

Reduciendo el número de componentes de una mezcla en modelos dinámicos

por
MAGDALENA CORDERO VALDAVIDA
Centro de Investigaciones Sociológicas

RESUMEN

El filtro de Kalman, como procedimiento de mínimos cuadrados, tiene el problema de la falta de robustez. Varios autores han tratado el tema de la robustificación del mismo mediante el uso de mezclas de distribuciones normales como modelo para el error tanto en la ecuación de observación como en la ecuación de estado. Pero la solución exacta del filtro no es operativa al ser el modelo explosivo ya que el número de componentes de la mezcla crece exponencialmente con t . Este trabajo trata de presentar distintas formas de reducción del número de componentes en la mezcla en cada etapa utilizando diversos criterios como: distancia de Kulback-Leibler, la poda, la unimodalidad y los momentos entre otros, presentándose algunos resultados comparativos.

Palabras clave: Mezclas de distribuciones, colapsar, filtro de Kalman robusto, aproximación de mezclas, observaciones atípicas.

Clasificación AMS: 62M20, 62F15.

1. INTRODUCCIÓN

Es ya un tema clásico, pero no por ello deja de ser atractivo, tratar de robustificar los procedimientos estadísticos. La presencia de valores atípicos en los datos destruye, en la mayoría de los casos, la validez de los resultados y, por tanto, el análisis de los mismos.

Una forma de acercarse al problema se basa en considerar un modelo original para los datos que permita la existencia de esas observaciones, para ello basta con establecer la hipótesis de que las observaciones proceden del modelo principal con un probabilidad α y permitir que exista un porcentaje de las mismas $(1-\alpha)$, que proceden de otra distribución (generalmente de mayor varianza) que también contribuye a la generación de los datos. El nuevo modelo para las observaciones tendrá una distribución del error que será una **mezcla** de distribuciones.

Las implicaciones del uso de mezclas son varias, la primera es que la estadística se ha desarrollado fundamentalmente bajo la hipótesis de normalidad, por lo que deberán desarrollarse nuevos procedimientos que permitan el uso de mezclas y establecer las propiedades de los nuevos estimadores. La segunda, y muy importante desde el punto de vista práctico es que, los procedimientos que utilizan mezclas de distribuciones no pueden calcularse de forma exacta ya que resulta computacionalmente imposible al crecer el número de componentes de la mezcla exponencialmente, siendo necesario algún mecanismo de reducción.

El presente trabajo se centra en los modelos dinámicos, en particular en los modelos de espacio de estado. Kalman (1960) introdujo un método que permite estimar el estado de un sistema, predecir una nueva observación y actualizar el vector de estado en cada momento del tiempo t , usando mínimos cuadrados de forma recursiva. A éste método se le conoce con el nombre de filtro de Kalman y ha sido ampliamente utilizado en la literatura, especialmente en la ingeniería y, últimamente, en la economía.

El problema del filtro de Kalman es que no es un procedimiento robusto. Dentro de los trabajos que intentan robustificarlo hay que destacar el de Alspach y Sorenson (1972), y de Masreliez (1975) quien da fórmulas para el filtro cuando tanto la ecuación de observación como la de estado tienen ruidos no normales.

Son los trabajos de Harrison y Stevens (1971, 1976), los que mejor presentan la metodología del uso del mezclas de distribuciones normales, introduciendo la terminología de modelos multiproceso.

El atractivo artículo, aunque difícilmente aplicable debido al gran coste computacional, de Kitagawa (1987), presenta una solución general independiente del modelo. Esta línea la complementa el autor con los trabajos relativos al filtro y el suavizado de modelos de espacios de estado utilizando mezclas (Kitagawa

(1989,1990)). Meinhold y Singpurwalla (1989) utilizan distribuciones t-Student para el error en su intento por robustificar el filtro de Kalman.

Es el artículo de Peña y Guttman (1989) el que presenta el desarrollo de las fórmulas del filtro de Kalman utilizando mezclas de forma más clara planteando el problema de la necesidad de colapsar utilizando criterios de optimalidad.

Cómo hacer que el número de componentes no crezca exponencialmente, fué la pregunta que se hicieron Box y Tiao (1968). Un cuarto de siglo después hay ya muchos autores que han intentado dar una respuesta. El presente trabajo es una recopilación de técnicas de reducción del número de componentes de mezclas de distribuciones normales en modelos dinámicos, se proponen algunas nuevas y se hace un estudio comparativo de ellas.

2. EL FILTRO DE KALMAN ROBUSTO

Supongamos que las observaciones de una serie de tiempo $y_t, t = 1, \dots, T$ se expresan de la forma

$$y_t = \mathbf{A}_t \theta_t + v_t \quad [1]$$

$$\theta_t = \Omega_t \theta_{t-1} + w \quad [2]$$

donde \mathbf{A}_t es una matriz $1 \times p$ de coeficientes conocidos y Ω_t es una matriz $p \times p$ conocida; v_t , y w_t son serialmente incorreladas; (1) es conocida como la *ecuación de observación* o *ecuación de medida* y a (2) se le denomina *ecuación de transición* o *ecuación de estado*; θ es conocido como *estado*. Cuando $v_t \sim N(0, C_t)$ y $w_t \sim N(0, R_t)$ el modelo anterior se resuelve utilizando el filtro de Kalman, un sistema de recursiones que permite estimar el vector de estado, y predecir nuevas observaciones.

Cuando la estimación del estado en cada momento del tiempo t se hace condicionada a las observaciones pasadas, se obtiene la estimación dada por el *filtro*, cuando se condiciona a valores futuros de la serie se obtiene el *suavizado*. Sea $\mathbf{Y}_t = (y_1, \dots, y_t)$. Las ecuaciones del filtro de Kalman tienen la siguiente expresión recursiva (ver Anderson y Moore (1979) para la demostración)

$$p(\theta_t | \mathbf{Y}_t) = N(\hat{\theta}_{t|t}, \mathbf{V}_{t|t})$$

$$E[\theta_t | \mathbf{Y}_t] = \hat{\theta}_{t|t} = \hat{\theta}_{t|t-1} + \mathbf{V}_{t|t-1} \mathbf{A}_t' M_t^{-1} (y_t - \mathbf{A}_t \hat{\theta}_{t|t-1}), \quad [3]$$

$$\text{var}[\theta_t | \mathbf{Y}_t] = \mathbf{V}_{t|t} = \mathbf{V}_{t|t-1} - \mathbf{V}_{t|t-1} \mathbf{A}_t' M_t^{-1} \mathbf{A}_t \mathbf{V}_{t|t-1}, \quad [4]$$

donde, $\hat{\theta}_{t|t-1} = \Omega_t \hat{\theta}_{t-1}$, $\mathbf{V}_{t|t-1} = \mathbf{R}_t + \Omega_t \mathbf{V}_{t-1} \Omega_t'$ y $\mathbf{M}_t = \mathbf{C}_t + \mathbf{A}_t \mathbf{V}_{t|t-1} \mathbf{A}_t'$.

Aunque este procedimiento es muy eficiente bajo la hipótesis de normalidad, resulta ser muy poco robusto ante la presencia de valores atípicos. Esta falta de robustez es debida a que la estimación del estado (3) es función lineal, y por tanto no acotada, de la diferencia entre la observación, y_t , y su predicción, sin que la varianza (4) dependa de y_t , por lo que un valor atípico en y_t afecta linealmente la estimación del estado.

Robustificando el filtro

Una técnica ya clásica de robustificar el filtro es utilizar mezclas de distribuciones normales (normal contaminada). Supongamos que y_t sigue el modelo (1)-(2) donde

$$v_t \sim \sum \gamma_i N(0, C_{t,i}), i = 1, 2, \dots, I$$

$$w_t \sim \sum \beta_j N(0, R_{t,j}), j = 1, 2, \dots, J$$

donde $\sum_i \gamma_i = \sum_j \beta_j = 1$. Sea $K = IJ$.

Sea $k=(i,j)$, $k=1, \dots, K$, el indicador resumen del modelo que interviene en cada componente del proceso. Sea i_t la variable que indica qué componente ha intervenido en la etapa t , esto es $i_t = k$, implica que el par (i,j) está interviniendo en la componente de interés. Entonces, la estimación del filtro en el instante 1 será de la forma

$$\begin{aligned} p(\theta_1 | \mathbf{Y}_1) &= \sum_{i_1=1}^K p(\theta_1 | \mathbf{Y}_1, i_1) p(i_1 | \mathbf{Y}_1) \\ &= \sum_{i_1=1}^K N(\hat{\theta}_{i_1}, \mathbf{V}_{i_1}) \alpha_{i_1} \end{aligned}$$

y, para el instante t ,

$$\begin{aligned} p(\theta_t | \mathbf{Y}_t) &= \sum_{i_1=1}^K \sum_{i_2=1}^K \dots \sum_{i_t=1}^K p(\theta_t | i_1, i_2, \dots, i_t, \mathbf{Y}_t) p(i_1, i_2, \dots, i_t | \mathbf{Y}_t) \\ &= \sum_{i_1=1}^K \sum_{i_2=1}^K \dots \sum_{i_t=1}^K N(\hat{\theta}_{i_1, i_2, \dots, i_t}, \mathbf{V}_{i_1, i_2, \dots, i_t}) \alpha_{i_1, i_2, \dots, i_t} \end{aligned} \quad [5]$$

El número de componentes, en el instante t es K^t y, para el total de la serie, K^T . Es obvio que, para una longitud realista de la serie, el cálculo exacto es imposible y es necesario buscar algún tipo de aproximación reduciendo el número de componentes de la mezcla durante el proceso.

3. APROXIMANDO LA ESTIMACIÓN DEL ESTADO

El primer paso para reducir el número de componentes en (5) es tener en cuenta que en el modelo lineal básico, cuando el tiempo transcurre, lo que a ocurrido en el pasado lejano se convierte en menos importante cara a hacer inferencia sobre el presente y el futuro. Utilizando esta idea se puede establecer el primer criterio de reducción. Supongamos que la estructura de la distribución mezclada en un momento del tiempo anterior a $t-h$ no es relevante cara a hacer estimaciones sobre el estado en el instante t . Se puede entonces aproximar (5) mediante

$$p(\theta_t | i_1, i_2, \dots, i_t, \mathbf{Y}_t) \sim p(\theta_t | i_{t-h}, i_{t-h+1}, \dots, i_t, \mathbf{Y}_t)$$

con lo cual

$$\begin{aligned} p(\theta_t | \mathbf{Y}_t) &\approx \sum_{i_1=1}^K \sum_{i_{t-1}=1}^K \dots \sum_{i_{t-h}=1}^K p(\theta_t | i_{t-h}, i_{t-h+1}, \dots, i_t, \mathbf{Y}_t) p(i_{t-h}, i_{t-h+1}, \dots, i_t | \mathbf{Y}_t) = \\ &= \sum_{i_1=1}^K \sum_{i_{t-1}=1}^K \dots \sum_{i_{t-h}=1}^K N(\hat{\theta}_{i_{t-h}, i_{t-h+1}, \dots, i_t}, \mathbf{V}_{i_{t-h}, i_{t-h+1}, \dots, i_t}) \alpha_{i_{t-h}, i_{t-h+1}, \dots, i_t} \end{aligned} \tag{6}$$

donde hemos supuesto que en la etapa $t-h$ el número de componentes de la mezcla, condicionada a todo el pasado de la serie, es K . De esta forma se reduce el número de componentes a K^{h+1} de forma constante en todo el proceso, para $t > h$.

Pero el proceso, calculado de forma exacta, en el momento $t-h$ tiene K^{t-h} componentes, con lo que deberá utilizarse algún método de colapsado para reducir su número a K . Siguiendo a Harrison y Stevens (1976), colapsaremos a K componentes priorizando las últimas h etapas en la mezcla. Esto es, si se tratara de colapsar (6), cara a calcular el estado $t+1$, se procedería de la forma

$$p(\theta_t | \mathbf{Y}_t) \sim \sum_{i_1=1}^K \sum_{i_{t-1}=1}^K \dots \sum_{i_{t-h+1}=1}^K N(\hat{\theta}_{i_{t-h+1}, \dots, i_t}^*, \mathbf{V}_{i_{t-h+1}, \dots, i_t}^*) \alpha_{i_{t-h+1}, \dots, i_t}^*$$

donde

$$\alpha_{i_{t-h+1}, \dots, i_t}^* = \sum_{i_{t-h}=1}^K \alpha_{i_{t-h}, \dots, i_t}$$

$$\hat{\theta}_{i_{t-h+1}, \dots, i_t}^* = \sum_{i_{t-h}=1}^K \hat{\theta}_{i_{t-h}, \dots, i_t} \frac{\alpha_{i_{t-h}, \dots, i_t}}{\alpha_{i_{t-h+1}, \dots, i_t}^*}$$

$$\mathbf{V}_{i_{t-h+1}, \dots, i_t}^* = \sum_{i_{t-h}=1}^K [\mathbf{v}_{i_{t-h}, \dots, i_{t+h}} + (\hat{\theta}_{i_{t-h+1}, \dots, i_t}^* - \hat{\theta}_{i_{t-h}, \dots, i_t}) (\hat{\theta}_{i_{t-h+1}, \dots, i_t}^* - \hat{\theta}_{i_{t-h}, \dots, i_t})'] \frac{\alpha_{i_{t-h}, \dots, i_t}}{\alpha_{i_{t-h+1}, \dots, i_t}^*}$$

Este procedimiento, aunque reduce el número de componentes de forma sustancial, sigue teniendo el inconveniente de que el número de componentes de la mezcla en cada etapa es bastante grande y, un poco de sentido común nos hace ver que innecesario, por ejemplo, para $h=3$, y $K=4$, el número de componentes de la mezcla es de 256, y para $h=2$, 64. Es obvio que muchas de estas componentes no son significativas, puesto que el objetivo final es manejar valores atípicos y éstos solamente se producen en un conjunto pequeño de las observaciones del proceso total, teniendo los términos de error distribución normal en el resto. Por ésto, los casos más comunes de aplicación de este procedimiento se refieren a $h=1$ (ver Harrison y Stevens (1976), Gordon y Smith (1988) y West y Harrison (1989)). El caso $h=0$ es equivalente a Peña y Guttman (1989).

4. COLAPSANDO MEZCLAS

La idea que subyace bajo la palabra colapsar es, además de reducir el número de componentes de la mezcla, tratar de encontrar una distribución más simple pero que no sea muy diferente de la función de distribución objetivo: la mezcla. Para ello hay que procurar que la "distancia" entre la función ajustada y la objetivo sea lo más pequeña posible, lo que implica utilizar algún criterio de distancia o divergencia.

Desde el punto de vista práctico, el procedimiento de colapsado que se utilice deberá ser sencillo computacionalmente, ya que al tener que aplicarlo a cada etapa del proceso, si se utiliza una técnica costosa se pierde la posibilidad de aplicarlo en series de muchas observaciones.

Utilizando la divergencia de Kullback-Leibler

El criterio de divergencia más comunmente utilizado es la divergencia de Kullback-Leibler (Kullback y Leibler (1951)). Sea $p_0(\theta)$ la verdadera distribución de θ .

Definición 1. Sean $p_0(\theta)$ y $p_1(\theta)$ dos distribuciones para la variable aleatoria θ . Se define la divergencia de Kullback-Leibler (KL) entre $p_0(\theta)$ y $p_1(\theta)$ bajo $p_0(\theta)$ como

$$KL(p_0, p_1) = E_0 \left\{ \log \frac{p_0(\theta)}{p_1(\theta)} \right\} = \int \log \frac{p_0(\theta)}{p_1(\theta)} p_0(\theta) d(\theta).$$

Supongamos que queremos colapsar una mezcla a una distribución de una sólo componente normal, si

$$p_0(\theta) = \sum_{i=1}^K p_i(\theta) \alpha_i \quad [7]$$

y si $p_i(\theta) = N(\mathbf{m}_i, \mathbf{V}_i)$, la distribución normal $p_0^*(\theta) = N(\mathbf{m}, \mathbf{V})$ que minimiza la divergencia KL (Peña y Guttman (1989)), es aquella de parámetros

$$\begin{aligned} \mathbf{m} &= \sum_{i=1}^K \mathbf{m}_i \alpha_i \\ \mathbf{V} &= \sum_{i=1}^K [\mathbf{V}_i + (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})'] \alpha_i \end{aligned} \quad [8]$$

En el caso en que se aproxime por L componentes ($L < K$), el método más intuitivo consiste en minimizar la divergencia KL

$$\begin{aligned} p_0(\theta) = \sum_{i=1}^K p_i(\theta) \alpha_i \Rightarrow p^*(\theta) = \sum_{j=1}^L p_j^*(\theta) \alpha_j^* \\ KL = \int \log \frac{\sum_{i=1}^K p_i(\theta) \alpha_i}{\sum_{j=1}^L p_j^*(\theta) \alpha_j^*} \sum_{i=1}^K p_i(\theta) \alpha_i d(\theta) \end{aligned} \quad [9]$$

pero, cuando $L > 1$, minimizar la integral (9) implica optimización no lineal con respecto a $L \left(p + \frac{p(p+1)}{2} \right)$ parámetros (donde p es la dimensión de θ), en la que cada evaluación de la función deberá resolverse mediante integración numérica. Un método más sencillo lo presenta Kitagawa (1990). Basado también en la divergencia KL este método consiste en comparar dos a dos las K densidades originales y colapsar a una aquéllas que minimicen la divergencia:

$$\begin{aligned} DKL(p_i, p_j) &= -2\alpha_i \alpha_j \{ KL(p_i, p_j) + KL(p_j, p_i) \} \\ &= \alpha_i \alpha_j \{ \mathbf{V}_i^{-1} \mathbf{V}_j + \mathbf{V}_j^{-1} \mathbf{V}_i + (\mathbf{m}_i - \mathbf{m}_j)' (\mathbf{V}_i^{-1} + \mathbf{V}_j^{-1}) (\mathbf{m}_i - \mathbf{m}_j) \} \end{aligned}$$

El objeto de introducir DKL es que KL no es simétrica, y en la comparación dos a dos, ninguna de las distribuciones es la "verdadera", o tiene prioridad sobre otra. Si las densidades que minimizan DKL son las de índices i y j , entonces, la nueva densidad será $N(\mathbf{m}_{ij}, \mathbf{V}_{ij})$ donde

$$\mathbf{m}_{ij} = \frac{\alpha_i \mathbf{m}_i + \alpha_j \mathbf{m}_j}{\alpha_i + \alpha_j}$$

$$\mathbf{V}_{ij} = \frac{\alpha_i \{ \mathbf{V}_i + (\mathbf{m}_{ij} - \mathbf{m}_i)(\mathbf{m}_{ij} - \mathbf{m}_i)' \} + \alpha_j \{ \mathbf{V}_j + (\mathbf{m}_{ij} - \mathbf{m}_j)(\mathbf{m}_{ij} - \mathbf{m}_j)' \}}{\alpha_i + \alpha_j}$$

Esta componente se une al conjunto de $K - 2$ densidades, y se vuelve a repetir el proceso hasta obtener una mezcla de L componentes.

Mirando la modalidad

Algunos autores como Meinhold y Singpurwalla (1989), utilizan el criterio de la modalidad para colsapsar mezclas: si la distribución mezclada es unimodal, se colapsa a una normal con media y varianza dadas por (8).

En el caso de una mezcla de dos componentes normales univariantes, el siguiente resultado debido a Robertson y Fryer (1969), caracteriza la unimodalidad.

Teorema 1. Sea x una variable aleatoria con función de distribución

$$p(x) = \alpha N(m_1, V_1) + (1 - \alpha)N(m_2, V_2),$$

supongamos, sin pérdida de generalidad, que $m_2 > m_1$ y definamos

$$m = \frac{m_2 - m_1}{V_1^{1/2}}, \quad V^{1/2} = \frac{V_2^{1/2}}{V_1^{1/2}}$$

entonces:

(i) p es unimodal si $0 < m \leq m_0$, donde

$$m_0 = \left[2(V^2 - V + 1)^{3/2} - (2V^3 - 3V^2 - 3V + 2) \right]^{1/2} / V^{1/2},$$

(ii) para $m > m_0$, p es bimodal si y sólo si, α está dentro del intervalo abierto (α_1, α_2) donde

$$\alpha_i^{-1} = 1 + \frac{V^{3/2} a_i}{m - a_i} \exp \left\{ -\frac{1}{2} a_i^2 + \frac{1}{2} \frac{(a_i - m)^2}{V} \right\}, \quad i = 1, 2$$

siendo (a_1, a_2) las raíces de la ecuación

$$(V - 1)a^3 - m(V - 2)a^2 - m^2a + mV = 0$$

con $0 < a_1 < a_2 < m$.

(iii) En otro caso, p es unimodal

Corolario 1. En las condiciones del teorema anterior se cumple que

(i) Si $m \leq 2\min(1, V^{1/2})$, p es unimodal.

(ii) Si $m \geq \frac{3}{2}\sqrt{3\min(1, V^{1/2})}$, entonces existe un intervalo (α_1, α_2) tal que p es bimodal para $\alpha_1 < \alpha < \alpha_2$.

El intervalo (α_1, α_2) introducido en la segunda parte del corolario nos lleva al llamado por O'Hagan (1981) *momento de indecisión*. Este momento de indecisión juega un papel importante en todo lo referente al tratamiento mediante mezclas de valores atípicos en modelos dinámicos. Pensemos en valores atípicos aditivos, la idea intuitiva es que si el valor del outlier es muy grande la componente de gran varianza de la mezcla será la que modelice esa observación, siendo las componentes de menor varianza irrelevantes, con un resultado final que será una distribución muy cercana a la normal. Cuando la observación no es un outlier claro, las distintas componentes de la mezcla pueden tener pesos relevantes, dando lugar a multimodalidad en la distribución final.

Los resultados relativos a la unimodalidad, presentan diversos problemas a la hora de aplicarlos como técnica de decisión para colapsar mezclas en modelos dinámicos. Primero, no se generalizan para una mezcla de K componentes, $K > 2$ por lo que habría que diseñar un mecanismo de comparación de las componentes dos a dos. Segundo, la unimodalidad no implica estar muy próximos a la normalidad; habría que tener en cuenta la bitangencialidad (que al menos una tangente corte dos puntos de la función de densidad), que es equivalente a la existencia de cuatro puntos de inflexión. La caracterización de la bitangencialidad es más compleja y Robertson y Fryer (1969) dan un análisis bastante completo de la misma. Y, tercero, los resultados multivariantes (Konstantellos (1980)) son muy limitados.

Colapsar por momentos

Introduzcamos un breve resumen de los momentos de una mezcla. Sea x una variable aleatoria univariante con función de distribución

$$x \sim \sum_{i=1}^I \alpha_i N(m_i, V_i) \quad [10]$$

donde $\sum_{i=1}^I \alpha_i = 1$, y $\alpha_i > 0$, entonces, la media de x será

$$m = \sum_{i=1}^I \alpha_i m_i \quad [11]$$

y la varianza

$$V = \sum_{i=1}^I \alpha_i (V_i + (m_i - m)^2) \quad [12]$$

El coeficiente de asimetría es

$$C_{as} = \frac{\sum_{i=1}^I \alpha_i (3V_i (m_i - m) + (m_i - m)^3)}{V^{3/2}}$$

y el coeficiente de apuntamiento será

$$C_{ap} = \frac{\sum_{i=1}^I \alpha_i (3V_i^2 + 6V_i (m_i - m)^2 + (m_i - m)^4)}{V^2}.$$

Mirando a la media y a la varianza

Sea $p(\theta)$ es una mezcla como en (7) entonces, si para dos componentes i, j , se cumple que

$$m_i = m_j \pm \varepsilon_1$$

$$V_i = V_j \pm \varepsilon_2$$

para ε_1 y ε_2 suficientemente pequeños, entonces las componentes i, j , se colapsan a una, reduciendo las componentes.

Este criterio es demasiado exigente ya que sólo permite colapsar aquellas distribuciones que sean aproximadamente idénticas, sin tener en cuenta los valores de α , por lo que renunciamos a colapsar las distribuciones cuya mezcla es aproximadamente una distribución normal.

Un procedimiento que considera también el vector de parámetros (m_i, V_i) es el presentado por West (1993), que consisten en: (1) ordenar las K componentes de la mezcla de acuerdo con el valor de α_i ; (2) calcular la distancia euclídea entre el vector de parámetros correspondiente al α mínimo y el resto de los vectores de

parámetros; y (3) colapsar las dos componentes que minimizan esta distancia y repetir el proceso hasta llegar al número de componentes deseado.

Este criterio tiene en cuenta α en el sentido de que trata de colapsar aquellas componentes cuyo peso es más pequeño, no obstante, tampoco tiene en cuenta la forma de la distribución, y obliga a trabajar con un número fijo de componentes durante todo el proceso, cuando en algunos casos puede que no sea necesario.

Mirando los coeficientes de asimetría y apuntamineto

Un criterio para decidir cuándo aproximar una mezcla de distribuciones normales por una única distribución normal es utilizar el criterio de los momentos. La utilización de los momentos de orden alto (asimetría y apuntamiento), ya fue utilizado por Mardia (1970) para hacer test de ajuste para distribuciones normales.

Una referencia para contrastar cómo de asimétrica es una distribución unimodal es compararla con la χ^2 . El estudio del coeficiente de asimetría no es suficiente, ya que podemos encontrarnos con funciones simétricas multimodales y, por tanto, muy alejadas de la normal. Para evitar ésto hay que considerar también el coeficiente de apuntamiento que, en el caso de multimodalidad, tendrá un valor inferior a 3. Una referencia para contrastar los valores del coeficiente de apuntamiento de una distribución simétrica, es compararlo con el de la distribución t-Student. El siguiente cuadro refleja algunos de estos valores.

Grados de libertad	Coficiente de Asimetría χ^2	Coficiente de Apuntamiento t-Student
5	1,265	9.000
20	0,632	3.375
50	0,400	3.130
100	0,283	3.062
500	0,126	3.012
1000	0,089	3.001

Momentos multivariantes

Un aspecto importante a favor del método de los momentos es su generalización multivariante. Sea \mathbf{x} un vector aleatorio con

$$E(\mathbf{x}) = \mathbf{m}$$

$$\text{Var}(\mathbf{x}) = E\{(\mathbf{x} - \mathbf{m})(\mathbf{x} - \mathbf{m})'\} = \mathbf{V}$$

Se definen los coeficientes de asimetría, C_{as} y apuntamiento, C_{ap} como

$$C_{as} = E\left\{\left[(\mathbf{x} - \mathbf{m})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{m})\right]^3\right\}$$

$$C_{ap} = E\left\{\left[(\mathbf{x} - \mathbf{m})' \mathbf{V}^{-1} (\mathbf{x} - \mathbf{m})\right]^2\right\}$$

Siendo \mathbf{x} e \mathbf{y} vectores aleatorios independientes idénticamente distribuidos. En el caso en que la distribución sea $N(\mathbf{m}, \mathbf{V})$ se cumple (Mardia (1989, pag. 41)) que $C_{as} = 0$ y $C_{ap} = p(p + 2)$ siendo p la dimensión de \mathbf{x} .

Si la distribución de \mathbf{x} es una mezcla de normales multivariantes, se tiene que:

$$p(\mathbf{x}) = \sum_{i=1}^I \alpha_i N(\mathbf{m}_i, \mathbf{V}_i)$$

$$\mathbf{m} = E(\mathbf{x}) = \sum_{i=1}^I \alpha_i \mathbf{m}_i$$

$$\mathbf{V} = \sum_{i=1}^I \alpha_i [\mathbf{V}_i + (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})']$$

$$C_{as} = \sum_{i,j=1}^I \alpha_i \alpha_j \left\{ \sum_{k=1}^p \lambda_{kij} D_{ij} + 6(\mathbf{m}_i - \mathbf{m})' \mathbf{V}^{-1} \mathbf{V}_i \mathbf{V}^{-1} [\mathbf{V}_j \mathbf{V}^{-1} (\mathbf{m}_j - \mathbf{m}) + (\mathbf{m}_i - \mathbf{m})] \mathbf{D}_{ij} + D_{ij}^3 \right\}$$

donde $D_{ij} = (\mathbf{m}_i - \mathbf{m})' \mathbf{V}^{-1} (\mathbf{m}_i - \mathbf{m})$ y λ_{kij} son los autovalores de $\mathbf{L}_i' \mathbf{V}^{-1} \mathbf{V}_j \mathbf{V}^{-1} \mathbf{L}_i$ con $\mathbf{V}_i = \mathbf{L}_i \mathbf{L}_i'$.

$$C_{ap} = \sum_{i=1}^I \alpha_i \left\{ 4(\mathbf{m}_i - \mathbf{m})' \mathbf{V}^{-1} \mathbf{V}_i \mathbf{V}^{-1} (\mathbf{m}_i - \mathbf{m}) + D_i^2 + 2D_i \sum_{k=1}^p \lambda_{ki} + 3 \sum_{k=1}^p \lambda_{ki}^2 + \sum_{k \neq l} \lambda_{ki} \lambda_{li} \right\}$$

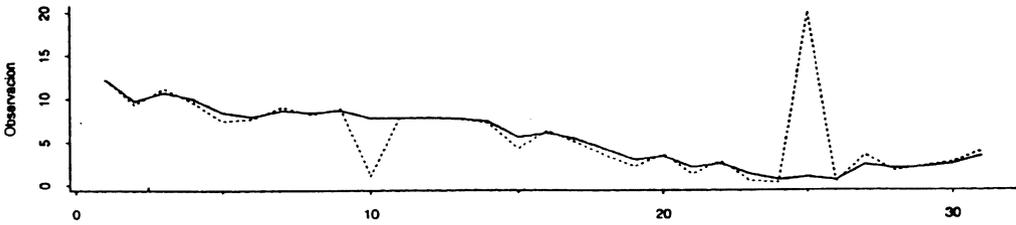
donde $D_i = (\mathbf{m}_i - \mathbf{m})' \mathbf{V}^{-1} (\mathbf{m}_i - \mathbf{m})$ y λ_{ki} son los autovalores de $\mathbf{L}_i' \mathbf{V}^{-1} \mathbf{L}_i$ y $\mathbf{V}_i = \mathbf{L}_i \mathbf{L}_i'$. La demostración de estos resultados se encuentra en el Apéndice I.

El gráfico 1 muestra los resultados de la utilización del criterio de los momentos para la decisión de colapsar o no en el caso de un outlier aditivo. Se puede observar que en los casos en que existe un outlier aditivo muy grande los momentos de la distribución del estado son muy próximos a los de la distribución normal, mientras que se alejan de ella cuando el valor que se observa no es un valor atípico claro.

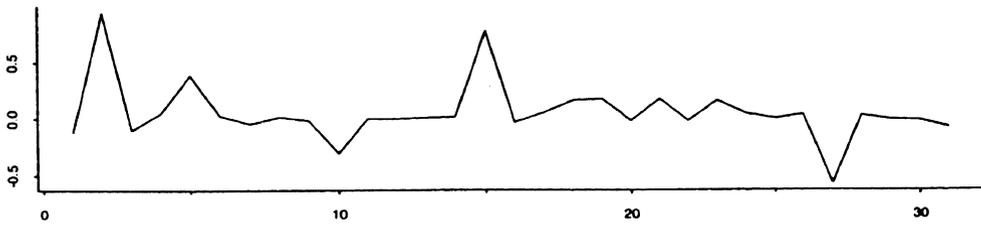
El gráfico 2 muestra un ejemplo de cambio de nivel.

Grafico 1. Manejando outliers aditivos con mezclas

Serie original y estimaci3n del estado



Asimetria



Apuntamiento

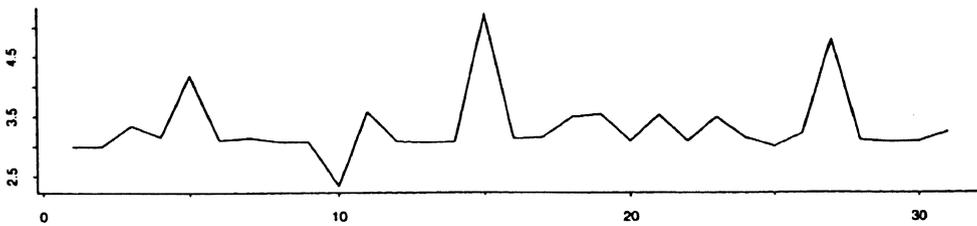
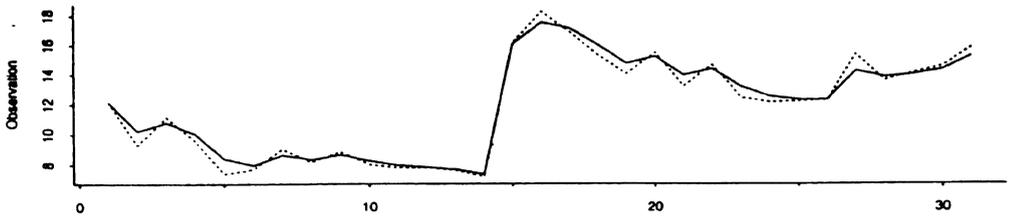
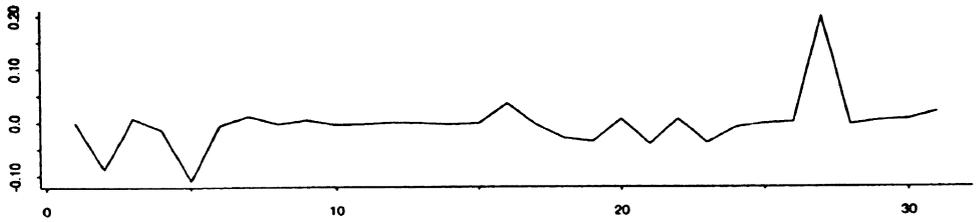


Grafico 2. Manejando cambios de nivel mediante mezclas

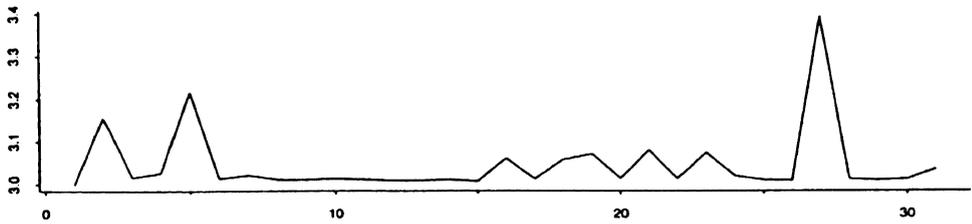
Serie original y estimacion del estado



Asimetria



Apuntamiento



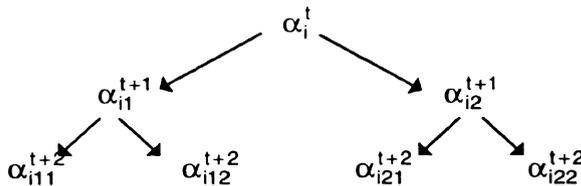
La poda

Un criterio para reducir el número de componentes en una mezcla se basa en eliminar aquellas componentes que tienen un peso muy pequeño en la distribución final. Esto es, si en (7) se cumple que

$$\alpha_i < \varepsilon$$

entonces la componente i -ésima se elimina de la mezcla ajustando las ponderaciones. Al valor ε le llamaremos umbral. Bruce y Martin (1992) fijan el umbral en 0.025.

Algunos autores como Kitagawa (1990) sugieren que éste podría no ser un buen método ya que se pueden estar eliminando componentes que, si bien en la etapa t tienen un peso muy pequeño, pueden ser madres de otras que en la etapa $t+1$ sean importantes. Aunque, la experiencia de la autora pone de manifiesto que esto no es un riesgo muy grande ya que en la estructura básica del modelo aparecen de nuevo todas las componentes, se podría hacer la poda en varias etapas. Esto es, siguiendo las reglas de los árboles de decisión de la inteligencia artificial: eliminar una rama a partir del momento t si sus hijos en el momento $t+1, \dots, t+h$ tienen todos una probabilidad de ocurrir muy pequeña. Dado el siguiente árbol



si $\alpha_i^t < \varepsilon, \alpha_{ij}^{t+1} < \varepsilon, (j = 1, 2)$, y $\alpha_{ijl}^{t+2} < \varepsilon$, entonces se elimina la componente i -ésima de la etapa t y se ajustan las probabilidades en $t, t+1$, y $t+2$.

Utilizando las mezclas sólo en determinadas partes del proceso

Llegados a este punto, queda claro que el modelo de mezclas es un modelo explosivo y, aún utilizando técnicas de aproximación, computacionalmente costoso. Se da también la paradoja de estar utilizando herramientas excesivamente costosas para modelizar valores atípicos que, por definición, aparecerán solamente en un porcentaje muy pequeño de los puntos del proceso.

Es natural pensar que debería generarse un mecanismo que detecte cuándo el modelo de errores normales (sin mezcla) deja de comportarse bien y entonces utilizar las mezclas. Para ello hay que contar con algún tipo de indicador que muestre cuándo nos hemos alejado de la normalidad: hay que monitorizar el proceso. Siguiendo a Ameen y Harrison (1985), el proceso sería:

i. Modelizar la serie según el modelo normal y utilizar un indicador como el *Factor de Bayes* o el *umbral en el error de predicción*. Mientras el indicador diga que el modelo estándar es bueno, no utilizar las mezclas.

ii. Cuando el indicador nos muestre que el modelo se deteriora al no predecir convenientemente, se introducen las mezclas.

iii. Continuar con las mezclas hasta que la situación se haya normalizado, esto es, hasta que no se "gane" nada por utilizar mezclas.

iv. Volver al modelo estándar como en el paso (i) y continuar el proceso.

En este esquema el uso de las mezclas se reduce a periodos de incertidumbre acerca de la serie.

El factor de Bayes

Supongamos que dos modelos tienen la misma estructura matemática pero diferentes parámetros. Llamemos al modelo M y sean M_0 y M_1 los dos modelos, donde M_0 refleja la situación estándar y M_1 las situaciones especiales. Sea

$$p(y_t | \mathbf{Y}_{t-1}, M_i), \quad i = 0, 1$$

la función de densidad de la predicción con relación a cada modelo.

Definición 2. El factor Bayes para M_0 frente a M_1 basado en la observación y_t se define como

$$H_t = \frac{p(y_t | \mathbf{Y}_{t-1}, M_0)}{p(y_t | \mathbf{Y}_{t-1}, M_1)} = \frac{p_0(y_t | \mathbf{Y}_{t-1})}{p_1(y_t | \mathbf{Y}_{t-1})}$$

Definición 3. El factor Bayes para M_0 frente a M_1 basado en las observaciones $y_t, y_{t-1}, \dots, y_{t-h+1}$ se define como

$$H_t(h) = \prod_{r=t-h+1}^t H_r = \frac{p_0(y_t, \dots, y_{t-h+1} | \mathbf{Y}_{t-h})}{p_1(y_t, \dots, y_{t-h+1} | \mathbf{Y}_{t-h})} \quad [13]$$

El criterio de decisión es el siguiente:

- $\log H_t(h) \approx 1$ hay indicios a favor del modelo 0
- $\log H_t(h) \approx 2$ hay fuertes indicios a favor del modelo 0
- $\log H_t(h) \approx -1$ hay indicios a favor del modelo 1
- $\log H_t(h) \approx -2$ hay fuertes indicios a favor del modelo 1

Umbral en el error de predicción

Sea

$$\hat{y}_{t+1|t}$$

el valor de la predicción para el instante $t + 1$ a partir de la información contenida en Y_t siguiendo el modelo gaussiano. Sea

$$e_{t+1} = \hat{y}_{t+1|t} - y_{t+1}$$

y notemos por e_{t+1}^* el error estandarizado. Entonces, si

$$e_{t+1}^* > \varepsilon$$

el modelo normal no es un buen predictor por lo que en la observación $t + 1$ deberán utilizarse mezclas. Puesto que estamos simplificando bastante el procedimiento, la elección de ε puede ser bastante rigurosa y situarse, incluso, por debajo de 2. El indicador del error en el umbral de predicción tiene una generalización multivariante inmediata. El gráfico 3 muestra un ejemplo de utilización del método. La observación 9 y 24 muestran un gran error de predicción por lo que la 10 y la 25 se han calculado utilizando mezclas. Los errores obtenidos en este caso han sido moderados, volviendo al modelo normal en la siguiente observación.

5. RESULTADOS COMPARATIVOS

Simulación

Para comparar los diferentes métodos de colapsado se ha realizado un estudio de los mismos basado en el error cuadrático. Para ello se han simulado 500 series de tiempo de longitud 50, siguiendo el modelo

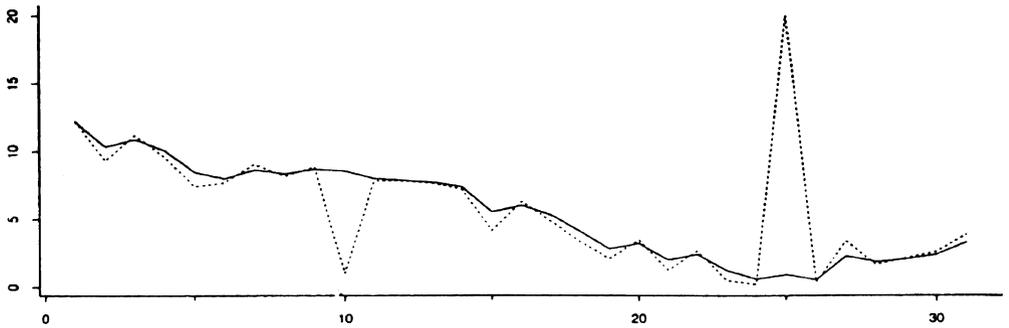
$$\begin{aligned} y_t &= \theta_t + v_t, \\ \theta_t &= \theta_{t-1} + w_t, \end{aligned}$$

con $v_t \sim N(0,1)$, y $w_t \sim N(0,1)$ con valor inicial $\hat{\theta}_0 = 10$. A cada serie se le añadieron cuatro valores atípicos aditivos según el modelo

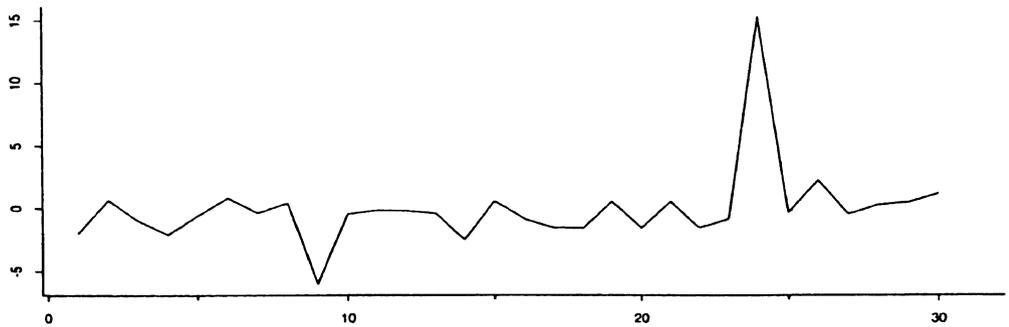
$$y_t^* = y_t + e_t, \quad i = 1,2,3,4 \quad e_t \sim N(0,10)$$

Gráfico 3. Utilizando el umbral en el error de predicción. Umbral=2

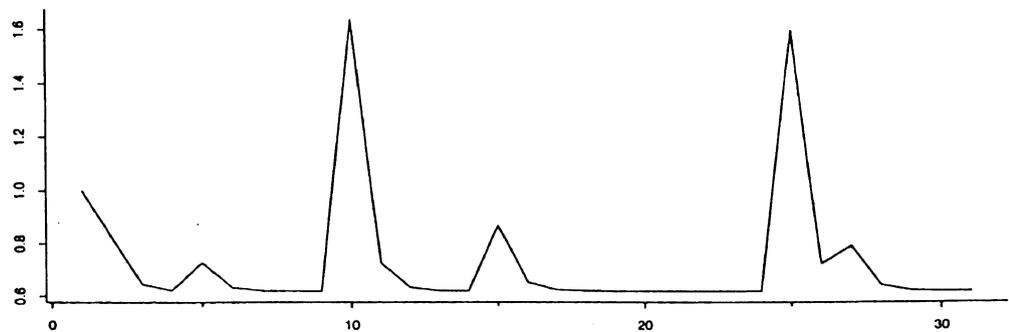
(a). Serie original y estimación del estado



(b). Error de predicción



(c). Varianza del estado



donde t_i fueron seleccionados de forma uniforme en el intervalo (4,50) para evitar la existencia de outliers en las primeras observaciones. El error que se consideró fue

$$EC = \sum_{i=1}^{50} (\theta_i - \hat{\theta}_{it})^2$$

siendo $\hat{\theta}_{it}$ los valores estimados para el estado en cada punto. Cada serie simulada se filtró utilizando el modelo

$$y_t = \theta_t + v_t^*, \quad v_t^* \sim 0.95N(0,1) + 0.05N(0,1)$$

$$\theta_t = \theta_{t-1} + w_t, \quad w_t \sim N(0,1)$$

y también se consideró el filtro de Kalman no robusto.

Los diferentes métodos de reducción utilizados se describen a continuación.

ST. Filtro de Kalman no robusto.

PG. Método de Peña y Guttman: colapsar la estimación del estado a una componente en cada etapa.

LG. Método de Lin y Guttman (1993): es una variante del método de Peña y Guttman que consiste en colapsar a una componente la función de verosimilitud en vez de la estimación del estado.

EP. Error de predicción. Utilizar mezclas solamente en aquellos puntos en que el error de predicción es mayor que un límite, en concreto, se utilizan mezclas si

$$\sqrt{\frac{(y_{t+1} - \theta_{tt})^2}{(1 + V_{tt})}} > 2$$

HS. Método de Harrison y Stevens. Utilizar siempre dos etapas y colapsar en la última. Esto es, considerar siempre dos componentes en la mezcla colapsando en la etapa anterior

$$\alpha_{t,j} = \sum_{i=1}^2 \alpha_{t,ij}$$

$$\hat{\theta}_{tt}^{(j)} = \sum_{i=1}^2 \frac{\alpha_{t,ij}}{\alpha_{t,j}} \hat{\theta}_{tt}^{(ij)}$$

$$V_{tt} = \sum_{i=1}^2 \frac{\alpha_{t,ij}}{\alpha_{t,j}} \left[V_{tt}^{(ij)} + (\hat{\theta}_{tt}^{(ij)} - \hat{\theta}_{tt}^{(j)})^2 \right]$$

donde i representa la componente de la mezcla en el instante $t-1$, y j representa la componente que interviene en el instante t .

PO. Método de la poda en una etapa. Se elimina una componente si $\alpha_{t,i} < 0.03$.

PO2. Método de la poda en dos etapas. Se elimina la componente y en el instante t si $\alpha_{t,i} < 0.05$ y $\alpha_{t+1,ij} < 0.05$, $j=1,2$ donde j representa la componente que interviene en el instante $t+1$.

MO. Método de la moda. Siguiendo a Meinhold y Singpurwalla, se ordenan las distribuciones según su media de menor a mayor y se comparan dos a dos consecutivas. Denotemos con exponente i los parámetros de la distribución que ocupa el lugar i -ésimo en la ordenación, entonces, si

$$(\hat{\theta}_{i|t}^{(i+1)} - \hat{\theta}_{i|t}^{(i)}) < 2\min(V_{i|t}^{i+1}, V_{i|t}^i)$$

la distribución es unimodal y se procede a colapsar. Aunque estos autores no describen como continúa el procedimiento, es obvio que la nueva distribución deberá entrar en el proceso de colapsado siendo ésta la que se compara con la distribución $i+2$. En caso de no colapsar las componentes $i, i+1$, la componente i permanecerá en la mezcla y se comparará $i+1$ con $i+2$ y así sucesivamente.

MOM. Método de los momentos. Se colapsa a una distribución normal si el coeficiente de asimetría es menor que 0.5 y el de apuntamiento es menor que 3.5 y mayor que 3.

Resultados obtenidos

El gráfico 4 muestra los Box-plot del error cuadrático medio no apreciándose diferencias significativas entre los diferentes métodos, y manifiestándose claramente la ventaja de utilizar los mismos frente al filtro de Kalman estándar. El gráfico 5 muestra los intervalos de confianza del error. Se observa cómo todos los modelos mejoran sustancialmente el filtro de Kalman, siendo el método de Lin y Guttman, el que presenta una diferencia significativa. Los demás son prácticamente idénticos, observándose una ligera mejoría en el método del error de predicción. Esto es debido a que al utilizar mezclas, podemos estar "empeorando" la estimación en aquellos puntos en que no hay valores atípicos.

Desde el punto de vista computacional los métodos más rápidos son aquellos que no permiten que el número de componentes crezca en ningún caso, esto es, PG, HS, EP.

Gráfico 4. Box-plot del error

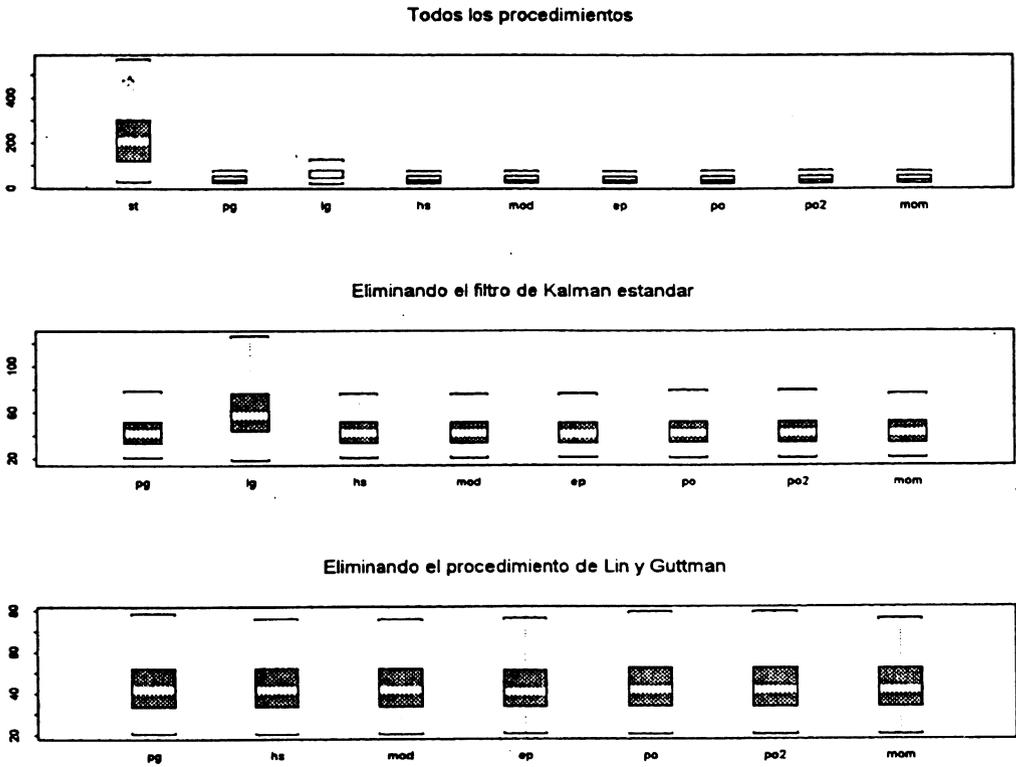
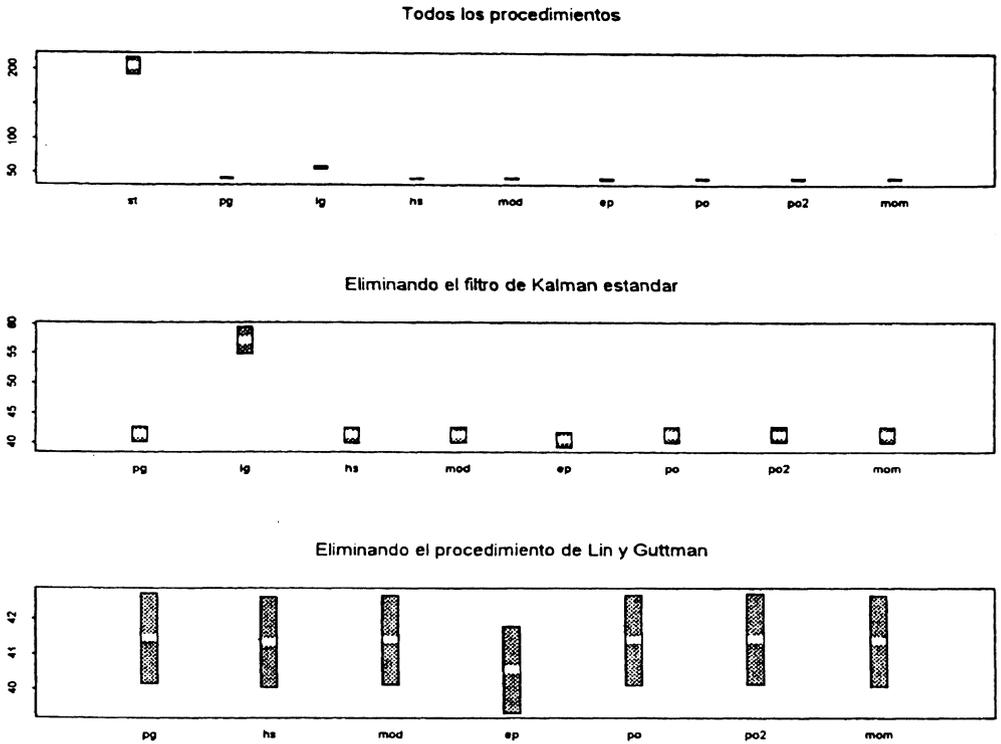


Gráfico 5. Intervalos de confianza robustos del error cuadrático medio $\alpha = 95\%$ 

El método menos eficiente computacionalmente es el método de los momentos, siendo el que es más rígido a la hora de colapsar, esto es debido a que una vez el número de componentes crece, no se permite una reducción poco a poco, sino que solamente se colapsa en aquellos casos en que se colapsa a una sola componente.

El método de la moda es también bastante lento. La condición suficiente que se utiliza es demasiado fuerte, impidiendo colapsar distribuciones que son unimodales, aunque no cumplen la condición.

6. CONCLUSIONES

Hemos presentado un conjunto de técnicas para colapsar mezclas en modelos dinámicos. La más novedosa y generalizable es el método de los momentos aunque implica un alto coste computacional, siendo quizá demasiado sensible según se observa en el gráfico 1.

El método de la unimodalidad no es generalizable y presenta problemas de tipo computacional al tener que hacerse comparaciones dos a dos, como también ocurre con la distancia de Kullback-Leibler.

Utilizar mezclas sólo en determinadas partes del proceso, con un indicador como el umbral en el error de predicción parece el más coherente con la falta de robustez del filtro de Kalman: es precisamente porque este error no está acotado por lo que, como veíamos en la Sección 2, el filtro no es robusto. Este método, además de no ser costoso computacionalmente, es el que mejores resultados ha ofrecido en el ejercicio de simulación realizado. Además tiene generalización multivariante y es independiente del número de componentes de la mezcla.

APÉNDICE I. CÁLCULO DE LOS MOMENTOS EN MEZCLAS DE NORMALES

Coeficiente de apuntamiento

Siguiendo a Mardia (1970), el coeficiente de apuntamiento multivariante se define de la forma

$$C_{ap} = E\left\{\left[(x - m)' V^{-1}(x - m)\right]^2\right\}$$

donde \mathbf{m} y \mathbf{V} son la media y la varianza del vector aleatorio \mathbf{x} . En el caso en que \mathbf{x} se distribuya como

$$p(x) = \sum_{i=1}^l \alpha_i p_i(x) = \sum_{i=1}^l \alpha_i N(m_i, V_i) \quad [14]$$

se tiene que

$$C_{ap} = \sum_{i=1}^l \alpha_i \int [(x-m)' V^{-1} (x-m)]^2 p_i(x) d(x) \quad [15]$$

$$= \sum_{i=1}^l \alpha_i \int [(x-m_i)' V^{-1} (x-m_i)]^2 p_i(x) d(x) \quad [16]$$

$$+ 2D_i \int (x-m_i)' V^{-1} (x-m_i) p_i(x) d(x) + \quad [17]$$

$$+ 4(m_i - m)' V^{-1} V_i V^{-1} (m_i - m) + D_i^2 \} \quad [18]$$

con $D_i = (m_i - m)' V^{-1} (m_i - m)$ y donde se cumple que

$$\int (x-m_i)' V^{-1} (x-m_i) (x-m_i)' V^{-1} (m_i - m) p_i(x) d(x) = 0$$

debido a la simetría de la función a integrar con respecto a $x - m_i$. Quedan pues por resolver las dos integrales (16) y (17). Sea $\int f(x) p_i(x) d(x) = E_i[f(x)]$. Al ser V_i una matriz simétrica, existe una matriz L_i triangular superior tal que

$$E_i(x - m_i)(x - m_i)' = V_i = L_i L_i'$$

Sea $z_i = L_i^{-1}(x - m_i)$ Como $L_i' V_i^{-1} L_i$ es simétrica, existe una matriz Λ_i diagonal tal que $P_i' L_i' V_i^{-1} L_i P_i = \Lambda_i$. Sea $w_i = P_i' z_i = P_i' L_i^{-1}(x - m_i)$ entonces

$$w_i \sim N(0, I), \quad w_{ij} \sim N(0, 1), \quad j = 1, \dots, p$$

por tanto,

$$\begin{aligned} E_i[(x - m_i)' V^{-1} (x - m_i)] &= E_i[(x - m_i)' L_i^{-1} P_i \Lambda_i P_i L_i^{-1} (x - m_i)] = \\ &= E[w_i' \Lambda_i w_i] = E\left[\sum_{k=1}^p \lambda_{ki} w_{ki}^2\right] = \sum_{k=1}^p \lambda_{ki} \end{aligned}$$

ya que si $w_{ki} \sim N(0,1)$, entonces $w_{ki}^2 \sim \chi_1^2$ y $E(w_{ki}^2) = 1$ siendo λ_{ki} , $k = 1, \dots, p$ los autovalores de $L_i V^{-1} L_i$. Queda así resuelta la integral (17). Nos queda por calcular

$$\begin{aligned} E_i \left\{ \left[(x - m_i)' V^{-1} (x - m_i) \right]^2 \right\} &= E \left\{ \left[w_i' \Lambda_i w_i \right]^2 \right\} = E \left\{ \left[\sum_{k=1}^p \lambda_{ki} w_{ki}^2 \right]^2 \right\} \\ &= 3 \sum_{k=1}^p \lambda_{ki}^2 + \sum_{k \neq l}^p \lambda_{ki} \lambda_{li} \end{aligned}$$

Sustituyendo estas expresiones en (18) se obtiene el resultado deseado.

Coefficiente de asimetría

Por definición, $C_{as} = E \left\{ \left[(x - m)' V^{-1} (y - m) \right]^3 \right\}$, siendo \mathbf{x} e \mathbf{y} vectores aleatorios independientes idénticamente distribuidos. Si \mathbf{x} se distribuye como (14) se tiene que

$$\begin{aligned} C_{as} &= \sum_{i=1}^I \alpha_i \sum_{j=1}^I \alpha_j \int \left[(x - m)' V^{-1} (y - m) \right]^3 p_i(x) p_j(y) dx dy \\ &= \sum_{i=1}^I \alpha_i \sum_{j=1}^I \alpha_j \int \left[(x - m_i + m_i - m)' V^{-1} (y - m_j + m_j - m) \right]^3 p_i(x) p_j(y) dx dy \\ &= \sum_{i=1}^I \sum_{j=1}^I \{ 3D_{ij} E_{ij} (x - m_i)' V^{-1} V_j V^{-1} (x - m_i) + \\ &\quad + 3D_{ij} (m_i - m)' V^{-1} V_i V^{-1} (m_j - m) + \\ &\quad + 3D_{ij} (m_i - m)' V^{-1} V_j V^{-1} (m_i - m) + \\ &\quad + 6D_{ij} (m_i - m)' V^{-1} V_j V^{-1} V_i V^{-1} (m_j - m) + \\ &\quad + D_{ij}^3 \} \end{aligned}$$

donde $D_{ij} = (m_i - m)' V^{-1} (m_j - m)$. Siguiendo los pasos ya descritos en el cálculo del coeficiente de apuntamiento se obtiene el resultado deseado.

REFERENCIAS

- ALSPACH, D.L. y SORENSON, H.W. (1972). «Nonlinear Bayesian Estimation Using Gaussian Sum Approximations». *IEEE Transactions on Automatic Control*, **AC-17**, 439-448.
- AMEEN, J.R.M. y HARRISON, P.J. (1985). «Discount Bayesian multiprocess modelling with cusums». *Time Series Analysis: Theory and Practice 5*, O.D. Anderson (De.). North-Holland, Amsterdam.
- ANDERSON, B.D.O. y MOORE, J.B. (1979). «Optimal Filtering», *Prentice-Hall*. New Jersey.
- BOX, G.E.P. y TIAO, G.C. (1968). «A Bayesian approach to some outlier problems», *Biometrika*, **55**, 119-129.
- BRUCE, A.G. y MARTIN, R.D. (1992). «Tree Based Robust Bayesian Estimation of time series structural models». *Census Technical Report*. U.S. Bureau of the Census. Whashington D.C.
- GORDON, K. y SMITH, A.F.M. (1988). «Modeling and Monitoring Discontiuous changes in Time Series. James Spall (Ed.)». *Bayesian Analysis of Time Series and Dinamic Models*. Marcel Dekker, Inc. New York.
- HARRISON, P.J., y STEVENS, C.F. (1971). «A Bayesian approach to short-term forecasting». *Oper. Res. Quart.*, **22**, 341-362.
- HARRISON, P.J., y STEVENS, C.F. (1976). «Bayesian forecasting (con discusión)». *Journal of the Royal Statistical Society, Ser. B* **38**, 205-247.
- KALMAN, R. E. (1960). «A new approach to linear filtering and prediction problems». *Transactions of ASME*, **82D**, 35-47.
- KITAGAWA, G. (1987). «Non-Gaussian State Space Modeling of Nonstationary Time Series». *Journal of the American Statistical Association*, **76**, 1032-1064.
- KITAGAWA, G. (1989). «Non-Gaussian Seasonal Adjustment». *Computers and Mathematics with applications*, **118**, 503-514.
- KITAGAWA, G. (1990). «The two filter formula for smoothing and an implementation of the Gaussian-sum smoother». *Technical report, The Institute of Statistical Mathematics*, 4-6-7 Minami-Azabu, Minato-ku Tokyo, Japan 106.
- KONSTANTELLOS, A.C. (1980). «Unimodality conditions for Gaussian sums». *IEEE Trans. Automatic Control*, **AC-25**, 838-839.
- KULLBACK, S. y LEIBLER, R. A. (1951). «On information and sufficiency». *Ann. Math. Statis.*, **22**, 79-86.
- LIN, D. y GUTTMAN, I. (1993). «Handling spuriousity in the Kalman filter». *Statistics and Probability Letters*, **16**, 259-268.

- MARDIA, K. V. (1970) «Measures of multivariate skewness and kurtosis with applications». *Biometrika*, **57**, 519-530.
- MARDIA, K. V. (1989). «Multivariate Analysis». *Academic Press*. Londres.
- MASRELIEZ, C.J. (1975). «Approximate Non-Gaussian Filtering with Linear State and Observation Relations». *IEEE Transactions of Automatic Control*, AC-20, 107-110.
- MEINHOLD, R.J. y SINGPURWALLA, N. D. (1989). «Robustification of Kalman Filter Models». *Journal of the American Statistical Association*, **84**, 479-486.
- PEÑA, D. y GUTTMAN, I. (1989) «Optimal Collapsing of Mixture Distributions in Robust Recursive Estimation». *Commun. Statist.-Theory Meth.*, **18(3)**, 817-833.
- O'HAGAN, A. (1981). «A moment of indecision». *Biometrika*, **68**, 329-330.
- ROBERTSON, C. A., y FRYER, J. G. (1969). «Some descriptive properties of normal mixtures». *Scandinavian Actuarial Journal*, **52**, 137-146.
- WEST, M. (1993). «Approximating Posterior Distributions by Mixtures». *Journal of the Royal Statistical Society, Ser. B* **55**, 409-422.
- WEST, M., y HARRISON, J. (1989). «Bayesian Forecasting and Dynamic Models.» *Springer-Verlag*. New York.

REDUCING THE NUMBER OF COMPONENTS OF A MIXTURE IN DYNAMIC MODELS

SUMMARY

The Kalman filter as a least square procedure, suffers from the lack of robustness. Some authors have tackled this problem of robustification by using normal distribution mixtures as a model for the error both in the observation and the state equations. However, the accurate filter solution is not operative with an explosive model, since the number of the mixture's components increases exponentially with t . This paper endeavours to present different ways of reducing the number of components of the mixture in each stage, using several approaches such as: Kulback-Leibler distance, pruning, unimodality and moments, giving some comparative results.

Key words: Mixture distributions, to collapse, robust Kalman filter, mixtures approach, outlier observations.

Classification AMS: 62M20, 62F15.

