Modelowanie i identyfikacja

Wykład 3: Problem identyfikacji i teoria estymacji

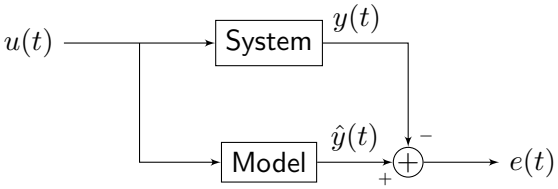
Gniewomir Sarbicki

Identyfikacja

Identyfikacją nazywamy proces budowy modelu matematycznego dla układu rzeczywistego na podstawie pomiarów sygnałów wejściowych i wyjściowych.

Bliskość modelu i systemu rzeczywistego określamy za pomocą **funkcji strat** - pewnej funkcji różnicy sygnałów wyjściowych modelu i układu rzeczywistego:

$$e(t) = \hat{y}(t) - y(t) \tag{1}$$

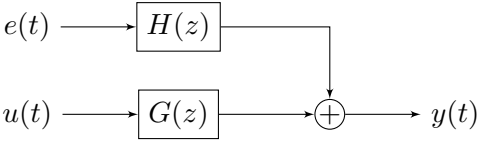


Najprostszym wyborem jest całka z kwadratu tej różnicy.

Na rozwiązanie zadania identyfikacji składają się następujące kroki:

- Wybór struktury modelu - równania różniczkowe, linearyzacja, dyskretyzacja.
- Wybór sygnałów wejściowych
- Estymacja parametrów modelu / estymacja stanu modelu - wyznaczanie parametrów / stanu modelu na podstawie danych doświadczalnych odpowiedzi systemu na sygnały wejściowe.
- Obliczenie funkcji strat
- Akceptacja modelu, lub rozbudowanie jego struktury.

Ogólna struktura modelu liniowego dyskretnego:



$$y(t) = G(z)u(t) + H(z)e(t), \tag{2}$$

gdzie:

- $u(t)$ - sygnał wejściowy
- $y(t)$ - sygnał wyjściowy
- $e(t)$ - zakłócenie o własnościach szumu białego
- $G(z)$ - transmitancja dyskretna toru sterowania
- $H(z)$ - transmitancja dyskretna toru zakłóceń

Jeżeli wyrazimy transmitancje jako stosunki wielomianów:

$$G(z) = \frac{B(z^{-1})}{\tilde{M}(z^{-1})}, \qquad H(z) = \frac{C(z^{-1})}{\tilde{Q}(z^{-1})}, \tag{3}$$

to równanie (2) możemy zapisać jako:

$$y(t) = \frac{B(z^{-1})}{\tilde{M}(z^{-1})}u(t) + \frac{C(z^{-1})}{\tilde{Q}(z^{-1})}e(t). \tag{4}$$

wygodnie jest pomnożyć obie strony równania przez A - najwyższy wspólny dzielnik wielomianów $\tilde{M}$ i $\tilde{Q}$:

$$A(z^{-1})y(t) = \frac{B(z^{-1})}{M(z^{-1})}u(t) + \frac{C(z^{-1})}{Q(z^{-1})}e(t). \tag{5}$$

Rozważa się następujące przypadki szczególne:

- model AR - *AutoRegresive* $(B = 0, C = M = Q = 1)$:

$$y(t) = \frac{1}{A(z^{-1})} e(t) \tag{6}$$

- model MA - *Moving Average* $(B = 0, A = M = Q = 1)$

$$y(t) = C(z^{-1}) e(t) \tag{7}$$

- Model ARMA *AutoRegresive and Moving Average* $(B = 0, M = Q = 1)$:

$$y(t) = \frac{C(z^{-1})}{A(z^{-1})} e(t) \tag{8}$$

- model FIR - *Finite Impulse Response* ($A = M = C = Q = 1$):

$$y(t) = B(z^{-1})u(t) + e(t) \tag{9}$$

- model ARX - *AutoRegresive with eXtra Input* ($M = C = Q = 1$):

$$y(t) = \frac{B(z^{-1})}{A(z^{-1})}u(t) + \frac{1}{A(z^{-1})}e(t) \tag{10}$$

- model ARMAX - *AutoRegresive and Moving Average with eXtra Input* ($M = Q = 1$):

$$y(t) = \frac{B(z^{-1})}{A(z^{-1})}u(t) + \frac{C(z^{-1})}{A(z^{-1})}e(t) \tag{11}$$

- model OE - *Output Error* ($A = C = Q = 1$):

$$y(t) = \frac{B(z^{-1})}{M(z^{-1})}u(t) + e(t) \tag{12}$$

- model BJ *Box-Jenkins* ($A = 1$).

$$y(t) = \frac{B(z^{-1})}{M(z^{-1})}u(t) + \frac{C(z^{-1})}{Q(z^{-1})}e(t) \tag{13}$$

$$
\begin{tikzcd}
\mathrm{GL} \arrow[r, "M = Q = 1"] \arrow[d, "A = 1"] & \mathrm{ARMAX} \arrow[r, "B = 0"] \arrow[d, "C = 1"] & \mathrm{ARMA} \arrow[r, "A = 1"] \arrow[d, "C = 1"] & \mathrm{MA} \\
\mathrm{BJ} \arrow[d, "C = Q = 1"] & \mathrm{ARX} \arrow[r, "B = 0"] \arrow[d, "A = 1"] & \mathrm{AR} & \\
\mathrm{OE} & \mathrm{FIR} & &
\end{tikzcd}
$$

Załóżmy że mamy ciąg liczb $\{X_i\}_{i=1}^N$ i.i.d. wylosowanych z rozkładu p_θ, o nieznanym parametrze $\theta \in \Theta$.

Estymatorem parametru θ nazywamy dowolną funkcję: $\hat{\theta} : \mathcal{X}^N \to \Theta$

Obciążeniem estymatora $\hat{\theta}$ nazywamy wielkość: $B_\theta[\hat{\theta}] = \mathbb{E}_{p_\theta}[\hat{\theta} - \theta]$.

Wariancją estymatora $\hat{\theta}$ nazywamy wartość oczekiwaną kwadratu różnicy estymatora i parametru estymowanego: $\mathrm{Var}_\theta[\hat{\theta}] = \mathbb{E}((\hat{\theta} - \theta)^2) - (\mathbb{E}(\hat{\theta} - \theta))^2$.

(średni kwadrat błędu estymatora jest sumą kwadratu obciążenia i wariancji).

Przykład: $\sum_{i=1}^N (X_i - \bar{X})^2/N$ jest obciążonym estymatorem wariancji rozkładu σ^2. $\sum_{i=1}^N (X_i - \bar{X})^2/(N-1)$ jest nieobciążonym estymatorem σ^2.

Dobrym wyborem estymatora byłby estymator o minimalnej wariancji wśród estymatorów nieobciążonych.

Problem: estymator nieobciążony może nie istnieć, reparametryzacja niszczy warunek nieobciążenia.

Dla ustalonego wektora parametrów rozkładu θ znamy funkcję $f(\vec{y}_N|\theta)$ gęstości prawdopodobieństwa otrzymania wyjścia $\vec{y}_N$. Ze wzoru na prawdopodobieństwo warunkowe otrzymujemy rozkład prawdopodobieństwa wektora parametrów θ:

$$f(\theta|\vec{x}_N) = f(\vec{x}_N|\theta)g(\theta)/f(\vec{x}_N), \quad \text{gdzie} \quad f(\vec{x}_N) = \int_\theta f(\vec{x}_N|\theta)g(\theta)\mathrm{d}\theta \tag{14}$$

Rozkład $g(\theta)$ jest rozkładem *a priori* parametrów układu. Wartość $\hat{\theta}$, która maksymalizuje (14) nazywamy *estymatorem maksymalnego prawdopodobieństwa*.

Estymator $\hat{\theta}$ może być:

- asymptotycznie nieobciążony: $\lim_{N\to\infty} B_\theta[\hat{\theta}_N] = 0$
- zgodny, jeżeli: $\lim_{N\to\infty} \hat{\theta}_N = \theta$

Przykład: Prawdopodobieństwo sukcesu w jednym losowaniu wynosi θ. Prawdopodobieństwo n sukcesów w N próbach wynosi:

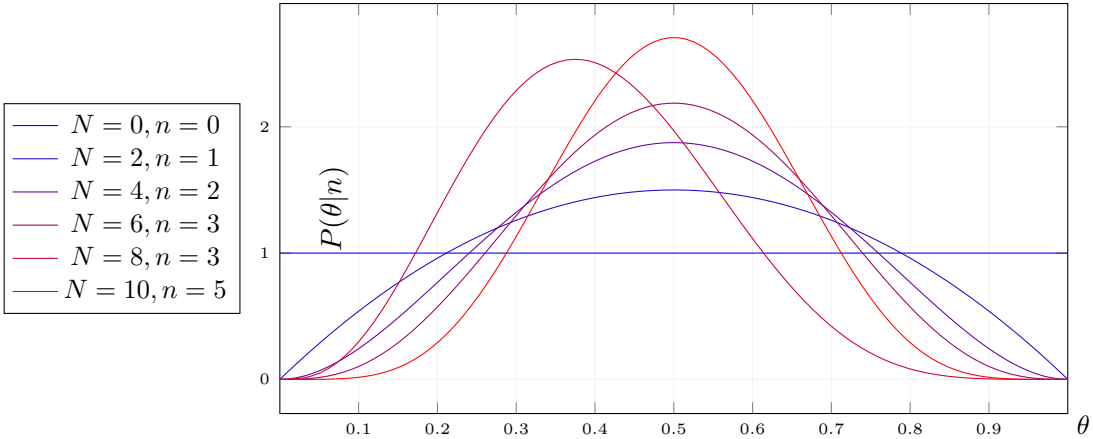
$$P(n|\theta) = \binom{N}{n}\theta^n(1-\theta)^{N-n} \tag{15}$$

Za rozkład *apriori* parametru θ weźmiemy rozkład jednostajny na odcinku $[0,1]$ (maksymalizujący entropię). Jeżeli w N próbach otrzymaliśmy n sukcesów, przy nieznanym parametrze θ, jego rozkład prawdopodobieństwa wynosi:

$$P(\theta|n) = \frac{\binom{N}{n}\theta^n(1-\theta)^{N-n}}{\int_0^1 \binom{N}{n}\theta^n(1-\theta)^{N-n}\mathrm{d}\theta} = (N+1)\binom{N}{n}\theta^n(1-\theta)^{N-n} \tag{16}$$

Powyższy rozkład osiąga wartość maksymalną dla wartości $\frac{n}{N}$. Jest to *estymator maksymalnego prawdopodobieństwa* dla θ. Jest on obciążony, ale asymptotycznie nieobciążony.
Jego wariancja wynosi $\frac{(n+1)(N-n+1)}{(N+2)^2(N+3)}$

Załóżmy, że otrzymaliśmy w dziesięciu próbach wyniki: $[0, 1, 1, 0, 1, 0, 0, 0, 1, 1]$. rozkład prawdopodobieństwa θ zmienia się w następujący sposób:



Rozkład jednostajny ma wadę - przestaje być jednostajny po zmianie parametru.

Definiujemy informację Fishera:

$$I_X(\theta) = \sum_{x \in \mathcal{X}} \left(\frac{\mathrm{d}}{\mathrm{d}\theta} \log_2 f(x|\theta) \right)^2 p(x|\theta) \tag{17}$$

Oznaczającą ilość informacji w zmiennej losowej X o parametrze θ.

Za rozkład startowy bierzemy (Jeffrey's prior):

$$g(\theta) = \sqrt{I_X(\theta)}/V, \quad V = \int_\Theta \sqrt{I_X(\theta)} \mathrm{d}\theta \tag{18}$$

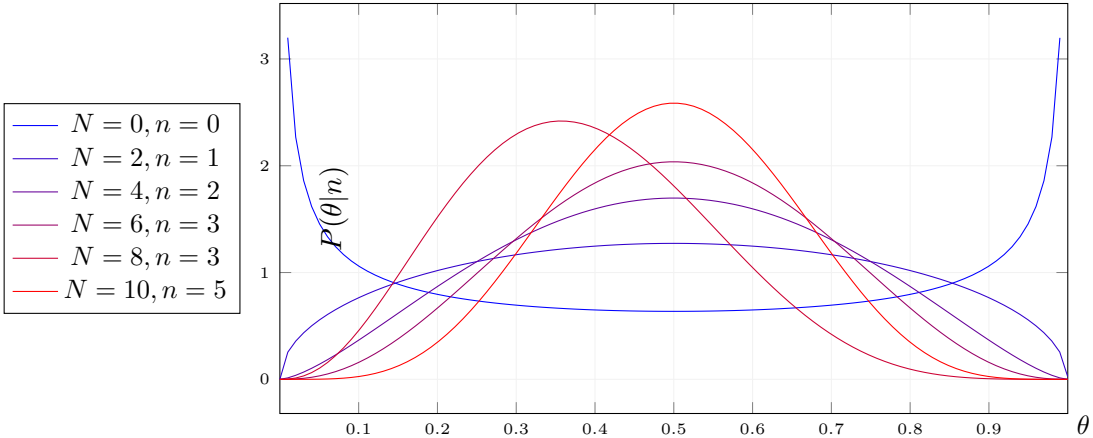
Dla rozkładu Bernouliego jest to:

$$p(\theta) = \frac{1}{\pi\sqrt{\theta(1-\theta)}} \tag{19}$$

a rozkład parametru θ po N próbach i n sukcesach wynosi:

$$P(\theta|n) = \binom{N}{n}\frac{\theta^n(1-\theta)^{N-n}}{\pi\sqrt{\theta(1-\theta)}} / \int_0^1 \binom{N}{n}\frac{\theta^n(1-\theta)^{N-n}}{\pi\sqrt{\theta(1-\theta)}}\mathrm{d}\theta \tag{20}$$

Dla dużych N rozkład $P(\theta|n)$ jest rozkładem Gaussa o wariancji $(nI_X(\theta))^{-1}$.


$N = 0, n = 0$
$N = 2, n = 1$
$N = 4, n = 2$
$N = 6, n = 3$
$N = 8, n = 3$
$N = 10, n = 5$
$P(\theta|n)$
0
1
2
3
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
θ