(t)$$

para $y(t)$ o sinal de saída, $u(t)$ o sinal de entrada e $e(t)$ um ruído branco. Conforme descrito na Atividade 4, para um vetor de parâmetros estimados θ, é possível escrever a equação genérica do preditor como:

$\hat{y}(t|\theta) = W_y(q,\theta)y(t) + W_u(q,\theta)u(t)$, para $W_y(q,\theta) = [1 - H^{-1}(q,\theta)]$ e $W_u(q,\theta) = H^{-1}(q,\theta)G(q,\theta)$

Nesse relatório, o objetivo é, a partir do conjunto de dados Z amostrados do processo, construir uma estimativa adequada de θ. A qualidade dessa estimativa é dada pela minimização do erro de predição:

$$\varepsilon(t,\theta) = y(t) - \hat{y}(t,\theta) = H^{-1}(q,\theta)[y(t) - G(q,\theta)u(t)]$$

O critério de limite para esse problema de otimização pode ser um escalar ou probabilístico.

Estimando os parâmetros do modelo por mínimos quadrados

Inicialmente, será avaliado um método para determinar o "quão grande" é o erro de predição $\varepsilon(t,\theta)$. Considerando $L(q)$ um filtro linear estável, é definido o sinal de erro filtrado:

$$\varepsilon_F(t,\theta) = L(q)\varepsilon(t,\theta) = [L^{-1}(q)H(q,\theta)]^{-1}[y(t) - G(q,\theta)u(t)]$$

tal que a função de custo é dada por:

$$V(\theta, Z) = \frac{1}{N}\sum_{t=1}^{N} \ell\left(\varepsilon_F(t, \theta)\right)$$

para $\ell(\cdot)$ um valor escalar positivo. O objetivo do filtro $L(q)$ é permitir flexibilidade para considerar (ou desconsiderar) propriedades momentâneas do erro de predição, tais como remoção de alta frequências desnecessários na modelagem do problema. Essa transformação de $\varepsilon(t, \theta)$ para $\varepsilon_F(t, \theta)$ é equivalente a transformar o modelo do ruído de $H(q, \theta)$ para $\overline{H}_L(q, \theta) = L^{-1}(q)H(q, \theta)$. Nesse estudo, será considerado $L(q)$ um polinômio mônico e unitário.

Dessa forma, a estimativa $\hat{\theta}$ é encontrada como o mínimo local da função de custo:

$$\hat{\theta} = \arg\min V(\theta, Z)$$

Uma função candidata para $\ell(\cdot)$ seria a norma quadrática $\ell(\varepsilon, \theta, t) = 0.5\varepsilon(t, \theta)^2$. Para $E_N(2\pi k/N, \theta), k = 0, 1, ...N-1$ a TDF de $\varepsilon(t, \theta)$ de N amostras, pela relação de Parseval, é possível aproximar a função de custo por:

$$V(\theta, Z) = \frac{1}{N}\sum_{t=1}^{N}\frac{1}{2}\varepsilon(t, \theta)^2 = \frac{1}{2N}\sum_{k=1}^{N-1}|E_N(2\pi k/N, \theta)|^2$$

Considerando $w(t, \theta) = G(q, \theta)u(t)$, conforme descrito no Teorema 2.1 de [1]:

$$W_N(\omega, \theta) = G(e^{j\omega}, \theta)U_N(\omega) + R_N(\omega)$$

Analogamente, para $s(t, \theta) = y(t) - w(t, \theta)$:

$$S_N(\omega, \theta) = Y_n(\omega) - G(e^{j\omega}, \theta)U_N(\omega) - R_N^*(\omega)$$

Sendo assim, para $\varepsilon(t, \theta) = H^{-1}(q, \theta)s(t, \theta)$:

$$E_N(\omega, \theta) = H^{-1}(e^{j\omega}, \theta)S_N(\omega, \theta) - \widetilde{R}_N(\omega)$$

Dessa forma, $|E_N(2\pi k/N, \theta)|^2$ é encontrado como;

$$|E_N(2\pi k/N, \theta)|^2 = |H(e^{2\pi jk/N}, \theta)|^{-2} \cdot \left|Y_n(2\pi jk/N) - G(e^{2\pi jk/N}, \theta)U_N(2\pi jk/N)\right|^2 + \overline{R}_N$$

Sabendo que $\hat{\hat{G}}_N = Y_N(\omega)/U_N(\omega)$, é possível reescrever a equação acima como:

$$|E_N(2\pi k/N, \theta)|^2 = Q_N(e^{2\pi jk/N}, \theta) \cdot \left|\hat{\hat{G}}(2\pi jk/N) - G(e^{2\pi jk/N}, \theta)\right|^2 + \overline{R}_N$$

para $Q_N(\omega, \theta) = |U_N(\omega)|^2/|H_N(2\pi k/N, \theta)|^2$.

Substituindo a expressão acima na equação da função de custo $V(\theta, Z)$, é evidente que $Q_N(\omega, \theta)$ é o fator de ponderação da média ponderada. Esse peso é então determinada pelo inverso da variância $\alpha_k = \dfrac{|U_N(2\pi k/N)|^2}{\Phi_v(2\pi k/N)}$, ou seja, o inverso da relação sinal ruído.

Para N grande, é possível aproximar o somatório por uma integral de tal forma que:

$$V(\theta, Z) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{2}Q_N(\omega, \theta) \cdot \left|\hat{\hat{G}}(e^{j\omega}) - G(e^{j\omega}, \theta)\right|^2 d\omega$$

Para solucionar o problema de iteração sobre a função de custo para encontrar um mínimo global, pode-se empregar regressão linear. É importante destacar que esse método não é válido para preditores não lineares, como ARMAX (apresentado na Atividade 04).

Regressão linear com mínimos quadrados

O preditor é escrito como:

$$\hat{y}(t|\theta) = \varphi^T\theta + \mu(t)$$

para $\varphi(t) = [-y(t-1)\ -y(t-2)...\ -y(t-n_a)\ u(t-1)\ u(t-2)...\ u(t-n_b)]^T$ o vetor de regressores e $\mu(t)$ um termo conhecido dependente dos dados em instantes anteriores.

Se considerarmos $\mu(t) = 0$, o erro de predição é dado por:

$$\varepsilon(t) = y(t) - \varphi^T\theta$$

Então, para $L(q) = 1$ e $\ell = 0.5\varepsilon^2$, a função de custo é:

$$V(\theta, Z) = \frac{1}{N}\sum_{t=1}^{N}\frac{1}{2}(y(t) - \varphi^T\theta)^2$$

Derivando a função de custo analiticamente, é possível comprovar que o mínimo local é dado pelo vetor de parâmetros:

$$\hat{\theta}_{LS} = \arg\min V(\theta, Z) = \left[\sum_{t=1}^{N}\varphi(t)\varphi^T(t)\right]^{-1}\sum_{t=1}^{N}\varphi(t)y(t) = P(t)\sum_{t=1}^{N}\varphi(t)y(t)$$

na condição que a inversa $P(t)$ existe.

Caso a processo identificado seja variante no tempo, é possível ponderar as amostras mais recentes na função de custo. Sendo assim, é necessário incluir o fator multiplicativo peso a seguir:

$$V(\theta, Z) = \frac{1}{N}\sum_{t=1}^{N}\alpha(t)(y(t) - \varphi^T\theta)^2$$

Análise da polarização da estimativa

Se considerarmos que há um ruído branco de média nula $e(t)$ na saída $y(t)$, é possível considerar esse erro na modelagem ao reescrever a equação de estimação como:

$$\hat{y}(t) = \varphi^T(t)\theta_0 + e(t)$$

Logo, encontra-se que:

$$\hat{\theta}_0 = P(t)\sum_{t=1}^{N}\varphi(t)\left[y(t) + e(t)\right]$$

Sendo assim, o erro entre a estimativa $\hat{\theta}$ não considerando o erro na modelagem e θ_0 considerando o erro na modelagem é dado por:

$$\tilde{\theta} = \hat{\theta} - \theta_0 = P(t)\sum_{t=1}^{N}\varphi(t)e(t)$$

A esperança desse erro é então dado por:

$$E(\tilde{\theta}) = P(t)\sum_{t=1}^{N}\varphi(t)E\left(e(t)\right) = 0$$

Logo, é possível concluir que a estimativa por mínimos quadrados para modelos com ruído branco na medição não possui polarização. Entretanto, se $e(t)$ não for é um ruído branco, não é possível garantir a convergência do algoritmo de recursão dos mínimos quadrados. Nesse caso, é possível caracterizar o ruído como:

$$e(t) = \kappa(q)e^*(t)$$

para $\kappa(q)$ um filtro linear $e^*(t)$ um ruído branco. Reescrevendo o modelo como:

$$A(q)\kappa^{-1}(q)y(t) = B(q)\kappa^{-1}(q)u(t) + e^*(t)$$

para $\kappa^{-1}(q) = L(q)$ na notação da seção anterior, é possível garantir que não irá existir polarização sobre o vetor de parâmetros estimados.

Análise probabilística

É possível estimar θ também a partir de uma análise probabilística. Dessa forma, a modelagem é adequada independentemente da formulação dos dados como processos estocásticos.

Supondo que as observações do sistema são uma variável aleatória y cuja função densidade de probabilidade (*probability density function* - PDF) é:

$$P(y \in A) = \int_{x \in A} f_y(\theta, x) dx$$

Logo, a estimativa $\hat{\theta}$ pode ser escolhido como o valor que maximiza a probabilidade do evento observado y, ou seja:

$$\hat{\theta}_{ML} = \arg\max f_y(\theta, y_*)$$

Essa função é determinística para θ a partir do momento que o valor de y_* é conhecido. Portanto, ela é nomeada função de probabilidade (*likelihood function*).

Por exemplo, considerando y uma variável aleatória independente com distribuição normal tal que $y \in N(\theta_0, \gamma_i)$, a PDF é dada por:

$$f_y(\theta, x) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\lambda_i}} \exp\left[-\frac{(x_i - \theta)^2}{2\lambda_i}\right]$$

O máximo global dessa função equivale a encontrar o máximo global de $\log f_y$. Dessa forma, a estimativa $\hat{\theta}_{ML}$ é:

$$\hat{\theta}_{ML} = \arg\max \log f_y(\theta, y_*) = \frac{\sum_{i=1}^{N} y(i)(\lambda_i)^{-1}}{\sum_{i=1}^{N} (\lambda_i)^{-1}}$$

Para modelos dinâmicos, no entanto, o sinal $y(t)$ é dado por:

$$y(t) = g(t, Z^{t-1}, \theta) + \varepsilon(t, \theta)$$

Como $\varepsilon(t, \theta)$ é um ruído com PDF $f_e(x, t, \theta)$ conhecida, é possível afirmar que:

$$\bar{f}_y(\theta, y) = \prod_{t=1}^{N} f_e\big(\varepsilon(t, \theta), t, \theta\big) = \prod_{t=1}^{N} f_e\big(y(t) - g(t, Z^{t-1}, \theta), t, \theta\big)$$

Logo, maximizar a função acima equivale a maximizar:

$$\frac{1}{N} \log \bar{f}_y(\theta, y) = \frac{1}{N} \sum_{t=1}^{N} \log f_e\big(\varepsilon(t, \theta), t, \theta\big)$$

Se definirmos $\ell(\varepsilon, \theta, t) = -\log f_e(\varepsilon, \theta, t)$, então nós podemos escrever:

$$\hat{\theta}_{ML} = \arg\min \sum_{t=1}^{N} \ell\big(\varepsilon(t, \theta), t, \theta\big)$$

Análise de correlação do erro do preditor com dados anteriores

Idealmente, o erro de predição $\varepsilon(t, \theta)$ deveria ser independente dos dados em instantes anteriores Z^{t-1}. Se essa correlação não for nula, então existe algum dado em $y(t)$ que não está sendo previsto pelo modelo proposto $\hat{y}(t|\theta)$.

Entretanto, é inviável testar se essa condição é verdadeira para o conjunto de dados coletados se N for grande ou $\varepsilon(t, \theta)$ uma transformação não linear. Dessa forma, é preferível selecionar uma sequência finita $\zeta(t)$ de Z^{t-1} e garantir que essa sequência é descorrelacionada de $\varepsilon(t,\theta)$. Matematicamente:

$$\frac{1}{N}\sum_{t=1}^{N}\zeta(t)\varepsilon(t,\theta) = 0$$

A melhor escolha para $\zeta(t)$ depende das propriedades do sistema. Estritamente, é necessário que as variáveis no vetor $\zeta(t)$ sejam descorrelacionadas do ruído $v(t) = H(q)e(t)$ adicionado na saída do modelo.

A solução pelo método dos mínimos quadrados é dada por:

$$\hat{\theta}_{LS} = \text{sol}\left\{\frac{1}{N}\sum_{t=1}^{N}\varphi(t)\left[y(t) - \varphi^T(t)\theta\right] = 0\right\}$$

Sendo assim, é possível adotarmos as variáveis $\zeta(t)$ (nomeadas variáveis instrumentais, do inglês *instrumental variables* - IV),de tal forma que:

$$\hat{\theta}_{LS}^{IV} = \text{sol}\left\{\frac{1}{N}\sum_{t=1}^{N}\zeta(t)\left[y(t) - \varphi^T(t)\theta\right] = 0\right\}$$

Assim, o vetor de parâmetros estimados $\hat{\theta}$ é encontrado por:

$$\hat{\theta}_{LS}^{IV} = \left[\sum_{t=1}^{N}\zeta(t)\varphi^T(t)\right]^{-1}\sum_{t=1}^{N}\zeta(t)y(t) = P(t)\sum_{t=1}^{N}\zeta(t)y(t)$$

É evidente que $P(t)$ deve ser inversível e que a esperança $E[\zeta(t)v(t)] = 0$ para que a estimativa $\hat{\theta}_{LS}^{IV}$ não seja polarizada. Para um modelo ARX, tem-se que:

$$y(t) + a_1y(t-1) + ... + a_{n_a}y(t-n_a) = b_1u(t-1) + ... + b_{n_b}u(t-n_b) + v(t)$$

Uma sugestão para a construção do vetor $\zeta(t)$ seria incluir os elementos não influenciados por $v(t)$, ou seja:

$$\zeta(t) = K(q)[-x(t-1) \ - x(t-2)... \ - x(t-n_a) \ u(t-1) \ u(t-2)... \ u(t-n_b)]^T$$

para $K(q)$ um filtro linear e $x(t)$ dependente da entrada a partir da modelagem pela equação:

$N(q)x(t) = M(q)u(t)$, para $N(q) = 1 + n_1q^{-1} + ... + n_{n_n}q^{-n_n}$ e $M(q) = m_0 + m_1q^{-1} + ... + m_{n_m}q^{-n_m}$

Logo, como $\zeta(t)$ pode ser gerada em malha aberta e depende apenas de valores anteriores de $u(t)$, as variáveis instrumentais independem do ruído $v(t)$.

Uma solução simplificada para encontrar os polinômios seria inicialmente aplicar o método de mínimos quadrados sobre o modelo ARMA completo (com o termo $v(t)$), e, em seguida, sobre os polinômios $N(q)$ e $M(q)$. O vetor de variáveis instrumentais é, então, facilmente encontradas adotando $K(q) = 1$.

Desenvolvimento

Nessa atividade, é proposta a estimação de uma função de transferência $G(z)$ conhecida. Essa planta foi simulada para uma entrada PRBS e um ruído branco $e(t)$ de média nula e variância unitária somado na saída.

O processo ARMA que define $y(t)$ é dado por:

$$y(t) - 1.5y(t-1) + 0.7y(t-2) = u(t-1) + 0.5u(t-2) + e(t)$$

Assim, é evidente que:

$$G_0(q) = \frac{z^{-1} + 0.5z^{-2}}{1 - 1.5z^{-1} + 0.7z^{-2}} = \frac{b_1 z^{-1} + b_2 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}}$$

A letra A do exercício proposto corresponde à Atividade 06. Dessa forma, nesse relatório, serão realizadas as letras B e C. Dessa forma, serão estimados os parâmetros via mínimos quadrados diretamente do vetor de regressão $\varphi(t)$. Em seguida, será utilizado o método das variáveis instrumentais sobre o vetor $\zeta(t)$. Os resultados serão comparados para duas condições de amostragem $N = 100$ e $N = 400$, três escolhas de $K(q)$, $M(q)$ e $N(q)$ e diferentes ordens do vetor de regressão $\varphi(t)$ ($n_a = n_b = 2$ e $n_a = n_b = 3$).

Resultados e discussão

Para o método dos mínimos quadardos, foi adotado o vetor de parâmetros $\theta = [a_1\ a_2\ b_1\ b_2]^T$ e $\varphi(t) = [-y(t-1)\ \ -y(t-2)\ u(t-1)\ u(t-2)]^T$. Observando a função de transferência, é evidente que $\theta_0 = [-1.5\ , 0.7\ , 1.0\ , 0.5]^T$

O resultado dos parâmetros estimados para $u(t)$ um sinal PRBS e $e(t) \in N(0,1)$, pelo método dos mínimos quadrados comum e para $N = 100$ é:

```
N = 100; myLS
```

```
Warning: The PRBS signal delivered is the 100 first values of a full sequence of length 127.
theta_LS = 1×4
   -1.5433    0.7759    1.0800    0.5326
```

Observando o resultado acima, é evidente que o resultado diverge do vetor esperado θ_0. Essa conclusão é justificada pela correlação não nula entre o vetor de regressão $\varphi(t)$ e o ruído $e(t)$. Para dispensar a modelagem do ruído e garantir uma boa estimação dos parâmetros do sistema, é possível empregar variáveis instrumentais.

Para $K(q) = 1$, $N(q) = 1$ e $M(q) = q^{-2}$, tem-se que $x(t) = u(t-2)$ e que o vetor $\zeta(t) = [-x(t-1)\ \ -x(t-2)\ u(t-1)\ u(t-2)]^T$. Sendo assim, é possível estimar os parâmetros estimados $\hat{\theta}$ como:

```
myIV_c1
```

```
theta_IV_c1 = 1×4
   -1.7205    1.1890    1.0126    0.2072
```

Esse resultado se aproxima mais lentamente do vetor esperado θ_0 do que o experimento anterior. Esse resultado pode ser otimizado com melhores escolhas de $K(q)$, $M(q)$ e $N(q)$.

Empregando $K(q) = 1$ e $M(q) = A(q)$ e $N(q) = B(q)$, para $A(q)$ e $B(q)$ os polinômios estimados pelo mínimos quadrados (variável `theta_LS`), é possível encontrar os valores estimados:

```
myIV_c2
```

```
theta_IV_c2 = 1×4
   -1.5938    0.8439    1.0796    0.4706
```

Comparando as variáveis `theta_LS` e `theta_IV_c2`, é evidente que a primeira está mais próxima dos parâmetros reais θ_0.

Considerando mais uma otimização, o último modelo define $K(q) = 1/(1 - 1.5q^{-1} + 0.7q^{-2})$, $M(q) = 1 - 1.5q^{-1} + 0.7q^{-2}$ e $N(q) = 1^{-1} + 0.5q^{-2}$. Esse é um caso especial do terceito teste, considerando que $A(q)$ e $B(q)$ são estimativas perfeitas dos polinômios de $G_0(q)$. O resultado obtido para o vetor de parâmetros estimados é:

```
myIV_c3
```

```
theta_IV_c3 = 1×4
   -1.6256    0.8665    1.1805    0.2358
```

Assim como do experimento anterior, é evidente que o modelo para método dos mínimos quadrados comum produz as estimativas mais próximas de θ_0. Será analisado se essa observação se mantém para maior número de amostras

Aumentaremos N para 400 amostras. Os resultados para os 4 experimentos, em sequância, são:

```
N = 400; myLS
```

```
Warning: The PRBS signal delivered is the 400 first values of a full sequence of length 511.
theta_LS = 1×4
   -1.4986    0.7131    1.0533    0.5010
```

```
myIV_c1
```

```
theta_IV_c1 = 1×4
   -1.4282    0.6171    1.0556    0.5705
```

```
myIV_c2
```

```
theta_IV_c2 = 1×4
   -1.5111    0.7269    1.0524    0.4880
```

```
myIV_c3
```

```
theta_IV_c3 = 1×4
   -1.5509    0.7587    1.0856    0.3199
```

Comparando os resultados acima, é evidente que ambos os experimentos para método dos mínimos quadrados comum e $K(q) = 1$ e $M(q) = A(q)$ e $N(q) = B(q)$ são os mais próximos do vetor θ_0. Analogamente, comparando os resultados para $N = 400$ e $N = 100$, é claro que, com o aumento do número de amostras, há um menor erro de predição.

No caso apresentado, ambas as abordagens (com e sem variáveis de instrumentação) são adequadas para encontrar as estimativas. Uma solução melhor para o conjunto de variáveis instrumentais pode ser obtida procurando um novo conjunto de $K(q)$, $M(q)$ e $N(q)$.

Analisando os resultados obtidos, é possível destacar que o as variáveis instrumentais (diferentemente do mínimos quadrados comum) dispensa a modelagem do ruído. Essa característica é adequada para situações em que não há interesse em concentrar esforços para modelá-lo.

Referências bibliográficas

[1] L. LJUNG. **System Identification: Theory for the User**. Pearson, 1998. 2nd edition, ISBN 9788131744956.