Procesos Estocásticos y Sistemas

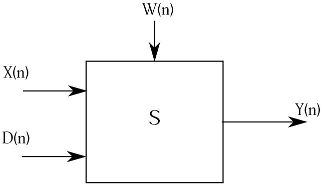
Cecilia Galarza

Procesos Estocásticos
Facultad de Ingeniería, Universidad de Buenos Aires

1er cuatrimestre 2025

Identificación de sistemas

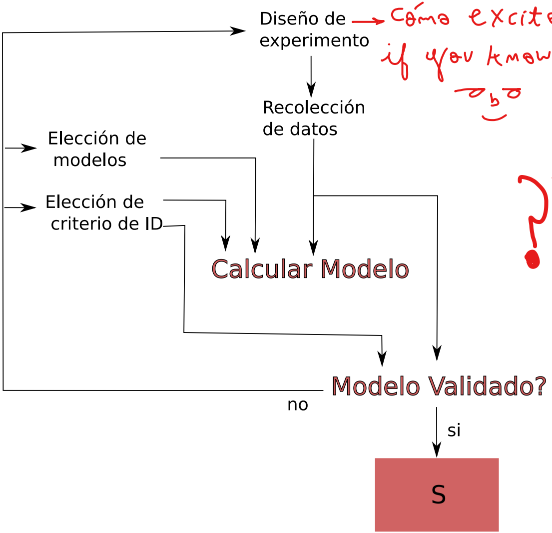
Un sistema general, tiene un proceso de entrada, uno de salida y sufre perturbaciones, algunas observables otras no.



$X(n)$, $Y(n)$: par (entrada, salida)
$D(n)$: perturbación observables
$W(n)$: perturbación no-observable (ruido)

Identificación de sistemas: procedimiento general

Se ve q' hay materias enteras de esto



Cómo excita el sistema
if you know what I mean

???

Identificación de sistemas

El problema de identificación del sistema S, utiliza realizaciones de $(X(n), Y(n))$, eventualmente de $D(n)$ para caracterizar la dinámica del sistema. A modo de introducción, vamos a desarrollar la solución cuando

- S es un sistema LTI caracterizado por su respuesta impulsiva $h(n)$ y la respuesta en frecuencia $H(\omega)$.
- $D(n) = 0$.
- $X(n)$ es un proceso ESA.

Identificación de sistemas usando $S_{XY}(\omega)$

- Supongamos que un sistema desconocido $H(\omega)$ es excitado con ruido blanco $X(n)$ con PSD $S_X(\omega) = \sigma^2$. Entonces,

$$R_{XY}(k) = h(k) * \underbrace{R_X(k)}_{\sigma^2\delta(k)} \implies S_{X,Y}(\omega) = \sigma^2 H(\omega).$$

- Sea $\widehat{S}_{X,Y}(\omega)$ una estimación de la PSD cruzada obtenida a partir de una realización de la salida del sistema $y(0), y(1), \cdots y(L)$. Planteamos

$$\widehat{H}(\omega) = \frac{\widehat{S}_{X,Y}(\omega)}{\sigma^2}.$$

Identificación de sistemas usando $S_{XY}(\omega)$

- Éste es un método de identificación *no paramétrico*.
- No es necesario conocer la estructura del sistema (cuántos polos, ceros, etc).
- La calidad de la estimación de H depende directamente de la calidad de la estimación de $S_{X,Y}$.

Identificación paramétrica

Supongamos que la respuesta en frecuencia $H(\omega)$ se modela como

¿Y K?

$$H(\omega) = \frac{1 + \sum_{k=1}^{M} b_k e^{-\jmath \omega k}}{1 + \sum_{k=1}^{N} a_k e^{-\jmath \omega k}} \quad , \quad \omega \in [-\pi, \pi]$$

Esto impone la cantidad de polos y ceros. Es más restrictivo

para N y M conocidos. Identificar $H(\omega)$ es hallar

$$\{\hat{a}_k, k = 1, \cdots N\} \quad \text{y} \quad \{\hat{b}_k, k = 1, \cdots M\}$$

a partir de las realizaciones $\{y(n), n = 0, \cdots L\}$.

Identificación paramétrica

Este problemita tan inocente es casi imposible de resolver con algoritmos.

Se podría plantear el problema de optimización

minimizo la diferencia entre $\hat{S}$ y S variando ak y bk

$$\min_{a_k, b_k} \int_{-\pi}^{+\pi} \left| \hat{S}_{XY}(\omega) - \sigma^2 H(\omega) \right|^2 d\omega$$

seguramente termine acá

Problema no-convexo

Acá quiero llegar

Pero éste es un problema no-convexo y tiene múltiples mínimos locales. Vamos a utilizar la estructura del modelo para salvar este problema cuando sea posible. Para ello, asociamos la salida del sistema a distintos tipos de procesos:

La idea sería: estimo S^ con realizaciones de y y luego busco los ak, bk que minimicen el error para hallar H(w)

- Si $\forall \omega, \sum_{k=1}^{M} b_k e^{-\jmath \omega k} = 0 \quad \longrightarrow \quad$ proceso AR.
- Si $\forall \omega, \sum_{k=1}^{N} a_k e^{-\jmath \omega k} = 0 \quad \longrightarrow \quad$ proceso MA.
- En el caso general, se asocia el modelo a un proceso ARMA.

Modelado por procesos autoregresivos

- Son sistemas sólo con polos, lo que permite modelar señales de banda angosta con picos colocando polos cerca de la circunferencia unitaria.

- El sistema $h(k)$ es causal y estable con respuesta impulsiva de duración infinita

$$Y(k) = \sum_{q=0}^{\infty} h(q)X(k-q) \quad (h[0] = 1).$$

- La estimación de los coeficientes del AR se obtienen resolviendo ecuaciones lineales, lo que lo hace un técnica atractiva.

Recordando Yule-Walker

Ecuaciones YW

$$R_Y(p) + \sum_{i=1}^{N} a_i R_Y(p-i) = \sigma^2 \delta(p) \qquad p = 0, \cdots N$$

Nota: Si $h(0) \neq 1$, entonces tomamos $\sigma^2 = \sigma_X^2 h(0)$.

Sistema de ecuaciones con incógnitas $\mathbf{a}_N$ y σ^2.

$$\begin{cases} \begin{bmatrix} R_Y(0) & \mathbf{r}_N^t \end{bmatrix} \begin{bmatrix} 1 \\ \mathbf{a}_N \end{bmatrix} = \sigma^2 \\ \mathbf{r}_N + \mathbf{R}_N \mathbf{a}_N = \mathbf{0} \end{cases}$$

Identificación utilizando ecuaciones YW

ESTARÍA BUENO ENTENDER ESTO

- Seleccionar el orden del modelo N
- Estima $\hat{R}_Y(k), k = 0 \cdots N$ utilizando el estimador sesgado
- Armar la matriz $\hat{\mathbf{R}}_N$ y el vector $\hat{\mathbf{r}}_N$.
- Estiman los coeficientes del AR y la varianza de la entrada

$$\hat{\mathbf{a}}_N = -\hat{\mathbf{R}}_N^{-1}\hat{\mathbf{r}}_N \qquad \sigma^2 = \hat{R}_Y(0) - \hat{\mathbf{r}}_N^t\hat{\mathbf{R}}_N^{-1}\hat{\mathbf{r}}_N$$

Solución alternativa

Volvemos a la descripción del proceso de salida del sistema con entrada ruido blanco de media nula,

$$\underbrace{Y(n) + \sum_{i=1}^{N} a_i Y(n-i)}_{V(n-1)} = X(n) \qquad \begin{cases} \mathbb{E}[X(n)] = 0 \\ R_X(k) = \sigma_X^2 \delta(k) \end{cases}$$

En el instante n y habiendo observado $y(n-1), y(n-2), \cdots,$ contamos con una realización del término $V(n-1)$. Luego,

$$\begin{aligned} \mathbb{E}\Big[[Y(n)\big| Y(n-1), \cdots, Y(n-N)\Big] &= \mathbb{E}\Big[X(n) - V(n-1)\big| V(n-1)\Big] \\ &\underset{\downarrow}{=} -v(n-1) \end{aligned}$$

$E[X(n)] = 0$

Solución alternativa: error de predicción

Predicción de un paso en adelante para $Y(n)$ (*one-step ahead prediction*)

Esto es la mejor apuesta que puedo hacer con esta información
Mentira, la mejor apuesta es todo al rojo, siempre y en todo lugar

$$\hat{y}(n|n-1) = \mathbb{E}[Y(n)|Y(n-1), \cdots Y(n-N)]$$ $=-V(n-1)$

El desarrollo anterior asume que también conocemos $\mathbf{a}_N$, es decir

$$\hat{y}(n|n-1;\mathbf{a}_N) = \begin{bmatrix} -y(n-1) & \cdots & -y(n-N) \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_N \end{bmatrix}$$

Vector de coeficientes

Error de predicción:

$$|y(n) - \hat{y}(n|n-1;\mathbf{a}_N)| = |y(n) + \sum_{i=1}^{N} a_i y(n-i)|$$

Solución alternativa: error de predicción de varianza mínima

Un criterio de identificación es hallar los coeficientes que minimicen la varianza del error de predicción, es decir

$$\min_{\mathbf{a}_N} \mathbb{E}\left[|Y(n) - \hat{y}(n|n-1; \mathbf{a}_N)|^2\right]$$

Desarrollando el error de predicción,

$$\left|Y(n) + \sum_{i=1}^{N} a_i y(n-i)\right| = \left|Y(n) + \begin{bmatrix} y(n-1) & y(n-2) & \cdots y(n-N) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}\right|$$

$$\hat{y}(n) = -V(n-1) = -\begin{bmatrix} y(n-1) & \cdots & y(n-N) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}$$

Solución alternativa: error de predicción de varianza mínima

A partir de la realización $y(0), \cdots, y(L)$, $L \gg N$, definimos:

Esto es la ecuación $\circledast$ para distintos instantes de tiempo

$$\boldsymbol{y} = \begin{bmatrix} y(N) \\ y(N+1) \\ \vdots \\ y(L) \end{bmatrix} \quad \mathbf{Y} = \begin{bmatrix} y(N-1) & y(N-2) & \cdots & y(0) \\ y(N) & \ddots & \cdots & y(1) \\ & \vdots & & \\ y(L-1) & y(L-2) & \cdots & y(L-N) \end{bmatrix}.$$

Para cada instante $N, N+1, \cdots L$, el error de predicción resulta

$$\text{ERROR} = \begin{bmatrix} y(N) \\ y(N+1) \\ \vdots \\ y(L) \end{bmatrix} + \begin{bmatrix} y(N-1) & y(N-2) & \cdots & y(0) \\ y(N) & \ddots & \cdots & y(1) \\ & \vdots & & \\ y(L-1) & y(L-2) & \cdots & y(L-N) \end{bmatrix} \mathbf{a}_N.$$

Solución alternativa: error de predicción de varianza mínima

- Aproximamos la varianza del error de predicción por $\|\boldsymbol{y} + \mathbf{Y}\,\mathbf{a}_N\|^2$.
- El problema a resolver es entonces

$$\min_{\mathbf{a}_N} \|\boldsymbol{y} + \mathbf{Y}\,\mathbf{a}_N\|^2$$

ESTE problema sí es convexo

Estimador LS

El estimador de los parámetros del AR por cuadrados mínimos es (si las columnas de $\mathbf{Y}$ son LI):

La puta madre que te remil parió

SOLUCIÓN ÚNICA

$$\hat{\mathbf{a}}_N = -(\mathbf{Y}^t\mathbf{Y})^{-1}(\mathbf{Y}^t\boldsymbol{y}).$$

Creo que es la pseudo inversa de los hijos de remil puta de moore-penrose

Breve repaso de álgebra para resolver $\min ||\boldsymbol{y} + \mathbf{Y}\,\mathbf{a}||^2$

- Sea $\mathbf{Y} \in \mathbb{R}^{p \times q}$ (en este caso, $p = L - N + 1$, y $q = N$).
- A partir de $\mathbf{Y}$ se definen dos espacios complementarios, $col(\mathbf{Y})$ y $ker(\mathbf{Y}^t)$ tal que $\forall\, \boldsymbol{y} \in \mathbb{R}^p$,

nucleo de Y^t

$$\boldsymbol{y} = \hat{\boldsymbol{y}} + \boldsymbol{y}_0 \qquad \hat{\boldsymbol{y}} \in col(\mathbf{Y})\ ,\ \boldsymbol{y}_0 \in ker(\mathbf{Y}^t), \quad \hat{\boldsymbol{y}} \perp \boldsymbol{y}_0.$$

- *Observación:* $\mathbf{Y}\,\mathbf{a} \in col(\mathbf{Y})$. Luego, se necesita encontrar el elemento de $col(\mathbf{Y})$ más cercano a $(-\boldsymbol{y})$, es decir, la projección ortogonal de $-\boldsymbol{y}$ sobre $col(\mathbf{Y})$. Ésta es $\hat{\boldsymbol{y}} = -\mathbf{Ya}$. $ker(Y^T) = \{ v : v^t . Y = 0,\ v \in \mathbb{R}^p \}$
- Integrando estas ideas, $Y^t . y_0 = 0$

$$\mathbf{Y}^t\boldsymbol{y} = \mathbf{Y}^t\left(\hat{\boldsymbol{y}} + \boldsymbol{y}_0\right) = \mathbf{Y}^t\hat{\boldsymbol{y}} = -\mathbf{Y}^t\mathbf{Y}\mathbf{a}_N \longrightarrow \mathbf{Y}^t\boldsymbol{y} = -\mathbf{Y}^t\mathbf{Y}\mathbf{a}$$

matriz cuadrada

- Si $\mathbf{Y}$ de rango completo, $\mathbf{Y}^t\mathbf{Y}$ invertible,

Se puede ret. Como diría Johnny Viale ¿LA res?

$$\mathbf{a} = -\left[\mathbf{Y}^t\mathbf{Y}\right]^{-1}\mathbf{Y}^t\boldsymbol{y}$$

Modelado AR: comparación técnicas

Usando YW con estimación de la autocorrelación

$$\hat{\mathbf{a}}_N = -\boxed{\hat{\mathbf{R}}_N}^{-1}\boxed{\hat{\mathbf{r}}_N}$$

Usando mínimos cuadrados

$$\hat{\mathbf{a}}_N = -\boxed{(\mathbf{Y}^t\mathbf{Y})}^{-1}\mathbf{Y}^t\underline{\boldsymbol{y}}$$

- $(\mathbf{Y}^t\mathbf{Y})$ es una estimación de $\hat{\mathbf{R}}_N$, $(\mathbf{Y}^t\boldsymbol{y})$ una estimación de $\mathbf{r}_N$
- El método LS no requiere estimar $\hat{R}_Y$.
- Pero, cuando L es pequeño, $\mathbf{Y}^t\mathbf{Y}$ no es necesariamente positiva
- Para ~~N~~ grande ambas técnicas obtienen resultados similares.
  L

Sistemas de promedio móvil

- Sistemas de sólo ceros, con polo múltiple en el origen

$$Y(n) = X(n) + \sum_{i=1}^{M} b_i X(n-i)$$

- Se asume que el *orden del modelo M* es conocido.
- A partir de la observación de $y(0), \cdots y(L)$, se necesita estimar

$$b_1, \cdots b_M.$$

Volvemos a la predicción de un paso en adelante

Volviendo a la definición del sistema

$$Y(n) = X(n) + \sum_{i=1}^{M} b_i X(n-i)$$

Como en el caso, anterior, obtenemos la predicción de un paso

$$\hat{y}(n|n-1; \mathbf{b}_M) = \mathbb{E}\left[Y(n)\middle|Y(n-1,\cdots)\right] = \sum_{i=1}^{M} b_i x(n-i)$$

donde $\mathbf{b}_M$ es el vector de parámetros y $x(k)$ es la realización de la entrada. El error de predicción resulta

$$\varepsilon(n; \mathbf{b}_M) = y(n) - \hat{y}(n|n-1; \mathbf{b}_M)$$

Volvemos a la predicción de un paso en adelante

Si $\mathbf{b}_M$ corresponde a los parámetros del sistema,

$$\varepsilon(n; \mathbf{b}_M) = x(n).$$

Luego,

$$\begin{aligned}
\varepsilon(n; \mathbf{b}_M) &= y(n) - \sum_{i=1}^{M} b_i \varepsilon(n-i; \mathbf{b}_M) \\
&= y(n) - \underbrace{\begin{bmatrix} \varepsilon(n-1; \mathbf{b}_M) & \varepsilon(n-2; \mathbf{b}_M) & \cdots & \varepsilon(n-M; \mathbf{b}_M) \end{bmatrix}}_{\boldsymbol{\varepsilon}(n-1; \mathbf{b}_M)^t} \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_M \end{bmatrix}}_{\mathbf{b}_M}
\end{aligned}$$

Identificación de menor error de predicción

Dada una observación $y(0), \cdots y(L)$ podemos plantear como antes la minimización de la varianza del error de predicción

$$\min_{\mathbf{b}_M} \sum_{n=M}^{L} \left| y(n) - \varepsilon(n; \mathbf{b}_M)^t \mathbf{b}_M \right|^2$$

- Ahora el problema es no-lineal dado que $\varepsilon(n; \mathbf{b}_M)$ depende de $\mathbf{b}_M$.
- Un modo de resolverlo de modo iterativo
  - Dado $\mathbf{b}_M^{(0)}$, obtener $\varepsilon(n; \mathbf{b}_M^{(0)})$
  - Resolver problema de mínimos cuadrados $\longrightarrow$ $\mathbf{b}_M^{(1)}$
  - Iterar

Modelado ARMA

- El modelo ARMA es más general. Tiene polos y ceros
- Sea $H(z)$ la transferencia del modelo a identificar (se asume estable):

$$H(z) = \frac{B(z)}{A(z)} = \frac{1 + b_1 z^{-1} + b_M z^{-M}}{1 + a_1 z^{-1} + a_N z^{-N}}$$

- El problema es obtener a_i y b_i.

Modelado ARMA: técnica LS en dos etapas

- Si fuera posible conocer la secuencia de entradas $x(k)$ al ARMA, podría utilizarse una técnica para estimar la respuesta del sistema como un problema de identificación de un sistema.
- La técnica LS en dos pasos se basa en estimar primero dicha secuencia.
- El modelo ARMA puede elegirse de fase mínima (ceros estables) de modo que tiene inversa estable y puede interpretarse a $X(k)$ como la salida mientras que $Y(k)$ es la entrada:

$$X(z) = \frac{A(z)}{B(z)} Y(z) = V(z) Y(z)$$

- Al ser causal, la respuesta de este “nuevo” sistema puede expandirse como una respuesta impulsiva infinita:

$$x(k) = y(k) + \sum_{p=1}^{\infty} v[p] y[k-p],$$

Modelado ARMA: técnica LS en dos etapas

- La respuesta $v(k)$ tiene infinitos coeficientes, no puede ser estimada con una cantidad finita de muestras.
- Sin embargo puede truncarse la respuesta a un orden K que contenga la mayor parte de la energía:

$$\hat{x}(k) = y(k) + \sum_{p=1}^{K} \tilde{v}[p]y[p-k].$$

- Se puede resolver el problema como un AR de orden K.
- Utilizando la solución $\tilde{v}[p]$, $p = 1, \cdots K$, se estima la entrada del ARMA: $\{\hat{x}[K+1], \ldots, \hat{x}[N]\}$.

Modelado ARMA: técnica LS en dos etapas (III)

- Se definen:

$$\boldsymbol{w} \triangleq \left[\begin{array}{ccc|ccc} a_1 & \dots & a_N & b_1 & \dots & b_M \end{array} \right]^t$$

y

$$\boldsymbol{u}(k) = \left[\begin{array}{ccc|ccc} y[k-1] & \dots & y[k-N] & -\hat{x}[k-1] & \dots & -\hat{x}[k-M] \end{array} \right]^t.$$

- Volviendo al modelo ARMA, tenemos:

$$y(k) + \boldsymbol{u}(k)^t \boldsymbol{w} = \hat{x}(k) \qquad k \geq K.$$

Modelado ARMA: técnica LS en dos etapas (IV)

Sea $T = K + M$. Usando una realización $y(0), \cdots y(L)$, definimos:

$$\mathbf{Z} \triangleq \begin{bmatrix} y[T-1] & \dots & y[T-N] & -\hat{x}[T-1] & \dots & -\hat{x}[T-M] \\ y[T] & \dots & y[T-N+1] & -\hat{x}[T] & \dots & -\hat{x}[T-M+1] \\ \vdots & & \vdots & \vdots & & \vdots \\ y[L-1] & \dots & y[L-N] & -\hat{x}[L-1] & \dots & -x[L-M] \end{bmatrix}$$

$$\boldsymbol{z} = \begin{bmatrix} y[T] & y[T+1] & \dots & y[L] \end{bmatrix}^t$$

$$\hat{\boldsymbol{x}} = \begin{bmatrix} \hat{x}[T] & \hat{x}[T+1] & \dots & \hat{x}[L] \end{bmatrix}^t$$

$$\min_{\boldsymbol{w}} \|\boldsymbol{z} + \mathbf{Z}\boldsymbol{w}\|^2$$

$$\hat{\boldsymbol{w}} = -(\mathbf{Z}^t\mathbf{Z})^{-1}(\mathbf{Z}^t\boldsymbol{z}).$$

Modelado ARMA: técnica LS en dos etapas (V)

Algunas observaciones:

- El estimador es positivo por construcción.
- Por el truncado del modelo AR, el estimador tiene sesgo. Tomando K grande se compensa esto.
- Sin embargo, K no debería ser tan grande que reduzca la exactitud del segundo paso.