

# **Análisis de Intervención Automática de Series Temporales**

## **Estimación bayesiana de estructuras deterministas en series temporales**

**Documentación técnica**

***Víctor de Buen Remiro***

vbuen@terra.es



## Contenido

1.	Esquema del algoritmo	1
2.	Condiciones de partida	3
3.	El conjunto de estructuras básicas	4
4.	El conjunto de observaciones de referencia	5
5.	Asignación de probabilidades a un conjunto de hipótesis	6
6.	Comparación de modelos con diferente número de parámetros	9
7.	Estimación lineal recursiva	12

## Resumen

El objeto de la presente nota es definir un método de Análisis de Intervención Automática (AIA) para series temporales. Resulta muy frecuente al observar datos de cualquier naturaleza (física, social, económica, etc ) que existan datos anómalos, valores fuera de lo común, que suelen ser debidos a errores de medida o de manipulación, comportamientos altamente no lineales, discontinuidades o cualquier tipo de causas endógenas o exógenas, conocidas o por determinar.

A la hora de analizar una serie temporal es muy importante separar este tipo de fenómenos del comportamiento normal, bien para intentar comprender las leyes subyacentes con vistas a la previsión de situaciones similares en el futuro, bien para filtrar la serie de los efectos que la distorsionan, si es que se sospecha de que se trata de errores, situaciones irrepetibles o impredecibles, o simplemente no se dispone de recursos para el análisis completo, unas veces por escasez muestral, y otras por razones más mundanas como su alto coste económico.

El propósito de este AIA no es eliminar esas anomalías sino identificarlas para poder hacer después una buena identificación y estimación del modelo ARIMA con Función de Transferencia.



Entidad	Análisis de Intervención Automática de Series Temporales
Asunto	Estimación bayesiana de estructuras deterministas en series temporales
Archivo	c:\home\bayes\aiia\aiia.doc
Edición	2002-12-31 13:06
Claves	Análisis de Intervención Automática, AIA, ARIMA, Series temporales, Estimación bayesiana,
Distribución	General



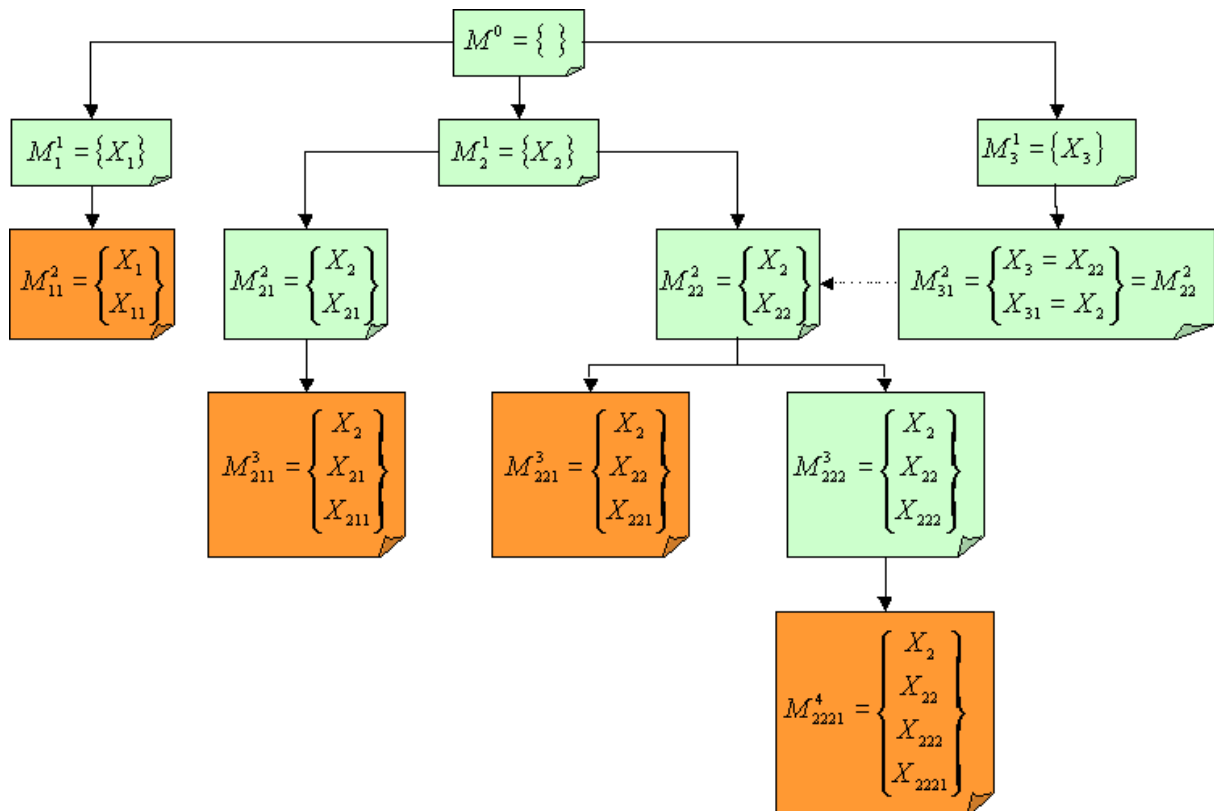
## 1. Esquema del algoritmo

El esquema del método AIA aquí expuesto es un árbol de decisión bajo un enfoque bayesiano. Una vez preparada la serie conveniente se busca el dato anómalo de mayor valor absoluto y se calcula la probabilidad de cada una de las estructuras de anomalías compatibles con él de entre un conjunto dado previamente, en términos del modo en que cada una explica mejor o peor dicho valor. Para cada opción con un mínimo de probabilidad se aplica recursivamente el método descrito.

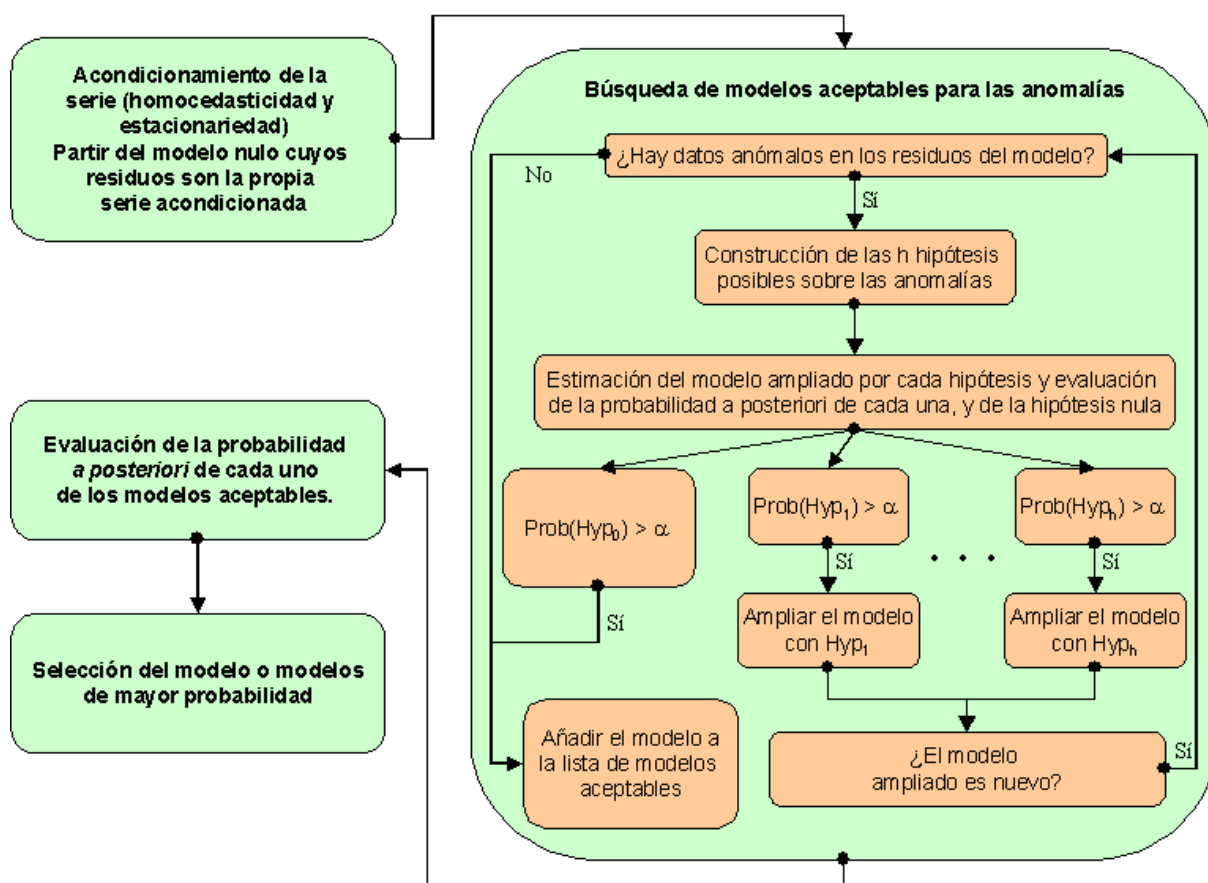
Para evitar ramificaciones innecesarias se debe observar si el modelo ya ha sido evaluado (lógicamente con las variables en otro orden). Del mismo modo, si existe un modelo con menos variables que da menos error se debe abandonar la rama.

Se debe empezar a evaluar siempre por la hipótesis más probable y continuar mientras la probabilidad sea aceptable. También se puede restringir el número de variables, el de ramas de un nodo o el tiempo de ejecución para evitar situaciones de ramificación excesiva.

He aquí un ejemplo gráfico del árbol de decisión:



Al final se obtiene un conjunto de modelos con diferente número de variables dependiendo del nivel del árbol, de entre las cuales hay que elegir el de mayor probabilidad *a posteriori*, o bien introducir algún criterio de información como el de Schwarz, aunque ésta solución no es tan plausible.



Otra forma de disminuir las ramificaciones y el tiempo de cálculo es el método *divide and conquer* que consiste en dividir la serie en tramos razonablemente independientes entre sí y aplicar el algoritmo a cada uno de ellos. Luego se seleccionan las anomalías detectadas en cada trozo que efectivamente no influyen en los demás. Por último se aplica el algoritmo a la serie completa pero partiendo del modelo unión resultante reestimado.



## 2. Condiciones de partida

Es necesario partir de una serie homocedástica y estacionaria que garantice varianza finita y constante y media constante, para poder trabajar bajo el supuesto de normalidad. No supone ninguna restricción exigir que la media sea nula.

Lo primero se consigue mediante transformaciones del tipo Box-Cox de la forma

$$[3.1] \quad T(z; a, b) = \begin{cases} (z - b)^a & \forall a > 0 \wedge z - b \geq 0 \\ \ln(z - b) & \forall a = 0 \wedge z - b > 0 \\ (z - b)^a & \forall a < 0 \wedge z - b > 0 \end{cases}$$

Si la serie es siempre positiva es habitual tomar el logaritmo que ofrece la ventaja de que los coeficientes de la estimación del modelo toman el significado aproximado de porcentaje de cambio.

Para lo segundo hay que encontrar una estructura de diferencias suficiente ya que no hay demasiado problema por sobre-diferenciar ligeramente la serie, pues no se debe olvidar que sólo queremos identificar las estructuras de anomalías. Precisamente, una vez identificadas se utiliza la serie filtrada sin diferenciar para buscar la estructura de diferencias más adecuada.

Teniendo en cuenta la o las posibles periodicidades buscaremos polinomios de la forma

$$[3.2] \quad \nabla(B) = (1 - B)^{d_0} (1 - B^{s_1})^{d_1} \dots$$

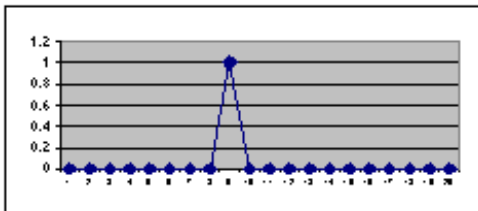
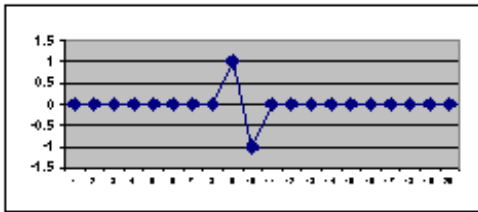
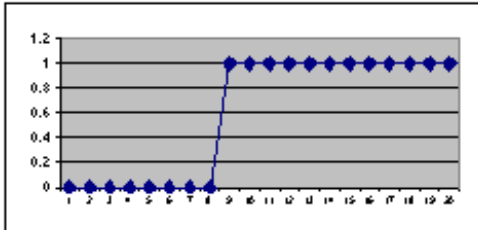
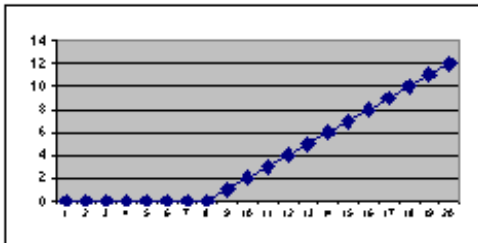
Llamando  $z_t^o$  a los datos originales se construye de esta manera la serie de partida

$$[3.3] \quad z_t = \nabla(B)T(z_t^o; a, b)$$

Basta con elegir la estructura que minimiza la varianza utilizando un estimador robusto de la misma como el que se describe en el apartado 3.

### 3. El conjunto de estructuras básicas

Una de las tareas más importantes es establecer las estructuras básicas o hipótesis a contrastar. Como es natural, estas estructuras dependerán de la forma que presenten los datos anómalos, si se trata de datos aislados o presentan cierta secuencia, si conforman ciclos o si tienen simetría. En definitiva, se trata de encontrar una base del espacio vectorial de todos los tipos anomalías que se presentan en la serie. Usualmente se toman las siguientes estructuras

Nombre	Función de respuesta	Forma
<p>Pulso</p> <p><math>Pulse(t)</math></p>	$\frac{P(B)}{Q(B)} = \frac{1}{1}$	
<p>Compensación</p> <p><math>Compens(t)</math></p>	$\frac{P(B)}{Q(B)} = \frac{1-B}{1}$	
<p>Escalón</p> <p><math>Step(t)</math></p>	$\frac{P(B)}{Q(B)} = \frac{1}{1-B}$	
<p>Tendencia</p> <p><math>Trend(t)</math></p>	$\frac{P(B)}{Q(B)} = \frac{1}{(1-B)^2}$	

## 4. El conjunto de observaciones de referencia

Antes de establecer cuáles son los datos anómalos caben dos posibilidades, o bien suponemos que la serie ya es ruido blanco, o bien le aplicamos un modelo ARMA más o menos sencillo para que la suposición sea más cercana a la realidad. Se considerarán como observaciones de referencia a aquellos datos cuyo valor absoluto sea mayor que cierto número de desviaciones típicas, como por ejemplo 3 ó 4. Cuanto menor sea este número más exhaustivo y al mismo tiempo más lento será el método. Se necesita una estimación de la desviación típica lo bastante robusta como para que no se vea demasiado afectada por dichos valores anormales.

Sean  $Y_1, \dots, Y_N$  las variables aleatorias resultantes de ordenar de menor a mayor  $N$  variables aleatorias  $X_1, \dots, X_N$  con distribuciones  $Normal(0, \sigma)$  independientes. Sea  $\xi$  una v. a. normal estándar independiente de las  $X_1, \dots, X_N$ , y sea  $K$  la variable aleatoria discreta definida por el número de elementos de  $\{X_1, \dots, X_N\}$  menores o iguales que  $\sigma\xi$ , o dicho de otro modo

$$[5.1] \quad K = \begin{cases} 0 & \Leftrightarrow \sigma\xi < Y_1 \\ k & \Leftrightarrow Y_k \leq \sigma\xi < Y_{k+1} \quad \forall k = 1..N \\ N & \Leftrightarrow Y_N \leq \sigma\xi \end{cases}$$

Es inmediato que  $K$  tiene distribución uniforme en  $\{0..N\}$ , luego

$$[5.2] \quad \begin{aligned} P[K \leq k] &= \frac{k+1}{N+1} \Rightarrow P[Y_k \leq \sigma\xi] = \frac{k+1}{N+1} \Rightarrow -\frac{1}{\sigma}Y_k = F_{01}^{-1}\left(\frac{k+1}{N+1}\right) \\ P[K \geq N-k] &= \frac{k+1}{N+1} \Rightarrow P[Y_{N-k} \geq \sigma\xi] = \frac{k+1}{N+1} \Rightarrow \frac{1}{\sigma}Y_{N-k} = F_{01}^{-1}\left(\frac{k+1}{N+1}\right) \end{aligned}$$

donde  $F_{01}^{-1}$  es la inversa de la función de distribución normal estándar. Entonces los valores de la serie  $x_1, x_2, \dots, x_N$  una vez ordenados  $y_1 \leq y_2 \leq \dots \leq y_N$  son una realización del vector de variables aleatorias  $(Y_1, \dots, Y_N)$  y se definen los estimadores de la desviación típica  $\sigma$

$$[5.3] \quad \left. \begin{aligned} \hat{\sigma}_{1,k} &= y_k / F_{01}^{-1}\left(\frac{k+1}{N+1}\right) \\ \hat{\sigma}_{2,k} &= -y_{N-k} / F_{01}^{-1}\left(\frac{k+1}{N+1}\right) \end{aligned} \right\} \hat{\sigma}_k = \frac{\hat{\sigma}_{1,k} + \hat{\sigma}_{2,k}}{2} = \frac{y_{N-k} - y_k}{-2F_{01}^{-1}\left(\frac{k+1}{N+1}\right)}$$

Es conveniente tomar  $k$  de forma que  $\frac{1}{2} - \frac{k+1}{N+1} \approx \frac{1}{3}$  y se suele tomar

$$[5.4] \quad k^{(6)} = \text{Round}[(N+1)/6] - 1 \Rightarrow -F_{01}^{-1}\left(\frac{k+1}{N+1}\right) \approx 0.96742 \Rightarrow \hat{\sigma}_{k^{(6)}} \approx \frac{1}{1.93484}(Y_{N-k^{(6)}} - Y_{k^{(6)}})$$

## 5. Asignación de probabilidades a un conjunto de hipótesis

En primer lugar hay que hacer notar la diferencia entre dato anómalo u observación de referencia de la anomalía y origen de la anomalía. Dado una observación de referencia en el instante  $t$ , para cada función de respuesta  $P(B)/Q(B)$  en concurso puede ocurrir que el origen de la anomalía se encuentre en ese punto, o bien antes o después, en cualquiera de los puntos en los que la serie

$$[6.1] \quad \frac{P(F)\nabla(F)}{Q(F)}Pulse(t)$$

se acerca a su máximo en valor absoluto.

Si partimos de un modelo de  $n-1$  variables al que llamaremos  $M_0$  se establece la hipótesis nula  $H_0$  de que el modelo  $M_0$  es adecuado. Dado un conjunto de estructuras de anomalías definidas cada una por un instante de origen, y una función de respuesta

$$[6.2] \quad \left( t_j, \frac{P_j(B)}{Q_j(B)} \right) \wedge t_j \in \{1 \dots N\}, \forall j = 1 \dots J$$

se establecen las hipótesis  $H_j$  correspondientes a añadir al modelo  $M_0$  cada una de las respectivas variables

$$[6.2] \quad X^{(k)} = \frac{P_j(B)\nabla(B)}{Q_j(B)}Pulse(t_j)$$

De esta forma hallamos las correspondientes sumas de cuadrados de residuos

$$[6.3] \quad S_{A_j}^2 = \sum_{t=1}^N a_{j,t}^2$$

La función de verosimilitud de los residuos condicionados a la hipótesis  $H_j$  y a la desviación típica es

$$[6.4] \quad L(Z | H_j, \sigma) = (2\pi\sigma^2)^{-\frac{1}{2}N} e^{-\frac{1}{2\sigma^2}A_j^T A_j} = (2\pi\sigma^2)^{-\frac{1}{2}N} e^{-\frac{1}{2\sigma^2}S_{A_j}^2}$$

La desviación típica  $\sigma$  es desconocida pero podemos hacer ciertas hipótesis, como por ejemplo que la función de densidad es decreciente. Debido a la gran simplicidad de los cálculos, una opción razonable es tomar como función de densidad  $f(\sigma) \propto \sigma^{-1}$

$$\begin{aligned}
 L(Z | H_j) &= \int_0^\infty \sigma^{-1} L(A | H_j, \sigma) d\sigma \\
 [6.5] \quad L(Z | H_j) &= \int_0^\infty \sigma^{-1} (2\pi)^{-\frac{1}{2}N} \sigma^{-N} e^{-\frac{1}{2\sigma^2} S_{A_j}^2} d\sigma \\
 L(Z | H_j) &= (2\pi)^{-\frac{N}{2}} \int_0^\infty \sigma^{-(N+1)} e^{-\frac{1}{2\sigma^2} S_{A_j}^2} d\sigma
 \end{aligned}$$

que se puede resolver con el cambio de variable

$$\begin{aligned}
 [6.6] \quad u &= \frac{1}{2} \frac{S_{A_j}^2}{\sigma^2} \quad \sigma = \left( \frac{S_{A_j}^2}{2u} \right)^{1/2} = \left( \frac{S_{A_j}^2}{2} \right)^{1/2} u^{-1/2} \\
 du &= -\frac{S_{A_j}^2}{\sigma^3} d\sigma \quad d\sigma = -\frac{1}{2} \left( \frac{S_{A_j}^2}{2} \right)^{1/2} u^{-3/2} du = \left( \frac{S_{A_j}^{2/3}}{2u} \right)^{3/2} du
 \end{aligned}$$

Resultando

$$\begin{aligned}
 L(Z | H_j) &= (2\pi)^{-\frac{N}{2}} \int_\infty^0 \left( \frac{S_{A_j}^2}{2u} \right)^{-\frac{N+1}{2}} e^{-u} \left( -\left( \frac{S_{A_j}^{2/3}}{2u} \right)^{3/2} \right) du \\
 [6.7] \quad L(Z | H_j) &= \frac{1}{2} (2\pi)^{-\frac{N}{2}} \left( \frac{S_{A_j}^2}{2} \right)^{-\frac{N}{2}} \int_0^\infty u^{\frac{N}{2}-1} e^{-u} du \\
 L(Z | H_j) &= \frac{1}{2} (2\pi)^{-N/2} \left( \frac{1}{2} S_{A_j}^2 \right)^{-N/2} \Gamma\left(\frac{N}{2}\right)
 \end{aligned}$$

En muchas ocasiones se utiliza el logaritmo de la función de verosimilitud para comparar las hipótesis puesto que el logaritmo es una función monótona creciente y nos da valores con menos problemas numéricos

$$[6.8] \quad \text{Log} L(Z | H_j) = \text{Log}\left(\frac{1}{2}\right) + \text{Log}\Gamma\left(\frac{N}{2}\right) - \frac{N}{2} \text{Log}(2\pi) - \frac{N}{2} \text{Log}\left(\frac{1}{2} S_{A_j}^2\right)$$

La probabilidad de los datos observados condicionados por nuestra hipótesis es proporcional a la función de verosimilitud y como en nuestro caso  $N$  no varía entre las diferentes hipótesis, a efectos de compararlas, se tiene que

$$[6.9] \quad P(Z | H_i) \propto S_{A_j}^{-N}$$

Obsérvese que dicho valor es el inverso de la raíz cuadrada del determinante de la matriz de covarianzas de los residuos.

Utilizando toda la información *a priori* de la se que disponga, se postulan las probabilidades  $P(H_j)$  con las que se obtienen las probabilidades *a posteriori* por el teorema de Bayes

$$[6.10] \quad P(H_j | Z) = \frac{P(Z | H_j)P(H_j)}{\sum_{k=0}^J P(Z | H_k)P(H_k)} = \frac{S_{A_j}^{-N} P(H_j)}{\sum_{k=0}^J S_{A_k}^{-N} P(H_k)}$$

Para evitar problemas numéricos se puede dividir todo por  $S_Z^{-N}$

$$[6.11] \quad P(H_j | Z) = \frac{(S_{A_j}/S_Z)^{-N} P(H_j)}{\sum_{k=0}^J (S_{A_k}/S_Z)^{-N} P(H_k)}$$

Si no se sabe nada sobre las probabilidades  $P(H_j)$  se suponen uniformes.

Para todas aquellas hipótesis cuya probabilidad supere cierto umbral se repite el proceso de búsqueda, que toma de este modo estructura recursiva arbórea.

## 6. Comparación de modelos con diferente número de parámetros

En el modelo lineal de  $N$  datos y  $n$  variables

$$[7.1] \quad Z = X\alpha + A$$

la estimación máximo verosímil del vector de variables es

$$[7.2] \quad \begin{aligned} X^T Z &= X^T X \hat{\alpha} \\ \hat{\alpha} &= (X^T X)^{-1} X^T Z \end{aligned}$$

La suma de cuadrados de los residuos es

$$[7.3] \quad \hat{A}^T \hat{A} = Z^T Z - \hat{\alpha}^T X^T X \hat{\alpha}$$

puesto que

$$[7.4] \quad \begin{aligned} \hat{A}^T \hat{A} &= (Z - X\hat{\alpha})^T (Z - X\hat{\alpha}) \\ \hat{A}^T \hat{A} &= Z^T Z - 2\hat{\alpha}^T X^T Z + \hat{\alpha}^T X^T X \hat{\alpha} \\ \hat{A}^T \hat{A} &= Z^T Z - 2\hat{\alpha}^T X^T X \hat{\alpha} + \hat{\alpha}^T X^T X \hat{\alpha} \\ \hat{A}^T \hat{A} &= Z^T Z - \hat{\alpha}^T X^T X \hat{\alpha} \end{aligned}$$

Si ponemos nombre a las sumas de cuadrados

$$[7.5] \quad \begin{aligned} S_z^2 &= Z^T Z = \|Z\|_2^2 \\ S_{\hat{A}}^2 &= \hat{A}^T \hat{A} = \|\hat{A}\|_2^2 \\ S_{X\hat{\alpha}}^2 &= (X\hat{\alpha})^T X\hat{\alpha} = \|X\hat{\alpha}\|_2^2 \end{aligned}$$

sus raíces cuadradas forman un triángulo rectángulo.

Para comparar la calidad de modelos con diferente número de parámetros se usa el siguiente hecho. La distribución de los parámetros es una multinormal

$$[7.6] \quad \alpha \approx N(\hat{\alpha}, \Sigma) \wedge \Sigma = \sigma^2 R^{-1} = (\sigma_{i,j})_{i,j=1..n} \wedge R = X^T X$$

Dada la descomposición de Choleski  $R = LL^T$  se puede introducir un cambio de variable en el modelo

$$[7.7] \quad Z = X\alpha + A = XL^{T-1}L^T\alpha + A = X^*\alpha^* + A \wedge X^* = XL^{T-1}; \alpha^* = L^T\alpha$$

de forma que

$$[7.7] \quad \Sigma^* = \sigma^2 (X^{*T} X^*)^{-1} = \sigma^2 (L^{-1} X^T X L^{T-1})^{-1} = \sigma^2 (L^{-1} L L^T L^{T-1})^{-1} = \sigma^2 I$$

y las nuevas variables sean independientes entre sí

$$[7.8] \quad \alpha^* \approx N(\hat{\alpha}^*, \sigma^2 I) \wedge \hat{\alpha}^* = L^T \hat{\alpha}$$

La suma de cuadrados de los residuos estandarizados no cambia pues los residuos del modelo son los mismos

$$[7.6] \quad \frac{S_{\hat{A}}^2}{\sigma^2} = \frac{(N-n)\hat{\sigma}^2}{\sigma^2} \wedge \hat{\sigma}^2 = \frac{S_{\hat{A}}^2}{N-n}$$

tiene una distribución ji-cuadrado con  $\nu = N - n$  grados de libertad y siendo  $\hat{\alpha}^*$  y  $S^2$  independientes.

Luego las variables

$$[7.8] \quad \beta_i = \sqrt{\nu} \frac{(\hat{\alpha}_i^* - \alpha_i^*)/\sigma}{S_{\hat{A}}/\sigma} = \frac{\hat{\alpha}_i^* - \alpha_i^*}{\hat{\sigma}}$$

tienen distribución t de Student con  $\nu$  grados de libertad independientes entre sí.

La condición de irreducibilidad es exigible como medida de calidad de un modelo de AIA. Si se establece la hipótesis  $H$  de que el modelo no es reducible, la probabilidad de  $H$  es el producto de las probabilidades de que una t de Student con  $\nu$  g.l.  $t_\nu$  alcance en módulo cada una de las componentes de  $w$ , condicionadas a que  $\alpha = 0$

$$[7.15] \quad P(H) = \prod_{i=1}^n P\left(t_\nu \leq \left| \frac{\hat{\alpha}_i^*}{\hat{\sigma}} \right| \right)$$

Dado un conjunto de modelos se pueden establecer las hipótesis  $\{H_j\}_{j=1..J}$  respectivas de que cada uno de ellos es no reducible, de modo que como se vio en el apartado 6

$$[7.16] \quad L(Z | H_j) P(H_j) = \left( \frac{1}{2} (2\pi)^{-N/2} \left( \frac{1}{2} S_{A_j}^2 \right)^{-N/2} \Gamma\left(\frac{N}{2}\right) \right) \left( \prod_{i=1}^{n_j} P\left(t_{N-n_j} \leq \left| \frac{\hat{\alpha}_{ji}^*}{\hat{\sigma}} \right| \right) \right)$$



$$[7.17] \quad P(H_j | Z) = \frac{L(Z | H_j)P(H_j)}{\sum_{k=1}^J L(Z | H_k)P(H_k)} = \frac{S_{A_j}^{-N} \prod_{i=1}^{n_j} P(t_{N-n_j} \leq |w_{ji}^0|)}{\sum_{k=1}^J S_{A_k}^{-N} \prod_{i=1}^{n_k} P(t_{N-n_k} \leq |w_{ki}^0|)}$$

Esto permite comparar modelos con distinto número de parámetros, incluido el modelo trivial o de ruido blanco, para el que se establece la hipótesis  $H_\phi$  de que no es reducible, lo que es absolutamente cierto, es decir

$$[7.18] \quad P(H_\phi) = 1$$

Este método cumple el precepto de la parquedad de parámetros pues castiga fuertemente las variables poco significativas o aquellas que introduzcan correlación en el modelo, sin necesidad de parámetros subjetivos de significación.

## 7. Estimación lineal recursiva

En cada iteración del algoritmo del AIA pasamos de un modelo lineal con cierto número de variables a otros modelos con una variable añadida correspondiente a cada una de las hipótesis a comparar. Parece por lo tanto interesante contar con un método de estimación que tenga en cuenta la parte ya estimada, es decir, un método recursivo de estimación lineal. Este método debe incluir el cálculo de la descomposición de Choleski de  $X^T X$  que se usa no sólo para resolver el sistema sino también para calcular la significación de los parámetros.

Dado el modelo lineal expresado en el apartado 5

$$[8.1] \quad Z = X\alpha + A$$

donde

$$[8.2] \quad \begin{aligned} \alpha &= (\alpha_i) \in \mathfrak{R}^{n \times 1} \\ X &= (x_{i,j}) \in \mathfrak{R}^{N \times n} \\ R &= X^T X = (r_{i,j}) \in \mathfrak{R}^{n \times n} \end{aligned}$$

definimos las submatrices

$$[8.3] \quad \begin{aligned} X^{(k)} &= (x_{i,j})_{j=k} \in \mathfrak{R}^{N \times 1} \\ X_{(k)} &= (X_{i,j})_{j \leq k} \in \mathfrak{R}^{N \times k} \\ R_{(k)} &= (r_{i,j})_{i,j \leq k} \in \mathfrak{R}^{k \times k} \end{aligned}$$

Entonces, podemos definir la sucesión de modelos

$$[8.4] \quad Z = X_{(k)}\alpha_{(k)} + A_{(k)}$$

donde

$$[8.5] \quad \alpha_{(k)} = (X_{(k)}^T X_{(k)})^{-1} X_{(k)}^T Z$$

$$[8.6] \quad A_{(k)} = Z - X_{(k)}\alpha_{(k)}$$

Obviamente

$$[8.7] \quad \begin{aligned} X_{(k)} &= \begin{pmatrix} X_{(k-1)} & X^{(k)} \end{pmatrix} \\ R_{(k)} &= X_{(k)}^T X_{(k)} \end{aligned}$$

y por tanto

$$[8.8] \quad R_{(k)} = \begin{pmatrix} R_{(k-1)} & X_{(k-1)}^T X^{(k)} \\ X^{(k)T} X_{(k-1)} & X^{(k)T} X^{(k)} \end{pmatrix}$$

Llamando

$$[8.9] \quad R_{(k,k-1)} = X^{(k)T} X_{(k-1)} = (r_{1,k} \quad r_{2,k} \quad \cdots \quad r_{k-1,k})$$

resulta

$$[8.10] \quad R_{(k)} = \begin{pmatrix} R_{(k-1)} & R_{(k,k-1)}^T \\ R_{(k,k-1)} & r_{k,k} \end{pmatrix}$$

Si tenemos la descomposición de Choleski para la iteración  $j$

$$[8.11] \quad R_{(j)} = L_{(j)} L_{(j)}^T$$

Entonces se cumple la relación recursiva

$$[8.12] \quad L_{(k)} = \begin{pmatrix} L_{(k-1)} & 0 \\ B & l \end{pmatrix}$$

puesto que

$$[8.13] \quad R_{(k)} = L_{(k)} L_{(k)}^T = \begin{pmatrix} L_{(k-1)} L_{(k-1)}^T & L_{(k-1)} B^T \\ B L_{(k-1)}^T & B B^T + l^2 \end{pmatrix} = \begin{pmatrix} R_{(k-1)} & R_{(k,k-1)}^T \\ R_{(k,k-1)} & r_{k,k} \end{pmatrix}$$

Igualando término a término

$$[8.14] \quad \begin{aligned} R_{(k,k-1)}^T &= L_{(k-1)} B^T \\ r_{k,k} &= B B^T + l^2 \Rightarrow l = \sqrt{r_{k,k} - B B^T} \\ |R_{(k)}| &= |L_{(k)}|^2 = |L_{(k-1)}|^2 l^2 \end{aligned}$$