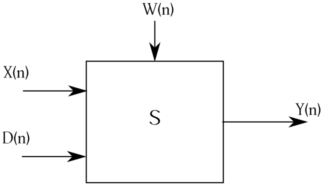
Modelado e Identificación

Cecilia Galarza

Procesos Estocásticos
Facultad de Ingeniería, Universidad de Buenos Aires

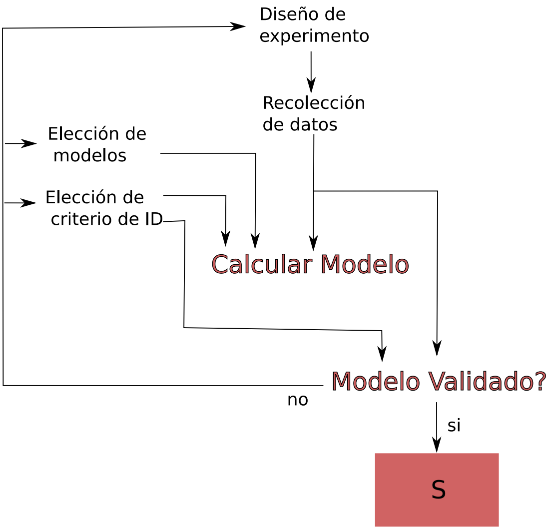
Identificación de sistemas

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



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,

$$S_{X,Y}(\omega) = \sigma^2 H(\omega).$$

- Sea $\widehat{S}_{X,Y}(\omega)$ una estimación de la PSD cruzada. 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 de espectros racionales

- Una función real $S(\omega)$ es una función racional de $e^{\jmath\omega}$ si se puede escribir como el cociente de polinomios en $e^{\jmath\omega}$,

$$S(\omega) = \frac{\displaystyle\sum_{k=-m}^{m} b_k e^{-\jmath\omega k}}{\displaystyle\sum_{k=-n}^{n} a_k e^{-\jmath\omega k}}.$$

- En particular, estamos interesados en el caso que los coeficientes satisfacen $b_k = b^*_{-k}$ y $a_k = a^*_{-k}$.
- Cuando estas funciones son positivas pueden aproximar la PSD de muchos procesos. Por esta razón, son muy atractivas para resolver el problema de identificación

Estimación Paramétrica

- Para $S(\omega)$ función racional, a_k y b_k son los parámetros que la definen.
- Se podría plantear el problema de optimización

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

- La solución a este problema tiene varios inconvenientes
  - Cómputo de $\hat{S}_{XY}(\omega)$.
  - En forma general, problema no-convexo, múltiples mínimos locales.
  - Si se quiere garantizar estabilidad, problema de optimización con restricciones.
  - .....

Identificación de sistemas: Observación

- Vimos que cuando $S_X(\omega) = \sigma^2$,

$$S_{XY}(\omega) = \sigma^2 H(\omega)$$

.

- Por otro lado, sabemos que

$$S_Y(\omega) = \sigma^2 |H(\omega)|^2$$

.

- De este modo, el modelo de una PSD $S(\omega)$ puede convertirse en un modelo para la señal misma.
- En este contexto, la identificación del sistema $H(\omega)$ o de la PSD $S(\omega)$ son dos problemas equivalentes.

Identificación con espectros racionales

- Vamos a utilizar la estructura del modelo para hacer la identificación. Esto da lugar a varios modelos:
  - Si $\forall\omega,\ \sum_{k=-n}^{n} a_k e^{-\jmath\omega k} = 1$ el sistema es modelado mediante un proceso de tipo MA (*moving average*).
  - Si $\forall\omega,\ \sum_{k=-m}^{m} b_k e^{-\jmath\omega k} = 1$, el sistema es modelado mediante un proceso autoregresivo (AR).
  - En el caso general, el modelo es 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 proceso puede describirse como la salida de un sistema $h(k)$ 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.

Modelado por procesos autoregresivos

- Vimos que el proceso AR de orden *n* (n-AR) se caracteriza por la ecuación en diferencias:

$$Y(k) + \sum_{i=1}^{n} a_i Y(k-i) = X(k). \tag{1}$$

- Se asume que el *orden del modelo n* es conocido.
- Se necesita estimar

$$a_1, \cdots a_n.$$

Modelado AR: ecuaciones de Yule-Walker

- Parto de (1) y multiplico ambos miembros por $Y^*(k-p)$, $p \geq 0$:

$$\left\{ Y(k) + \sum_{i=1}^{n} a_i Y(k-i) \right\} Y^*(k-p) \; = \; X(k)Y^*(k-p).$$

- Tomo esperanza:

$$\mathbb{E}\left[Y(k)Y^*(k-p)\right] + \sum_{i=1}^{n} a_i \mathbb{E}\left[Y(k-i)Y^*(k-p)\right] = \mathbb{E}\left[X(k)Y^*(k-p)\right]$$

- Obtengo ecuación en diferencias para R_Y

$$R_Y(p) + \sum_{i=1}^{n} a_i R_Y(p-i) = \mathbb{E}[X(k)Y^*(k-p)]. \tag{2}$$

Modelado AR: ecuaciones de Yule-Walker

- El sistema es causal, luego,

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

- Reemplazo en $\mathbb{E}[X(k)Y^*(k-p)]$

$$\mathbb{E}[X(k)Y^*(k-p)] = \sum_{q=0}^{\infty} h^*[q]\underbrace{\mathbb{E}[X(k)X^*(k-p-q)]}_{R_X(p+q)}.$$

- $X(k)$ es ruido blanco y $R_X(q) = \sigma_X^2\delta(q)$. Luego,

$$\mathbb{E}[X(k)Y^*(k-p)] = \sigma_X^2 \sum_{q=0}^{\infty} h^*[q]\delta(p+q) = \sigma_X^2 h^*(-p).$$

- Sistema causal, luego $h(-p) = 0$ si $p > 0$. Reemplazo en (2).

Modelado AR: ecuaciones de Yule-Walker

Ecuaciones de YW

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

donde $\sigma^2 = \sigma_X^2 h(0)$.

Modelado AR: ecuaciones de Yule-Walker

Con las ecuaciones YW planteamos un sistema de $n+1$ ecuaciones con $n+1$ incógnitas:

$$\begin{bmatrix} R_Y(0) & R_Y(-1) & \dots & R_Y(-n) \\ R_Y(1) & R_Y(0) & \dots & R_Y(-n+1) \\ \vdots & & \ddots & R_Y(-1) \\ R_Y(n) & \dots & & R_Y(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad (3)$$

Despejar las incógnitas

$$a_1, \cdots a_n, \sigma^2.$$

Modelado AR: técnica de la autocovarianza

- Particionamos (3)

$$\left[\begin{array}{c|ccc} R_Y(0) & R_Y(-1) & \dots & R_Y(-n) \\ \hline R_Y(1) & R_Y(0) & \dots & R_Y(-n+1) \\ \vdots & & \ddots & R_Y(-1) \\ R_Y(n) & \dots & & R_Y(0) \end{array}\right] \left[\begin{array}{c} 1 \\ \hline a_1 \\ \vdots \\ a_n \end{array}\right] = \left[\begin{array}{c} \sigma^2 \\ \hline 0 \\ \vdots \\ 0 \end{array}\right]$$

- Definimos:

$$\mathbf{r}_n \triangleq \begin{bmatrix} R_Y(1) \\ \vdots \\ R_Y(n) \end{bmatrix} \qquad \boldsymbol{\theta}_n = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

$$\mathbf{R}_n = \begin{bmatrix} R_Y(0) & \dots & R_Y(-n+1) \\ & \ddots & R_Y(0) \end{bmatrix}$$

Modelado AR: técnica de la autocovarianza

Las ecuaciones de Y-W quedan reformuladas del siguiente modo:

$$\begin{cases} \begin{bmatrix} R_Y(0) & \mathbf{r}_n^* \end{bmatrix} \begin{bmatrix} 1 \\ \boldsymbol{\theta}_n \end{bmatrix} = \sigma^2 \\ \mathbf{r}_n + \mathbf{R}_n \boldsymbol{\theta}_n = \mathbf{0} \end{cases}$$

donde las incógnitas son $\boldsymbol{\theta}_n$ y σ^2.

Modelado AR: técnica de la autocovarianza

Si conocemos la función de autocorrelación $R_Y(k)$, entonces podemos resolver el sistema anterior

$$\boldsymbol{\theta}_n = -\mathbf{R}_n^{-1}\mathbf{r}_n \qquad \sigma^2 = R_Y(0) - \mathbf{r}_n^*\mathbf{R}_n^{-1}\mathbf{r}_n$$

Por lo general, $R_Y(k)$ tiene que ser estimada a partir de una realización del proceso de salida

$$y(1), \cdots y(N) \quad , \quad N \gg n.$$

Modelado AR: técnica de la autocorrelación

Técnica de la autocorrelación

- Dado el orden del modelo n
- Se estima $\hat{R}_Y(k), k = 0 \cdots n$ utilizando el estimador sesgado
- A partir de estos valores, se obtienen la matriz $\hat{\mathbf{R}}_n$ y el vector $\hat{\mathbf{r}}_n$.
- Se estiman los coeficientes del AR y la varianza de la entrada

$$\hat{\boldsymbol{\theta}}_n = -\hat{\mathbf{R}}_n^{-1}\hat{\mathbf{r}}_n \qquad \sigma^2 = \hat{R}_Y(0) - \hat{\mathbf{r}}_n^*\hat{\mathbf{R}}_n^{-1}\hat{\mathbf{r}}_n$$

Modelado AR: solución usando cuadrados mínimos (LS)

Un enfoque alternativo es el siguiente. Para un par de índices N_1, N_2, buscamos los coeficientes $a_i, i = 1, \cdots n$ que minimizan la siguiente función de costo:

$$\sum_{k=N_1}^{N_2} \left| y(k) + \sum_{i=1}^{n} a_i y(k-i) \right|^2 \qquad 1 \leq N_1 \leq N_2 \leq N.$$

Para resolver este problema, no es necesario estimar $\hat{R}_(k)$ primero. Pero hay que resolver un problema de optimización.

Modelado AR: solución usando cuadrados mínimos (LS)

Por qué utilizar la función de costo anterior?

Sea $X(k)$ el proceso de entrada de media nula y descorrelacionado en el tiempo, y $x(k)$ la realización asociada con la realización $y(k)$. Luego,

$$y(k) + \sum_{i=1}^{n} a_i y(k-i) = x(k)$$

- $\displaystyle\sum_{k=N_1}^{N_2} \left| y(k) + \sum_{i=1}^{n} a_i y(k-i) \right|^2$ es un estimador de $\mathbb{V}(X(k))$.
- Si hallamos el modelo que minimiza la varianza de la entrada, estamos atribuyendo todas las variaciones de $Y(k)$ al sistema.

$$\min \sum_{k=N_1}^{N_2} \left| y(k) + \sum_{i=1}^{n} a_i y(k-i) \right|^2$$

- Definimos el siguiente vector y la matriz:

$$\boldsymbol{y} = \begin{bmatrix} y(N_1) \\ y(N_1+1) \\ \vdots \\ y(N_2) \end{bmatrix} \quad \mathbf{Y} = \begin{bmatrix} y(N_1-1) & y(N_1-2) & \cdots & y(N_1-n) \\ y(N_1) & \ddots & \cdots & y(N_1-n+1) \\ & \vdots & & \\ y(N_2-1) & y(N_2-2) & \cdots & y(N_2-n) \end{bmatrix}$$

- $y(k) + \sum_{i=1}^{n} a_i y(k-i) = (\boldsymbol{y} + \mathbf{Y}\boldsymbol{\theta})_k$
- La solución por mínimos cuadrados busca minimizar la norma

$$||\boldsymbol{y} + \mathbf{Y}\boldsymbol{\theta}||^2$$

Modelado AR: solución usando cuadrados mínimos (LS)

Estimador LS

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

$$\hat{\theta}_{ls} = -(\mathbf{Y}^*\mathbf{Y})^{-1}(\mathbf{Y}^*\boldsymbol{y}).$$

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

- Sea $\mathbf{Y} \in \mathbb{C}^{p \times q}$ (en este caso, $p = N_2 - N_1 + 1$, y $q = n$).
- Sabemos que $\forall \boldsymbol{y} \in \mathbb{C}^p$, existen $\hat{\boldsymbol{y}} \in col(\mathbf{Y})$ y $\boldsymbol{y}_0 \in ker(\mathbf{Y}^*)$ tal que

$$\boldsymbol{y} = \hat{\boldsymbol{y}} + \boldsymbol{y}_0 \quad , \quad \hat{\boldsymbol{y}} \perp \boldsymbol{y}_0.$$

- $\hat{\boldsymbol{y}}$ es el vector de $col(\mathbf{Y})$ que mejor aproxima $\boldsymbol{y}$.
- Idea: Como $\hat{\boldsymbol{y}} \in col(\mathbf{Y})$, elijo $\boldsymbol{\theta} \in \mathbb{C}^q$ tal que $\mathbf{Y}\boldsymbol{\theta} = -\hat{\boldsymbol{y}}$.
- Por definición de núcleo de una matriz

$$\mathbf{Y}^*\boldsymbol{y} = \mathbf{Y}^*\left(\hat{\boldsymbol{y}} + \boldsymbol{y}_0\right) = \mathbf{Y}^*\hat{\boldsymbol{y}}$$

- Luego,

$$\mathbf{Y}^*\boldsymbol{y} = -\mathbf{Y}^*\mathbf{Y}\boldsymbol{\theta}$$

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

$$\boldsymbol{\theta}_{ls} = -\left[\mathbf{Y}^*\mathbf{Y}\right]^{-1}\mathbf{Y}^*\boldsymbol{y}$$

Modelado AR: comparación técnicas

Usando la autocorrelación

$$\hat{\theta}_n = -\hat{\mathbf{R}}_n^{-1}\hat{\mathbf{r}}_n$$

Usando mínimos cuadrados

$$\hat{\theta}_{ls} = -(\mathbf{Y}^*\mathbf{Y})^{-1}(\mathbf{Y}^*\boldsymbol{y}).$$

- Por construcción, la matriz $\hat{\mathbf{R}}_n$ es definida positiva.
- La matriz de LS $\mathbf{Y}^*\mathbf{Y}$ no es necesariamente positiva.
- El modelo obtenido por LS puede ser inestable (en ese caso, un modelo estable puede obtenerse reflejando los polos inestables).
- Para N grande ambas técnicas son parecidas porque los estimadores son más parecidos.

Modelado ARMA

- El modelo ARMA es más general. Tiene polos y ceros
- La desventaja es que la dependecia de los datos y los parámetros es no lineal, lo que complica la solución del mismo.
- Sea $H(z)$ la transferencia del modelo a identificar (se asume estable):

$$H(z) = \frac{B(z)}{A(z)} = \frac{b_0 + b_1 z^{-1} + b_m z^{-m}}{1 + a_1 z^{-1} + a_n z^{-n}}$$

- 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:

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

- 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:

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

  Luego puede tratarse el problema de predecir la muestra $y(k)$ en funcion de las K muestras previas.
- Pueden obtenerse la respuesta v usando las técnicas Y-W o LS vistas antes.
- Con esto se estima la entrada del ARMA: $\{\hat{x}[K+1], \ldots, \hat{x}[N]\}$
- A continuación se utiliza esto para identificar el modelo.

Modelado ARMA: técnica LS en dos etapas (III)

- Se definen:

$$\mathbf{w} \triangleq \left[\begin{array}{ccc|ccc} a_1 & \dots & a_n & b_1 & \dots & b_m \end{array} \right]^T$$

y

$$\mathbf{u}(k) = \left[\begin{array}{ccc|ccc} y[k-1] & \dots & y[k-n] & -x[k-1] & \dots & -x[k-m] \end{array} \right]^T$$

y se escribe la ecuación en diferencias del ARMA como:

$$y(k) + \mathbf{w}^T \mathbf{u}(k) = x(k)$$

- El objetivo es estimar **w**.
- La entrada x fue estimada para $k \geq K$.
- Por lo tanto, definiendo:

$$L = K + m$$

podemos escribir las ecuaciones en forma matricial para:

$$L + 1 \leq k \leq N.$$

Modelado ARMA: técnica LS en dos etapas (IV)

- Matricialmente:

$$\mathbf{Z} \triangleq \begin{bmatrix} y[L] & \dots & y[L-n+1] & -\hat{x}[-L] & \dots & -\hat{x}[L-m+1] \\ y[L+1] & \dots & y[L-n+2] & -\hat{x}[-L+1] & \dots & -\hat{x}[L-m+2] \\ \vdots & & \vdots & \vdots & & \vdots \\ y[N-1] & \dots & y[N-n] & -\hat{x}[N-1] & \dots & -x[N-m] \end{bmatrix}$$

$$\mathbf{z} = \begin{bmatrix} y[L+1] & y[L+2] & \dots & y[N] \end{bmatrix}^T$$

$$\hat{\boldsymbol{x}} = \begin{bmatrix} \hat{x}[L+1] & \hat{x}[L+2] & \dots & \hat{x}[N] \end{bmatrix}^T$$

el sistema queda:

$$\mathbf{z} + \mathbf{Z}\mathbf{w} \approx \hat{\boldsymbol{x}}$$

La solución que mejor aproxima a $\hat{\boldsymbol{x}}$ por LS es:

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

Modelado ARMA: técnica LS en dos etapas (V)

- Por último, se estima σ_x^2 como:

$$\hat{\sigma}_x^2 = \frac{1}{N-L}||\mathbf{z} + \mathbf{Z}\hat{\mathbf{w}}||^2$$

donde el vector **w** es el del estimador LS:

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

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.
- Por otro lado, la dificultad mayor dificultad es cuando los ceros están cerca de la circunferencia unidad, pues K debe ser grande en ese caso.