Tema 2: Fundamentos del problema de estimación

1. Formulación del problema de estimación

2. Solución del problema de estimación

3. Estimación lineal mínimo cuadrática

En la mayor parte de las situaciones reales, las perturbaciones que afectan a un sistema dinámico son aleatorias, por lo que tanto el estado del sistema como las observaciones del mismo tienen también carácter aleatorio. Estos sistemas se denominan Sistemas Estocásticos.

Como en cualquier sistema dinámico, el punto de partida para explicar el comportamiento de un sistema estocástico es conocer el modelo matemático que lo rige, así como las condiciones iniciales que posee al comienzo del estudio. La variable de interés vendrá representada ahora por un vector aleatorio, el vector estado, y dispondremos de una serie medidas u observaciones relacionadas con el mismo, que estarán perturbadas también aleatoriamente. Uno de los principales problemas que se plantea en el análisis de estos sistemas es la estimación de la trayectoria del estado a partir de las observaciones disponibles en cada instante de tiempo. Centrándonos en sistemas discretos se trata, por tanto, de estimar un vector aleatorio (el estado en cada instante de tiempo) a partir de un conjunto finito de vectores relacionados con el mismo (las observaciones).

1 Formulación del problema de estimación

Consideremos un sistema cuyo estado se manifiesta a través del tiempo mediante un proceso estocástico n-dimensional discreto, $\{x(k)\,;\ k \in I\}$, definido sobre un espacio probabilístico $(\Omega, \mathcal{A}, P)$, donde I es un conjunto de índices discreto, finito o infinito: $I = \{0, \ldots, N\}$ o $I = \{0, 1, \ldots\}$.

Supongamos que se está interesado en conocer el valor de $x(k)$ para algún k fijo, pero $x(k)$ no es directamente accesible para el investigador, sino que sólo se dispone de una serie de medidas u observaciones $z(0), \ldots, z(j)$, relacionadas de algún modo con $x(k)$, y se quiere utilizar estas observaciones para estimar el valor de $x(k)$. Supondremos que estas medidas, $\{z(i)\,;\ i = 0, \ldots, j\}$, son vectores aleatorios m-dimensionales definidos también sobre $(\Omega, \mathcal{A}, P)$.

Puesto que sólo disponemos de las medidas $z(0), \ldots, z(j)$ para estimar $x(k)$, un estimador de $x(k)$ basado en dichas medidas, que notaremos mediante $\widehat{x}(k/j)$, será una función vector valuada n-dimensional Φ_k de las mismas; esto es,

$$\widehat{x}(k/j) = \Phi_k\left[z(i)\,;\ i = 0, \ldots, j\right].$$

Por consiguiente, el problema se reduce a determinar Φ_k de forma que el estimador sea óptimo en algún sentido. Si $j < k$, el problema se denomina de *predicción*; si $j = k$, de *filtrado*, y si $j > k$, de *suavizamiento*.

Todo problema de estimación debe resolverse en base a algún criterio que permita valorar la bondad de los estimadores y, para ello, se considera el error de estimación, definido como la diferencia entre el verdadero valor de la variable a estimar y el estimador,

$$\widetilde{x}(k/j) = x(k) - \widehat{x}(k/j).$$

En el caso ideal en que $\widetilde{x}(k/j) = 0$, el estimador sería óptimo, pero, generalmente, $\widetilde{x}(k/j) \neq 0$, por lo que debemos asignar una penalización o medida del error que se comete al aproximar $x(k)$ por $\widehat{x}(k/j)$. Con este fin, se considera una función de pérdida o penalización, $L = L\left[\widetilde{x}(k/j)\right]$, con las siguientes propiedades:

1. L es una función real de n variables.
2. $L(0) = 0$.
3. $L\left[\widetilde{x}^b(k/j)\right] \geq L\left[\widetilde{x}^a(k/j)\right]$, siempre que $\rho\left[\widetilde{x}^b(k/j)\right] \geq \rho\left[\widetilde{x}^a(k/j)\right]$, siendo ρ una función real de n variables, no negativa y convexa.
4. $L\left[\widetilde{x}(k/j)\right] = L\left[-\widetilde{x}(k/j)\right].$

Estas propiedades garantizan que las pérdidas asignadas a distintos errores de estimación son coherentes, y toda función que las verifique se denomina *función de pérdida admisible.*

Debido al carácter aleatorio de $x\,(k)$ y $\widehat{x}\,(k/j)$, se sigue que $\widetilde{x}\,(k/j)$ es también un vector aleatorio y, por tanto, $L\,[\widetilde{x}\,(k/j)]$ será una variable aleatoria; una medida usual de la pérdida global es la pérdida media o esperada,

$$J\,[\widetilde{x}\,(k/j)] = E\Big[L\,[\widetilde{x}\,(k/j)]\Big],$$

que es una función no decreciente de la pérdida. De este modo, el estimador óptimo bajo dicha función de pérdida es aquel que minimiza la pérdida esperada del error de estimación.

El problema de estimación formulado puede resumirse como sigue:

Dadas las medidas $z\,(0)\,,\ldots,z\,(j)$ relacionadas con $x(k)$, se trata de encontrar un estimador, $\widehat{x}\,(k/j) = \Phi_k\,[z\,(i)\,;\; i=0,\ldots,j]$, que minimice la pérdida media $J\,[\widetilde{x}\,(k/j)] = E\Big[L\,[\widetilde{x}\,(k/j)]\Big]$, donde L es una función de pérdida admisible.

Notemos que, como consecuencia de las propiedades de la esperanza condicionada,

$$E\Big[L\,[\widetilde{x}\,(k/j)]\Big] = E\Big[E\Big[L\,[\widetilde{x}\,(k/j)]\,/z\,(0)\,,\ldots,z\,(j)\Big]\Big],$$

y, puesto que la esperanza es un operador no decreciente, de esta expresión se sigue que minimizar $E\Big[L\,[\widetilde{x}\,(k/j)]\Big]$ equivale a minimizar $E\,[L\,[\widetilde{x}\,(k/j)]\,/z\,(0)\,,\ldots,z\,(j)]$. Por tanto, el problema puede reformularse considerando como medida global de la pérdida asociada a un estimador, la pérdida media condicional a las observaciones,

$$\overline{J}\,[\widetilde{x}\,(k/j)] = E\Big[L\,[\widetilde{x}\,(k/j)]\,/z\,(0)\,,\ldots,z\,(j)\Big].$$

2 Solución del problema de estimación

Bajo determinadas condiciones sobre la distribución condicionada de $x\,(k)$ dadas las observaciones $\{z\,(0)\,,\ldots,z\,(j)\}$ o sobre la función de pérdida, la solución teórica del problema de estimación formulado en el apartado anterior puede ser caracterizada de forma simple, como se establece en los siguientes resultados.

Teorema 1 (Sherman). *Si la distribución condicionada del vector $x\,(k)$ dadas las observaciones $\{z\,(0)\,,\ldots,z\,(j)\}$, es simétrica respecto de su media y la función de distribución es convexa para todo valor menor o igual que la media, entonces, para cualquier función de pérdida admisible, el estimador óptimo es*

$$\widehat{x}\,(k/j) = E\,[x\,(k)\,/z\,(0)\,,\ldots,z\,(j)]\,.$$

- Una consecuencia inmediata que se deduce del teorema anterior es que, si los procesos $\{x(k)\,;\; k \in I\}$ y $\{z(i)\,;\; i \in I\}$ verifican las hipótesis de este teorema y son independientes, entonces el estimador óptimo, para cualesquiera k y j, es

$$\widehat{x}(k/j) = E[x(k)].$$

- Si se reemplaza la hipótesis de convexidad de la función de distribución por la de que L sea convexa, el estimador óptimo sigue siendo la esperanza condicionada.

- Si la distribución condicionada de $x(k)$, dadas las observaciones $\{z(0), \ldots, z(j)\}$, es gaussiana, se verifican las hipótesis del teorema y podemos enunciar el siguiente resultado.

  Corolario (Doob). *Si $\{x(k)\,;\; k \in I\}$ y $\{z(i)\,;\; i = 0, \ldots, j\}$ son conjuntamente gaussianos, entonces, para cualquier función de pérdida admisible y cualquier k, el estimador óptimo de $x(k)$ es*

$$\widehat{x}(k/j) = E\left[x(k)/z(0), \ldots, z(j)\right].$$

  Este resultado es también cierto aunque la función de pérdida no sea simétrica, es decir, la cuarta propiedad de las funciones de pérdida no es necesaria.

- Como consecuencia del corolario anterior, y teniendo en cuenta la expresión de la esperanza condicionada bajo hipótesis de normalidad, se tiene que, en el caso gaussiano, el estimador óptimo es lineal; concretamente, en el caso de gaussianidad se tiene:

$$\widehat{x}(k/j) = E[x(k)] + P_{x(k)Z_j} P_{Z_j Z_j}^{-1} \Big(Z_j - E[Z_j] \Big),$$

  donde

$$\begin{aligned} Z_j &= \left(z^T(0), \ldots, z^T(j)\right)^T \\ P_{x(k)Z_j} &= E\Big[\Big(x(k) - E[x(k)]\Big)\Big(Z_j - E[Z_j]\Big)^T\Big] \\ P_{Z_j Z_j} &= E\Big[\Big(Z_j - E[Z_j]\Big)\Big(Z_j - E[Z_j]\Big)^T\Big], \end{aligned}$$

  y el estimador posee las siguientes propiedades:

  1. $\widehat{x}(k/j)$ es lineal, es decir, es combinación lineal de las observaciones disponibles.
  2. $\widehat{x}(k/j)$ y $\widetilde{x}(k/j)$ son vectores n-dimensionales gaussianos.

3. $\widetilde{x}(k/j)$ es independiente de cualquier combinación lineal de las observaciones disponibles; en particular, $\widetilde{x}(k/j)$ es independiente de $\widehat{x}(k/j)$, y por tanto,
$$E\left[\widetilde{x}(k/j)\,\widehat{x}^T(k/j)\right] = 0.$$
4. $\widehat{x}(k/j)$ es el único estimador óptimo.

El corolario de Doob garantiza que, para procesos gaussianos, el estimador óptimo bajo cualquier función de pérdida admisible es la esperanza condicionada. En lo que sigue, vamos a centrarnos en una función de pérdida específica, pero ganaremos generalidad en la clase de procesos estocásticos que pueden considerarse (procesos de segundo orden), pues, para dicha función de pérdida, el estimador óptimo sigue siendo la esperanza condicionada. Concretamente, vamos a considerar una *función de pérdida cuadrática* (admisible):
$$L[\widetilde{x}(k/j)] = \widetilde{x}^T(k/j)\,\widetilde{x}(k/j),$$
con lo que la pérdida esperada es
$$J[\widetilde{x}(k/j)] = E\left[\widetilde{x}^T(k/j)\,\widetilde{x}(k/j)\right].$$
El estimador que minimiza $J[\widetilde{x}(k/j)]$, se denomina *estimador de mínimos cuadrados* o de *menor error cuadrático medio.*

Para este tipo de funciones de pérdida se tiene el siguiente resultado.

Teorema 2 (Doob). *Si la función de pérdida es cuadrática, entonces el estimador óptimo es*
$$\widehat{x}(k/j) = E[x(k)/z(0), \ldots, z(j)].$$

Comparando el Teorema 1 y el Teorema 2, es obvio que el segundo, a pesar de ser más restrictivo respecto a la clase de funciones de pérdida, es más general que el primero en lo que a la clase de procesos estocásticos considerados se refiere. Ambos teoremas son complementarios en el sentido de que, si un problema satisface las hipótesis de Teorema 1, pero no las del Teorema 2, o viceversa, el estimador óptimo es la esperanza condicionada, en cualquiera de los casos.

3 Estimación lineal mínimo cuadrática

En muchas ocasiones, el cálculo del estimador óptimo no es simple, y el problema de estimación debe reconducirse a la búsqueda del óptimo en una subclase de estimadores.

Se ha indicado en la sección anterior que, en el caso gaussiano, el estimador óptimo bajo cualquier función de pérdida admisible, y en particular bajo función de pérdida cuadrática, es la esperanza condicionada y, por tanto, es una función lineal de las observaciones. Así, en el caso gaussiano, el estimador óptimo coincide con el estimador lineal de menor error cuadrático medio y la restricción a la clase de estimadores lineales no supone pérdida de generalidad en caso de gaussianidad. En esta sección presentamos una interpretación geométrica del estimador lineal de menor error cuadrático medio, que se obtiene mediante el Teorema de la Proyección Ortogonal.

Consideremos las observaciones $z(0), \ldots, z(j)$, y denotemos por $\mathcal{Y}(j)$ el subespacio lineal del espacio de Hilbert $\mathcal{L}^2_{\mathbb{R}^n}(\Omega, A, P)$[1] generado por dichos vectores. En virtud del Teorema de la Proyección Ortogonal, todo vector $x \in \mathcal{L}^2_{\mathbb{R}^n}(\Omega, A, P)$ se puede descomponer de forma única como suma de dos vectores, un vector $\overline{x} \in \mathcal{Y}(j)$, y un vector $\widetilde{x}$ ortogonal a $\mathcal{Y}(j)$ (es decir, ortogonal a cualquier vector de $\mathcal{Y}(j)$). El vector $\overline{x}$ se denomina *proyección ortogonal* de x sobre $\mathcal{Y}(j)$ y es aquel vector de $\mathcal{Y}(j)$ (es decir, aquella función lineal de $z(0), \ldots, z(j)$) que minimiza $E\left[(x-\bar{w})^T(x-\bar{w})\right]$ en $\mathcal{Y}(j)$.

En efecto, si $\bar{w}$ es otro vector de $\mathcal{Y}(j)$, se tiene que

$$E\left[(x-\bar{w})^T(x-\bar{w})\right] = E\left[[(x-\overline{x}) + (\overline{x}-\bar{w})]^T[(x-\overline{x}) +]\overline{x}-\bar{w})]\right).$$

Puesto que $\widetilde{x} = x - \overline{x}$ es ortogonal a $\mathcal{Y}(j)$, en particular es ortogonal a $(\overline{x}-\bar{w})$ y, por tanto,

$$\begin{aligned} E\left[(x-\bar{w})^T(x-\bar{w})\right] &= E\left[(x-\overline{x})^T(x-\overline{x})\right] + E\left[(\overline{x}-\bar{w})^T(\overline{x}-\bar{w})\right] \\ &\geq E\left[(x-\overline{x})^T(x-\overline{x})\right]. \end{aligned}$$

Luego $\overline{x}$ minimiza la pérdida cuadrática media, $E\left[(x-\bar{w})^T(x-\bar{w})\right]$, en la clase de funciones lineales de $z(0), \ldots, z(j)$.

Por tanto, la proyección ortogonal, $\overline{x}$, es el estimador lineal de mínimos cuadrados de x basado en $\{z(0), \ldots, z(j)\}$ y $\widetilde{x}$ es el error de estimación.

Para el problema de estimación formulado anteriormente, estos resultados se resumen en el siguiente teorema.

Teorema 3. *Sean $\{x(k);\ k \in I\}$ y $\{z(k);\ k \in I\}$ procesos estocásticos con media cero. Supongamos que observamos $z(0), \ldots, z(j)$ y que se da alguna de las condiciones siguientes:*

[1] $\mathcal{L}^2_{\mathbb{R}^n}(\Omega, A, P)$ denota el conjunto de clases de equivalencia (inducidas por la relación de igualdad casi segura) de variables aleatorias n-dimensionales, definidas sobre el espacio probabilístico (Ω, A, P), con momento de segundo orden finito. $\mathcal{L}^2_{\mathbb{R}^n}(\Omega, A, P)$ es un espacio de Hilbert con el producto escalar $< Z, W >= E\left[Z^T W\right]$.

(i) Los procesos $\{x(k);\ k \in I\}$ *y* $\{z(k);\ k \in I\}$ *son conjuntamente gaussianos.*

(ii) La función de pérdida considerada es cuadrática y nos restringimos a la clase de funciones lineales de las observaciones.

Entonces el estimador óptimo de $x(k)$ *dadas* $z(0), \ldots, z(j)$ *viene dado por la proyección ortogonal de* $x(k)$ *sobre* $\mathcal{Y}(j)$.

A partir de este resultado, la obtención del estimador lineal de menor error cuadrático medio puede realizarse haciendo uso del siguiente resultado.

Lema de Proyecciones Ortogonales. *El estimador lineal de menor error cuadrático medio,* $\widehat{x}(k/j)$, *es la única combinación lineal de* $z(0), \ldots, z(j)$ *tal que* $x(k) - \widehat{x}(k/j)$ *es ortogonal a cualquier otra función lineal de* $z(0), \ldots, z(j)$,

$$E\left[(x(k) - \widehat{x}(k/j))^T\ \bar{w}\right] = 0, \quad \forall\ \bar{w} \in \mathcal{Y}(j).$$

Ya que $E\left[x^T z\right] = trazaE\left[x z^T\right]$, la condición del Lema de Proyecciones Ortogonales se puede expresar de forma equivalente como

$$E\left[(x(k) - \widehat{x}(k/j))\ \bar{w}^T\right] = 0, \quad \forall\ \bar{w} \in \mathcal{Y}(j),$$

lo que indica que el estimador lineal de menor error cuadrático medio es la única combinación de $z(0), \ldots, z(j)$ tal que *el error de estimación es incorrelado con cualquier función lineal de* $z(0), \ldots, z(j)$.

Entonces, notando que la incorrelación con los elementos de $\mathcal{Y}(j)$ equivale a la incorrelación con los vectores que lo generan, $z(0), \ldots, z(j)$, la búsqueda del estimador lineal de menor error cuadrático medio se reduce a encontrar la solución (en $\mathcal{Y}(j)$) de la siguiente ecuación, denominada *ecuación de Wiener-Hopf*:

$$E\left[x(k) z^T(i)\right] = E\left[\widehat{x}(k/j)\, z^T(i)\right], \quad \forall i = 0, \ldots, j.$$

A partir de este ecuación es inmediato deducir que

$$\widehat{x}(k/j) = \widehat{x}(k/j-1) + K(k,j)\left[z(j) - \widehat{z}(j/j-1)\right],$$

siendo $\widehat{z}(j/j-1)$ el estimador lineal de menor error cuadrático medio de $z(j)$ basado en $z(0), \ldots, z(j-1)$, y $K(k,j)$ un factor a determinar en cada situación. Esta ecuación constituye el punto de partida para los algoritmos recursivos de estimación lineal de menor error cuadrático medio que se obtienen en los siguientes temas.