

Orden de un Modelo Estadístico: Estimación y Tests

D. Dacunha-Castelle *

El estudio del orden de un modelo estadístico es un problema práctico que aparece en todo tipo de aplicaciones: estadística clásica, series cronológicas, mezclas de poblaciones, cadenas de Markov escondidas.

En realidad, todos los modelos estadísticos construidos con un número finito de parámetros tienen que ver con el problema del orden que es, *grosso modo*, el del número de parámetros que hay que introducir en el modelo. Frecuentemente estos parámetros tienen un sentido práctico y son interpretables en el dominio considerado. Si una población es heterogénea (por ejemplo en genética) el problema puede ser saber cuántas poblaciones homogéneas la componen y, para cada población homogénea, saber cuál es su papel en la mezcla.

En este caso, los parámetros del modelo tienen un sentido biológico claro. Pero esto no es siempre así en econometría. La mayor parte del tiempo los parámetros no tienen otro sentido que el de medir correlaciones parciales entre variables dentro de un cuadro temporal. El modelo se construye tanto para predecir el futuro como para analizar el pasado. Esta banalidad no es únicamente matemática: si uno quiere tener un modelo que haga buenas predicciones, uno no puede ir simultáneamente muy lejos en la descripción del pasado utilizando un gran número de parámetros. Su estimación imprecisa implicaría entonces malas cualidades predictivas para el modelo. Hace falta, entonces, un principio heurístico de parsimonia, buscar un equilibrio entre la calidad en términos descriptivos y explicativos del ajuste y la calidad de la precisión obtenida con la ayuda del modelo.

Para hablar de equilibrio es necesario ponderar el interés que tenemos en cada una de las cualidades. De hecho, este punto de vista es raramente el adoptado por los estadísticos porque no es fácil, ni en la práctica ni en la teoría, de implementar... Es preferible pensar que en todos los casos existe un modelo verdadero que depende de un cierto número de parámetros. El objetivo es entonces obtener este modelo verdadero. Evidentemente, no hay respuesta exacta a menos que el tamaño de la muestra sea infinito de modo que podamos aplicar, de manera adecuada, la ley de los grandes números. Caemos entonces en los

*Université Paris-Sud, Orsay.

problemas de velocidad de la estimación y de tests que no han tenido hasta ahora sino soluciones parciales, sobre todo en el aspecto práctico. Además, parece razonable pensar que puntos de vista más amplios, más centrados sobre la noción de complejidad en el cuadro de la estadística no-paramétrica, permitirán aportar un esclarecimiento radicalmente diferente a la cuestión del orden de un modelo, si éste no está construido con ayuda de un número finito de parámetros impuestos, en genética de poblaciones, por las restricciones de la biología.

1 Modelos paramétricos

Sea $(P_\alpha^n)_{\alpha \in A}$ una familia de probabilidades sobre un mismo espacio (Ω, \mathcal{A}) , $n \in \mathbb{N}$. Para cada n , P_α^n define un modelo y para α fijo la familia $(P_\alpha^n)_{n \in \mathbb{N}}$ es coherente. El índice n será omitido a menos que sea necesario. Llamemos $A = \bigcup_{p \in P} A_p$ donde P es un conjunto finito o infinito con un orden parcial. Por convención $A_p \subset \mathbb{R}^{k(p)}$. El orden del modelo es entonces p . En todos los casos prácticos, el orden parcial es muy simple: $q > p$ equivale a $\mathbb{R}^{k(q)}$ sub-espacio de $\mathbb{R}^{k(p)}$: “ todo modelo de orden $q < p$ es un modelo de orden p ”.

Por razones prácticas el cero juega un rol particular.

Si denotamos por $0_{k(p)-k(q)}$ el cero de $\mathbb{R}^{k(p)} - \mathbb{R}^{k(q)}$ entonces si $q < p$, supondremos que existe una inyección $A_q \rightarrow A_p$ definida por $\theta_\alpha \rightarrow \tilde{\theta}_\alpha$ y tal que $(\theta_\alpha, 0_{k(p)-k(q)}) = \tilde{\theta}_\alpha$ para todo $\theta_\alpha \in A_q$.

Identificabilidad : Un modelo es identificable si $\alpha \neq \alpha' \Rightarrow P_\alpha \neq P_{\alpha'}$. En realidad esta noción puede ser restrictiva y mal adaptada. Diremos que un modelo es identificable en 0 si $\bar{\alpha} \neq \bar{\alpha}'$ implica $P_{\bar{\alpha}} \neq P_{\bar{\alpha}'}$, donde $\bar{\alpha}$ designa la clase de equivalencia de α definida por:

$$\theta_\alpha \sim \theta_{\alpha'} \text{ si existe } \bar{\theta}_\beta \neq 0 \text{ (en } \mathbb{R}^{k(q)}) \text{ tal que } \theta_\alpha = \bar{\theta}_\beta \text{ (en } \mathbb{R}^{k(p)}, p \geq q) \\ \text{ y } \theta_{\alpha'} = \theta_\beta \text{ (en } \mathbb{R}^{k(p')}, p' \geq q)$$

Módulo esta relación de equivalencia, la inyección $A_q \rightarrow A_p$ es una identidad para las probabilidades asociadas.

1.1 Ejemplos

1.1.1 Familias exponenciales de densidad f_α respecto a una medida D y observaciones independientes.

$$f_\alpha(x) = Z^{-1}(\alpha) \exp\left\{-\sum_{j=1}^p \alpha_j T_j(x)\right\}$$

El orden del modelo es p y $\theta = (\alpha_1, \dots, \alpha_p)$, $\alpha = (p, \theta)$. El orden parcial es simplemente el orden de los subconjuntos finitos de $A = (\alpha_1, \dots, \alpha_p, \dots)$, A

puede estar acotado o no. Z es la función de partición y T_1, \dots, T_p la familia de observables.

1.1.2 Regresión.

Familia de relaciones lineales. Una observación tomada de una sucesión independiente de pares de variables:

$$Y_{ij} = \sum_{i=1}^p \alpha_i X_i + \epsilon_{ij}$$

(ϵ_{ij}) un ruido blanco. El orden es p , número de factores aleatorios (o eventualmente controlados), $\theta = (\alpha_1, \dots, \alpha_p)$. El orden es el de los subconjuntos de $A = (\alpha_1, \dots, \alpha_p, \dots)$.

En esta dirección podríamos citar los modelos de Cox.

1.1.3 Mezclas de población.

Sea (G_λ) una familia de leyes de probabilidad $\lambda \in \Lambda \subset \mathbb{R}^l$. La observación es una sucesión de variables independientes de ley

$$Q_\alpha(dx) = a_1 G_{\lambda_1}(dx) + \dots + a_p G_{\lambda_p}(dx)$$

$a_1 + \dots + a_p = 1$, $0 < a_l \leq 1$ con $a_l < 1$ si $p > 1$, p es el orden de mezcla, $\theta = (a_1, \dots, a_p, \lambda_1, \dots, \lambda_p)$, $\alpha = (\theta, p)$.

Otra manera de escribir los parámetros es interesante: α se considera como medida puntual sobre Λ : $\alpha = \sum_{j=1}^p a_j \delta_{\lambda_j}$.

Las medidas puntuales operan sobre la familia (G_λ) por $\alpha(G) = \sum_{j=1}^p a_j G_{\lambda_j}$

El modelo es esencialmente no-identificable pero hay diversos tipos de no-identificabilidad.

Para comenzar las hay de tipo patológico, que pueden ser consideradas como bastante triviales. Por ejemplo, si (G_λ) es la familia de leyes de Bernoulli: $b(\lambda)$, entonces $\mu(G)$ es una ley de Bernoulli: $b(\sum_{j=1}^p a_j \lambda_j)$ y así distintas mezclas dan la misma ley.

El modelo es intrínsecamente no-identificable si la aplicación $\mu \rightarrow \mu(G)$ no es inyectiva. El estudio de este tipo de no-identificabilidad intrínseca se hace en [10] y [11]. Otra forma trivial de no-identificabilidad se debe a que la medida de mezcla es invariante por permutación de los parámetros. Pero el modelo de mezcla clásico no es identificable de manera no trivial. La medida $1\delta_{\lambda_0} + 0\delta_{\lambda_1}$ es igual para todos los (λ_1, M) a $M\delta_{\lambda_0} + (1 - M)\delta_{\lambda_0}$ y entonces $(M, \lambda_0, \lambda_0)$ equivale a $(0, \lambda_0, \lambda_1)$.

Un ejemplo cercano al de mezclas de población es el de las Cadenas de Markov Escondidas que presentan problemas de no-identificabilidad similares, pero más difíciles de tratar que los de mezclas de poblaciones.

1.1.4 Procesos ARMA Gaussianos

La observación (γ_n) se obtiene por una ecuación de recurrencia

$$\gamma_n - a_1\gamma_{n-1} - \dots - a_p\gamma_{n-p} = \epsilon_n - b_1\epsilon_{n-1} - \dots - b_q\epsilon_{n-q}$$

donde (ϵ_n) es una sucesión de variables gaussianas independientes, (a_1, \dots, a_p) , (b_1, b_2, \dots, b_q) son parámetros reales tales que los polinomios

$$P(z) = \sum_{j=1}^p a_j z^j \text{ y } Q(z) = \sum_{k=1}^q b_k z^k$$

tienen sus raíces fuera del círculo unitario. El orden del modelo es $(p, q) \in N^2$, N^2 esta dotado del orden parcial $(p, q) \leq (p', q') \Leftrightarrow p \leq p' \text{ y } q \leq q'$;

$$\theta = (a_1, \dots, a_p, b_1, \dots, b_q, \sigma^2) \text{ donde } \sigma^2 = E\epsilon^2.$$

Como en el caso de las mezclas, existe un parámetro más intrínseco, la densidad espectral f (que juega entonces el papel que jugaba la medida de mezcla) donde $f(\lambda) = \sigma^2 \left| \frac{Q(e^{-i\lambda})}{P(e^{-i\lambda})} \right|^2$.

El modelo no es identificable con los parámetros naturales P, Q, σ^2 porque si R es un polinomio cualquiera (PR, QR, σ^2) dan la misma densidad espectral y en consecuencia la misma probabilidad que (P, Q, σ^2) . De hecho, la no-identificabilidad se debe en buena parte al carácter gaussiano de ϵ ; volveremos a esto luego.

En lo que sigue, vamos a interesarnos esencialmente en el caso de modelos no trivialmente no-identificables. Consideraremos sobre todo el caso de mezclas de poblaciones y de procesos ARMA gaussianos, pero muchas ideas pueden extenderse a otros modelos que presentan una no-identificabilidad esencial. Vamos a estudiar sucesivamente tres situaciones: la estimación del orden por métodos principalmente algebraicos, los tests sobre el orden, test de verosimilitud y, finalmente, la estimación simultánea del orden y los parámetros por métodos de verosimilitud o de verosimilitud aproximada.

2 Estimación del orden de una mezcla de poblaciones y de una cadena de Markov escondida por métodos de momentos [1]

Comenzamos con una familia de leyes $(G_\gamma)_{\gamma \in \Gamma}$ sobre \mathbb{R}^p donde $\Gamma \subset \mathbb{R}^k$, $G_{\gamma_1} \neq G_{\gamma_2}$ para $\gamma_1 \neq \gamma_2$.

Una mezcla será una probabilidad $Q : Q = \sum_{i=1}^r \pi_i G_{\gamma_i}$, donde r es el orden de la mezcla y $\mu = \sum_{i=1}^r \pi_i \delta_{\gamma_i}$ es una probabilidad sobre Γ llamada medida de mezcla.

El problema es estimar r sin preocuparnos por el momento de estimar los parámetros (π_1, \dots, π_r) y $(\gamma_1, \dots, \gamma_r)$.

La idea utilizada aquí es la siguiente : los momentos de una medida cuyo soporte tiene exactamente r puntos tienen propiedades algebraicas particulares.

Nos interesa inicialmente el caso en el cual $\Gamma \subset \mathbb{R}$; sea $\Phi_p = (\gamma_d)_{1 \leq d \leq 2p}$ y $\mu(\Phi_p) = \int_{\Lambda} \Phi_p(\gamma) d\mu(\gamma)$ para toda medida μ sobre Γ ; sea $c^p = (c_1, \dots, c_{2p}) \in \mathbb{R}^{2p}$ y $H(c^p, p)$ la matriz de Hankel de orden p asociada a c^p , es decir la matriz $(p+1) \times (p+1)$ definida por $H(c^p, p)_{ij} = c_{i+j-2}^p$, $1 \leq i, j \leq p+1$ con $c_0^1 = 1$. Si μ tiene momentos de orden $2p$ finitos entonces la matriz de Hankel de μ es $H(\mu(\Phi_p), p)$ (ver [26], [27]).

Sea $K_p = \{c, c \in \mathbb{R}^{2p} \text{ tal que existe una medida positiva } \mu \text{ con } c = \mu(\Phi_p)\}$.

Pongamos $L(c, p) = \det H(c, p)$ para $c \in \mathbb{R}^{2p}$

Tenemos entonces

Proposición 1 : $L(c, p) = 0$ si y solamente si toda medida positiva tal que $\mu(\Phi_p) = c$ es discreta y tiene soporte con a lo sumo p puntos.

La demostración es inmediata (ver [27]).

Por lo tanto el orden r puede ser caracterizado por $r = \inf(p, L(c^p, p) = 0)$

Sean ahora las observaciones X_1, \dots, X_n de igual distribución marginal Q .

Supongamos que a partir de estas observaciones sabemos estimar $c^p = \mu(\Phi_p)$ por \hat{c}_n^p ; $L(\hat{c}_n^p, p)$ no se anulará aun si μ es de orden p a causa de la fluctuación estadística. Estamos en una situación clásica en estadística.

Llamaremos función de penalización a una función $\bar{A} : N \rightarrow \mathbb{R}^+$ tal que $\bar{A}(p, n)$ sea estrictamente creciente en p (para todo n) y tienda a cero cuando n tiende a infinito (para todo p). Lo usual es tomar $\bar{A}(p, n) = A(p)l(n)$. Ponemos entonces $J_n(p) = |L(C_n^p, \hat{p})| + A(p)l(n)$ y $\hat{r}_n = \arg \min_{p \in N} J_n(p)$

Tenemos entonces el siguiente teorema

Teorema 1 [1]

Supongamos que $\forall p \in N, \hat{c}_n^p \rightarrow c^p$ y $\frac{1}{l(n)}[L(\hat{c}_n^p, p) - L(c^p, p)] \rightarrow 0$ c.s. entonces $\hat{r}_n \rightarrow r$ c.s.

(El mismo resultado vale reemplazando c.s. por convergencia en probabilidad).

No en todos los problemas de penalización que estudiaremos (y esto tiene un sentido muy general) sabemos definir una función optimal $l(n)$. Argumentos de tipo Bayesiano o de complejidad llevan con frecuencia a escoger $A(p)l(n) = \frac{p \log n}{\sqrt{n}}$. Pero estos argumentos no son aún realmente convincentes y el recurso de la experimentación numérica es necesario a falta de un criterio de riesgo general que pueda ser aplicado a la estimación de \hat{r}_n para medir su calidad. El resultado siguiente precisa el error en r .

Supongamos que \hat{c}_n^p sea dado por una ley de grandes números del tipo

$$\hat{c}_n^p = f_p\left(\frac{1}{n} \sum_{i=1}^n \Phi_p(X_i)\right)$$

Tenemos entonces el siguiente teorema

Teorema 2 [1]

Supongamos que para todo $p \leq r$ la función $a \rightarrow L(f_p(a), p)$ es lipschitziana y que $\int \exp\{\langle t, \Phi_p(x) \rangle\} dQ(x)$ está definida en una vecindad de cero. Si además $\sqrt{nl}(n) \rightarrow \infty$, existe $d > 0$ y $n_0 \in N$ tal que para $n \geq n_0$,

$$P(\hat{r}_n \neq r) \leq \exp(-dnl^2(n)).$$

Este teorema exige en el caso estudiado aquí que Q tenga soporte acotado, pero veremos una extensión importante más adelante.

Demos ahora una aplicación importante por su generalidad. Supongamos que $dG_j(x) = dG(x - \gamma_j)$, donde G es una probabilidad conocida sobre \mathbb{R} .

Podemos entonces poner $X_n = Y_n + M_n$ donde (Y_n) y (M_n) son sucesiones de variables independientes entre ellas, y donde las (Y_n) son variables independientes de ley G y las (M_n) variables no necesariamente independientes de ley μ .

$$\text{Tenemos } Q(x^k) = \sum_{l=0}^k \frac{k!}{l!(k-l)!} G(x^l) \mu(x^{k-l})$$

Como $G(x^l)$ es conocida, obtenemos así un sistema triangular que permite estimar $\mu(x^k)$ a partir de la estimación de $Q(x^k)$, por ejemplo por un estimador empírico

$$\hat{Q}_n(x^k) = \frac{1}{n} \sum_{i=1}^n X_i^k \text{ si la ley de grandes números vale.}$$

Hipótesis sobre el muestreo: suponemos que observamos $X_n = Y_n + M_n$, donde M_n es un proceso ergódico que toma un número finito de valores, Y_n es una sucesión de variables aleatorias independientes, de ley conocida. (Y_n) y (M_n) son sucesiones independientes.

Proposición 2 [1]: Bajo estas condiciones de muestreo si $G(x^p) < \infty$ para todo p , el teorema 1 vale si $\sqrt{n}(\log \log n)^{-1}l(n) \rightarrow \infty$ para la convergencia c.s. y $\sqrt{n}l(n) \rightarrow \infty$ para la convergencia en probabilidad.

La proposición siguiente permite extender el resultado de \mathbb{R} a \mathbb{R}^s .

Proposición 3 [1]: Sea μ una probabilidad discreta sobre \mathbb{R}^s que tiene r puntos en su soporte. Para todo $v \in \mathbb{R}^s, \|v\| = 1$ llamemos $v\mu$ a la ley de $\langle v, \gamma \rangle, \gamma \in \mathbb{R}^s$ de ley μ . Entonces $v\mu$ es discreta con exactamente r puntos de soporte, salvo para a lo sumo $\frac{r(r-1)}{r}$ valores de v (para los cuales el soporte tiene menos de r puntos).

Tenemos $E[\langle v, \gamma \rangle^k] = \sum_{k_1+\dots+k_s=k} \frac{k!}{k_1! \dots k_s!} \prod_{i=1}^s v_i^{k_i} E \prod_{i=1}^s (\gamma^{k_i})$ y la proposición 2 lleva el caso \mathbb{R}^s al caso \mathbb{R} . Podemos entonces extender este resultado a los casos siguientes:

- 1- G depende de una varianza desconocida σ_0 , entonces $Q(dx) = \sum_{i=1}^r \pi_i dG(\frac{x-\gamma_i}{\sigma_0})$
- 2- El modelo tiene ruido $Z_n = X_n + \sigma_0 U_n$, entonces $Z_n = M_n + Y_n + \sigma_0 U_n$.

Tendremos un modelo de mezcla si (M_n) es una sucesión de variables independientes y un modelo más general si M_n es un proceso ergódico con marginal discreta.

Para realizar esta extensión, necesitamos una hipótesis bien sobre G , bien sobre el ruido.

Definición: Llamaremos \mathcal{L} a la clase de leyes de probabilidad G tales que si U es una variable de ley G , para todo $c \in]0, 1[$ existen dos variables independientes U y V_c , U de ley G con $U = cU + V_c$.

Esta clase ha sido caracterizada como la clase de leyes infinitamente divisibles de densidad de Levy f , tal que $xf(x)$ es decreciente en \mathbb{R}^+ y \mathbb{R}^- y la parte gaussiana es arbitraria.

Observamos que los ruidos con valores discretos están excluidos.

Si aplicamos la relación precedente para estimar $\mu(\gamma^k)$ a partir de $\widehat{Q}_n(x^l)$, obtenemos evidentemente $\mu(\gamma^k)$ que dependen de σ desconocido a partir de

$$(*) Q(x^k) = \sum_{l=0}^k \frac{k! \theta^{-l}}{l!(k-l)!} G(x^l) \mu(\theta^{k-l}, \sigma)$$

Observamos que $H(\sigma, p)$ es la matriz de Hankel construida con los pseudomomentos $\mu(\theta^i, \sigma)$.

Teorema 3 [1]

Si G (o U) pertenece a L entonces $\forall \sigma < \sigma_0, \forall p < r, \det H(\sigma, p) > 0$ para $\sigma = \sigma_0, \det H(\sigma, p) > 0$ si y solamente si $p > r$.

Pongamos entonces

$$K_n(p, \sigma) = \left| \det \hat{H}_n(\sigma, p) \right| + A(p)l_1(n) + \sigma l_2(n)$$

donde $l_1(n)$ y $l_2(n)$ son funciones positivas que tienden a cero en el infinito y $\hat{H}_n(\sigma, p)$ es la matriz de Hankel construida a partir de soluciones de (*) cuando reemplazamos $Q(x^k)$ por $\hat{Q}_n(x^k)$.

Definimos entonces \hat{r}_n y $\hat{\sigma}_n$ por:

$$K_n(\hat{r}_n, \hat{\sigma}_n) = \min \{K(p, \sigma), p \in N, \sigma \in S\} \text{ (} S \text{ acotado)}$$

Proposición 4 [1]: El teorema 1 es válido bajo estas condiciones para \hat{r}_n . Además $\hat{\sigma}_n$ converge a σ_0 bajo las mismas condiciones que las impuestas en el teorema 1.

Este tipo de razonamiento puede extenderse en muchas direcciones.

- 1- Podemos reemplazar las matrices de Hankel por las matrices de Toeplitz, con exponenciales en lugar de potencias. En este caso los momentos están acotados y uno puede siempre obtener las desigualdades exponenciales indicadas en el teorema 2.
- 2- Podemos utilizar otros criterios distintos de los criterios de Hankel y Toeplitz [24].
- 3- Podemos generalizar estos resultados a mezclas sobre la varianza, la covarianza o el parámetro de una familia exponencial en el sentido estadístico del término [1].
- 4- Podemos desarrollar ideas similares a propósito del problema, fundamental en teoría de señal de la separación de fuentes, cuando la señal es de tipo discreto [28] [29].
- 5- Existen métodos bastante similares para identificar el orden de los procesos ARMA, en ese caso las matrices tipo Hankel utilizan las autocorrelaciones y las autocorrelaciones parciales ([14],[25]).

3 Test para el orden de un modelo

Después de haber determinado de manera algo imprecisa el orden de un modelo, podemos pasar a una etapa más fina que consiste en hacer un test del orden. Para esto, buscamos utilizar un proceso que en los casos identificables usuales es asintóticamente optimal en sentidos diversos, que son los de los tests del cociente de verosimilitudes. Pero la no-identificabilidad hace que las cosas sean bastantes complejas.

Examinemos el caso simple de una mezcla de a lo sumo 2 poblaciones. Queremos probar que la muestra representa una población contra la hipótesis de que la muestra está formada por 2 poblaciones. Sea f_0 la densidad bajo la hipótesis nula y $\pi f_{\gamma_1} + (1 - \pi)f_{\gamma_2}$ bajo la hipótesis alternativa $0 < \pi < 1$, $\gamma_1 \neq \gamma_2$, $\gamma_1, \gamma_2 \in \Gamma$, Γ compacto.

Podemos suponer que $\gamma_2 < \gamma_1$ por ejemplo, y estudiar la singularidad en 0, que hace que la información de Fisher sea degenerada separando dos casos, de manera de tener una heurística (ver sobre este tópico [2]).

a) El caso $\pi \rightarrow 0$, $\gamma_2 \rightarrow 0$, γ_1 fijo para el desarrollo del logaritmo de la verosimilitud.

$$\text{Pongamos } \bar{f}_\gamma = \frac{f_\gamma - f_0}{f_0} \text{ y } \dot{f}_0 = \frac{f'_0}{f_0}, \ddot{f}_0 = \frac{f''_0}{f_0}$$

donde f' es la derivada respecto a γ ; tenemos como logaritmo del cociente de verosimilitud para una observación (X_1, \dots, X_n)

$$T_n = \sum_{i=1}^n \log(1 + \pi \bar{f}_{\gamma_1}(X_i) + (1 - \pi) \bar{f}_{\gamma_2}(X_i))$$

de donde un desarrollo del tipo

$$\sum_{i=1}^n (\pi \bar{f}_{\gamma_1}(X_i) + (1 - \pi) \gamma_2 \dot{f}_0(X_i)) \text{ al 1er orden. Si aplicamos sin cuidado}$$

la teoría clásica de la verosimilitud, utilizaríamos el desarrollo del log al 2do orden, el modelo será regular en (π_1, γ_2) y con γ fijo, como $\pi > 0$, el máximo de verosimilitud será de tipo $G_{\gamma_1}^2 1_{(G_{\gamma_1} > 0)}$ donde G_γ es un proceso gaussiano centrado de covarianza regular. Pero este tipo de resultado es inexacto si no tenemos mayor precaución. El estadístico $\sup_{\gamma_1} G_{\gamma_1}^2 1_{(G_{\gamma_1} > 0)}$ no toma en cuenta

el hecho de que γ_1 puede variar con n (se trata de hacer un test que tome en cuenta todos los valores de γ_1 en cada etapa). γ_1 no es entonces un parámetro de ruido. Y si desarrollamos los cálculos veremos que si γ_1 tiende a cero muy rápidamente, entonces no es posible detener el desarrollo en el 2do orden y es necesario un análisis más fino.

b) En el 2do caso, donde π está fijo, γ_1 y γ_2 tienden a cero. Como tenemos 2 parámetros podemos tomar como desarrollo al 1er orden

$(\pi \gamma_1 + (1 - \pi) \gamma_2) \dot{f}_0 + \frac{1}{2}(\pi \gamma_1^2 + (1 - \pi) \gamma_2^2) \ddot{f}_0$ considerando como nuevo parámetro $\pi \gamma_1 + (1 - \pi) \gamma_2$ y $\frac{1}{2}(\pi \gamma_1^2 + (1 - \pi) \gamma_2^2)$ y el modelo es regular para π fijo.

Pero aun aquí es necesario ser cuidadoso porque es indispensable derivar al 2do orden para aplicar la teoría clásica y el hecho de que la nueva parametrización no sea diferenciable presenta problemas. Este tipo de análisis que encontramos en los otros casos simples de modelos no-identificables nos ha llevado a proponer otro enfoque, mucho más unificador de este tipo de problema. Es el fundamentado sobre los modelos localmente cónicos.

4 Modelos localmente cónicos [3]

Sea (X_1, \dots, X_n) una muestra de ley $P_0^{(n)}, P_0^{(n)} \in \mathcal{P}_n$, conjunto dominado por una medida positiva $D^{(n)}$.

Las condiciones siguientes definen un modelo estadístico localmente cónico. 1- Suponemos que \mathcal{P}_n se parametriza a través de dos parámetros θ y β , $(\theta, \beta) \in [O, M] \times \mathcal{B}$ de modo que $\mathcal{P}_n = \{P_{\theta, \beta}^{(n)}, (\theta, \beta) \in \mathcal{T}\}$, $[O, M] \times \mathcal{B}$ con la topología

producto y \mathcal{T} con clausura $\overline{\mathcal{T}}$ compacta.

2- Para $\theta = 0$, el modelo puede ser no-identificable en β , pero lo es en θ para $\theta = 0$, es decir en particular que $P_{\theta, \beta}^{(n)} = P_0^{(n)} \Leftrightarrow \theta = 0$

Para cada $c > 0$, definimos $\mathcal{B}^c = \{\beta \in \mathcal{B}, \exists \theta \leq c, (\theta, \beta) \in \overline{\mathcal{T}}\}$ y ponemos $\theta_\beta = \sup \{t, [0, t] \times \{\beta\} \subset \overline{\mathcal{T}}\}$.

Suponemos además que:

$\forall \beta \in \mathcal{B}$, o bien $\theta_\beta > 0$ o $[0, c] \times \beta \cap \overline{\mathcal{T}}$ es vacío para un $c > 0$.

En todas las aplicaciones, trabajaremos con $\tilde{\mathcal{B}} = \bigcap_{c>0} \mathcal{B}^c$ y podremos suponer $\tilde{\mathcal{B}} = \mathcal{B}$ sin restringir la generalidad del modelo. β podemos pensarla como una dirección en la cual nos acercamos a 0, cuando $\theta \rightarrow 0$.

Ahora daremos ejemplos de aplicación de la noción de modelo localmente cónico y de tests de verosimilitud que podemos aplicarles.

En lo que sigue \mathcal{T}_0 y \mathcal{T}_1 serán sub-conjuntos de \mathcal{T} .

Contrastaremos $(\theta, \beta) \in \mathcal{T}_0$ contra $(\theta, \beta) \in \mathcal{T}_1$. Pondremos para $i = 0$ ó 1 ,

$$T_n(i) = \sup_{(\theta, \beta) \in \mathcal{T}_i} L_n(\theta, \beta)$$

donde L_n es el logaritmo de la verosimilitud cuando el modelo tiene como densidad $g_{\theta, \beta}^{(n)}$ y el estadístico de test será $T_n(i) - T_n(0)$, la hipótesis $H_0 : (\theta, \beta) \in \mathcal{T}_0$ será rechazada al nivel α cuando $T_n(i) - T_n(0) > C_\alpha$.

Ejemplos de modelos paramétricos identificables

Consideremos una muestra de v.a.i. X_1, \dots, X_n de ley con densidad $g_0 = g_{\gamma_0}$. El modelo regular (en sentido local) es decir $t \rightarrow g_{\gamma_0 + t u}$ es de clase C^2 para $t > 0$, g_0 c.s. en todas las direcciones u tales que $u \in U(\gamma^0) = \{u \in \mathbb{R}^p, T(\gamma, u)$ contiene un intervalo $[0, u_\gamma], u_\gamma \neq 0\}$ con $T(\gamma, u) = \{t \in \mathbb{R}, \gamma + t u \in \mathcal{T}\}$

Suponemos además que existen funciones h, l_1, l_2 tales que para todo $\gamma = \gamma^0 + t u$, $h \in U(\gamma^0)$ tenemos:

$$|\log g_\gamma| \leq h, \left| \frac{1}{g_\gamma} \frac{\partial^i g_\gamma^0 + t u}{\partial t^i} \right| \leq l_i, \quad i = 1, 2$$

con $h \in L^1(g_0 \nu)$, $l_1 \in L^2(g_0 \nu)$, $l_2 \in L^1(g_0 \nu)$

La parametrización es entonces $\gamma = \gamma^0 + \theta \beta$ donde $\theta = \sqrt{(\gamma - \gamma^0)^T I(\gamma^0) (\gamma - \gamma^0)}$, $\beta = \frac{\gamma - \gamma^0}{\theta}$ y $I(\gamma^0)$ es la matriz de información de Fischer: \mathcal{D} es el espacio de gradientes,

$$\mathcal{D} = \overline{\mathcal{D}} = \left\{ d = \frac{1}{g_0} \sum_{i=1}^p b_i d_i, \text{ donde } d_i = \left(\frac{\partial g_{\theta, \beta_i}}{\partial \theta} \right)_{\theta=0} \text{ y } \sum_{i=1}^p b_i \beta_i \in \mathcal{B} \right\}$$

$\beta_1 \dots \beta_p$ son p direcciones linealmente independientes.

Teorema 5: Bajo las condiciones precedentes $\sup_{\gamma \in \Gamma} L_n(\gamma) - L_n(0)$ converge en ley a

$$\frac{1}{2} \sup_{u \in I(\gamma_0)^{\frac{1}{2}} \mathcal{B}} (\langle u, V \rangle)^2 1_{\langle u, V \rangle \geq 0}$$

donde V es un vector gaussiano estándar de dimensión p .

Este teorema permite generalizar el resultado clásico sobre χ^2 . La ley límite es $\chi^2(p)$ si γ^0 está en el interior de Γ y si \mathcal{B} y $I(\gamma^0)^{\frac{1}{2}} \mathcal{B}$ contienen todas las direcciones.

En los otros casos, en particular si γ^0 es un punto frontera de Γ , la ley límite se escribe utilizando la proyección $\bar{P} = P_{I(\gamma^0)^{\frac{1}{2}} \mathcal{B}}$ como $\frac{1}{2} (\langle V, \bar{P}V \rangle)^2 1_{\langle V, \bar{P}V \rangle \geq 0}$.

Este resultado puede extenderse al caso donde existen parámetros fantasmas.

Ejemplos de modelos de perturbación [2]

Tenemos un ejemplo muy simple de modelo no-identificable. El conjunto de densidades que corresponden a \mathcal{T} está dado por:

$$\mathcal{G} = \{g_{\pi, \gamma} = (1 - \pi)f_{\gamma_0} + \pi f_{\gamma}\}$$

$g_0 = f_{\gamma_0}$ es conocido y γ_0 está en el interior de Γ , conjunto compacto de $\mathbb{R}^p, \gamma \in \Gamma$

Sea $H = L^2(g_0\theta)$ y definamos (θ, β) por

$$\theta = \left\| \frac{g_{\pi, \gamma} - g_0}{g_0} \right\|_H = \pi \left\| \frac{f_{\gamma} - f_{\gamma_0}}{f_{\gamma_0}} \right\|_H \quad \text{y } \beta = \gamma.$$

Sea $\bar{f}_{\beta} = \frac{f_{\beta} - f_0}{f_0}$. Tenemos entonces

$$g_{\theta, \beta} = g_0 \left(1 + \theta \frac{\bar{f}_{\beta}}{\|\bar{f}_{\beta}\|_H} \right)$$

Obtenemos un modelo lineal en θ con β fijo. Podemos aplicar la teoría ordinaria de máximo de verosimilitud si f_{γ} es C_1 en el interior de Γ y si

$$\forall \gamma \in \Gamma, \left| \frac{1}{f_{\gamma}} \frac{\partial f_{\gamma}}{\partial f_i} \right| \leq l, i = 1, \dots, p$$

con $l \in H$.

$$\text{Ponemos } \mathcal{D} = \left\{ \frac{\bar{f}_{\beta}}{\|\bar{f}_{\beta}\|_H}, \beta \in \Gamma \right\}.$$

Teorema 6 [2]

Si \mathcal{D} es una clase de Donsker [22] y si el proceso canónico gaussiano (ξ_d) sobre \mathcal{D} tiene trayectorias continuas, entonces la ley límite del cociente del logaritmo del cociente de verosimilitudes es la ley de $\sup_{d \in \mathcal{D}} \xi_d^2 1_{\xi_d > 0}$.

Recordemos que el proceso canónico ξ_d es el proceso gaussiano centrado, normalizado tal que $E(\xi_{d_1}, \xi_{d_2}) = \langle d_1, d_2 \rangle_H$.

Este tipo de resultados se extiende a modelos no-paramétricos de perturbación del tipo: contrastar $H_0 : \{g_0\}$ contra $H_1 : \{g_{\theta, \beta} = g_0 + \theta\beta\}$ o $H'_1 : \{(1 - \epsilon)g_0 + \epsilon g_1\}$ cuando β está en un conjunto funcional que tiene las propiedades de Donsker adecuadas, por ejemplo

$$\mathcal{B} = \left\{ \beta \in H^4 \cap H^2 \cap L^2\left(\frac{1}{g_0 \delta}\right), \|\beta\|_{L^2} = 1, \|\beta\|_{H^4} \leq K_1, \|\beta\|_{H^2} \leq K_2, \beta(0) = 0 \right\}$$

donde H^p es el espacio de Sobolev con la norma $\sum_{\gamma=0}^p \|f^{(\gamma)}\|_2$ [2].

Ejemplo del test de una población contra dos poblaciones [2]

Hemos indicado en la introducción las dificultades que un enfoque usual produce. Vamos entonces a comenzar por detallar una parametrización localmente cónica.

Sea $\Gamma \subset \mathbb{R}$, Γ compacto, $g_0 = f_{\gamma_0}$ y $g = \{\pi f_{\gamma_1} + (1 - \pi) f_{\gamma_2}\}$

Ponemos $\beta = (\gamma, \delta)$, $\gamma \in \Gamma$, $\delta \in [0, M]$ y $g_{\theta, \beta} = \frac{\theta}{N(\beta)} f_{\gamma} + (1 - \frac{\theta}{N(\beta)}) f_{\gamma_0 + \frac{\theta}{N(\beta)} \delta}$

con $N(\beta) = \left\| \delta \frac{1}{g_0} \frac{\partial f_{\gamma}}{\partial \gamma} 1_{\gamma=\gamma_0} + \frac{f_{\gamma} - g_0}{f_0} \right\|_H$ donde H es $L^2(g_0 \nu)$

Suponemos que

1) $N(\beta) = 0$ si y solamente si $\gamma = \gamma_0$ y $\delta = 0$ (propiedad de independencia lineal de las funciones $\frac{f_{\gamma} - g_0}{g_0}$ y de sus límites)

2) $f_{\gamma} \in C^5$ y $\sup_{|\gamma - \gamma_0| < \epsilon} \left| \frac{f_{\gamma}^i}{g_0} \right| \in H$ para un $\epsilon > 0$ y $i \in [2, 5]$,

$$\mathcal{D} = \left\{ d, d = \frac{1}{N(\beta)} \frac{f_{\gamma} - f_{\gamma_0} + \delta f_{\gamma_0}'}{f_{\gamma_0}} \right\},$$

distinguiamos además $d_i = \frac{\frac{f_{\gamma_0}^{(i)}}{f_{\gamma_0}}}{\left\| \frac{f_{\gamma_0}^{(i)}}{f_{\gamma_0}} \right\|_H}$, $i = 1, 2$ y $u = \langle d_1, d_2 \rangle_H$

Teorema 7 [2]

Bajo las hipótesis precedentes y si \mathcal{D} es una clase de Donsker, si ξ_d es un proceso canónico con trayectorias continuas entonces el estadístico de test converge

en ley a $\sup_{d \in \mathcal{D}} \left\{ \frac{1}{2} \sup_{d \in \mathcal{D}} \xi_d^2 1_{\xi_d \geq 0}, \frac{1}{2} \xi_{d_1}^2 + \frac{1}{2} \xi_{d_{2-u d_{1/\sqrt{1-u^2}}}}^2 1_{d_{2-u d_{1/\sqrt{1-u^2}}} > 0} \right\}$

Comentario

La demostración consiste en utilizar un desarrollo de Taylor en θ con β fijo. Pero este desarrollo no puede detenerse en el segundo orden, es necesario desarrollar hasta el orden 5 y la parte útil sobre la cual se hace la optimización es entonces un polinomio en $\frac{\theta}{N(\beta)}$ de grado 4. Esto refleja la no-identificabilidad y parece ser un fenómeno bastante general que se encuentra en otros modelos.

El desarrollo se detiene gracias a un control de los parámetros, que también es de tipo mucho más general que el caso de mezclas, y que se traduce aquí por: existen constantes A y B tales que si $\Phi = \frac{\delta}{N(\beta)}$, entonces $\delta \leq \frac{A}{\Phi^{\frac{1}{3}}}$, $N(\beta) = \frac{B}{\Phi^{\frac{2}{3}}}$.

Este tipo de control es posiblemente uno de los puntos más interesantes del modelo localmente cónico. En la presentación heurística de este modelo hecha en la introducción, hemos señalado la dificultad importante que había para separar el tratamiento en dos casos que corresponden a dos sub-modelos. Esta separación se obtiene utilizando de manera bastante complicada el control de los parámetros señalado anteriormente.

5 Prueba de una mezcla de q poblaciones contra una mezcla de p poblaciones

Mantenemos las mismas notaciones $(f_\gamma)_{\gamma \in \Gamma}$ donde Γ , compacto en \mathbb{R}^k es el conjunto de las densidades de probabilidad posibles.

$$g_0 = \sum_{l=1}^q \pi_l^0 f_{\gamma^l, 0}$$

La alternativa es

$$g_p = \{g_{\pi, \alpha} = \sum_{i=1}^p \pi_i f_{\gamma^i}\}$$

que suponemos identificables en sentido débil, es decir que la medida $\mu = \sum_{i=1}^p \pi_i \delta_{\gamma^i}$ define una mezcla única. La parametrización localmente cónica se define de la manera siguiente:

a) $\mathcal{B} = (\lambda_1, \dots, \lambda_{p-q}, \gamma^1, \dots, \gamma^{p-q}, \delta^1, \dots, \delta^q, \rho_1, \dots, \rho_q)$ tales que

$$\lambda_i \geq 0, \gamma^i \in \Gamma, i = 1, \dots, p-q, \delta^l \in \mathbb{R}^k, \rho_l \in \mathbb{R}, l = 1, \dots, q$$

y

$$\sum_{i=1}^{p-q} \gamma_i + \sum_{l=1}^q \rho_l = 0 \text{ y } \sum_{i=1}^{p-q} \lambda_i^2 + \sum_{i=1}^q \rho_i^2 + \sum_{i=1}^q \|\delta^i\|^2 = 1$$

b) $H = L^2(g_0 \nu)$ y

$$N(\beta) = \left\| \sum_{l=1}^q \pi_l^0 \sum_{i=1}^k \delta_i^l \frac{1}{g_0} \frac{\delta f_\gamma}{\delta \gamma^i} \Big|_{\gamma=\gamma^l, 0} + \sum_{i=1}^{p-q} \gamma_i \frac{f_{\gamma^i}}{g_0} + \sum_{l=1}^q \rho_l \frac{f_{\gamma^l, 0}}{g_0} \right\|_H$$

entonces

$$g(\theta, \beta) = \sum_{i=1}^{p-q} \gamma_i \frac{\theta}{N(\beta)} f_\gamma^i + \sum_{l=1}^q \pi_l^0 = \rho_l \frac{\theta}{N(\beta)} f_{\gamma^l, 0 + \frac{\theta}{N(\beta)} \delta^l}$$

con

$$\pi_l^0 + \rho_l \frac{\theta}{N(\beta)} \geq 0$$

Esta parametrización puede parecer compleja pero es local y adaptada al problema del test: perturbamos de un lado los parámetros de g_0 y añadimos una perturbación independiente que es mezcla de otras $p - q$ densidades, mezcla que va a tender a cero. Una dificultad técnica consiste en escoger una buena permutación de los parámetros para asegurar la identificabilidad en $\theta = 0$. El algoritmo de selección es el siguiente: calculamos

$$\min_{l=1\dots q \ i=1\dots p} \|\gamma^{l,0} - \gamma^i\|,$$

sean l_1 e i_1 los puntos donde este mínimo se alcanza; definimos entonces la perturbación Γ por $\sigma(p - q + l_1) = i_1$ e iteramos eliminando los puntos l_1 e i_1 . Completamos así la permutación. La definición conduce a la condición

$$\frac{\theta}{N(\beta)} \leq \rho + 2q \sup_{\gamma \in \Gamma} \|\gamma\|^2$$

y entonces $N(\theta) \rightarrow 0 \Rightarrow \theta = 0$. El espacio de los gradientes $\mathcal{D} \subset S(H)$, esfera unitaria de H , es entonces el espacio de

$$\frac{1}{N(\beta)} \left(\sum_{l=1}^q \pi_l^0 \sum_{i=1}^k \delta_i^l \frac{D_i^l f_{\gamma^{l,0}}}{g_0} + \sum_{l=1}^{p-q} \gamma_l \frac{f_{\gamma^l}}{g_0} + \sum_{l=1}^q \rho_l \frac{f_{\gamma^{l,0}}}{g_0} \right)$$

con $\beta \in B$; ξ_d es el proceso gaussiano común en una clase de Donsker \mathcal{D} [2] (para verificarlo en todos los casos clásicos, se utiliza un cálculo de entropía o de entropía con corchetes, que es fácil porque \mathcal{D} es de tipo paramétrico finito). Suponemos que los conjuntos de funciones

$$\left(\left(\frac{f_{\gamma^l}}{g_0} \right)_{l=1,\dots,p_1}, \left(\frac{f_{\gamma^{l,0}}}{g_0} \right)_{l=1,\dots,q}, \left(\frac{D_i^l f_{\gamma^{l,0}}}{g_0} \right)_{l=1,\dots,q, i=1,\dots,k}, \left(\frac{D_{ij}^2 f_{\gamma^{l,0}}}{g_0} \right)_{\substack{l=1,\dots,p_2 \\ i,j=1,\dots,k}} \right)$$

son linealmente independientes en H para toda selección de $p_1, p_2, p_1 + p_2 \leq p - q$ y γ^l distintos de $\gamma^{l,0}$. Ponemos:

$$\begin{aligned} \mathcal{D}_1 &= \left\{ \sum_{l=1}^q \sum_{i=1}^k \lambda_{l,k} D_i^1 \frac{f_{\gamma^{l,0}}}{g_0} \right\} \\ \mathcal{D}_2 &= \left\{ \sum_{l=1}^{p_1} \mu_l \frac{f_{\gamma^l}}{g_0} + \sum_{l=1}^q \tilde{\rho}_l \frac{f_{\gamma^{l,0}}}{g_0} + \sum_{l=1}^q \lambda_{l,1} \frac{D_i^1 f_{\gamma^{l,0}}}{g_0} \right. \\ &\quad \left. + \sum_{l=1}^{p_2} \sum_{i,j=1}^k \tau_l a_i a_j \frac{D_{ij}^2 f_{\gamma^{l,0}}}{g_0} \right\} \end{aligned}$$

con $p < p - q - 1$, $p_1 + p_2 \leq p - q$, $\mu_l > 0$, $\sum_{l=1}^q \mu_l + \sum_{l=1}^q \tilde{\rho}_l = 0$, $\tau_l \geq 0$. Entonces las funciones de \mathcal{D}_1 y \mathcal{D}_2 son ortogonales.

Teorema 8 Bajo estas hipótesis y las hipótesis de dominación introducidas en el teorema 7, el estadístico de test converge en ley a

$$\sup \left\{ \sup_{d \in \mathcal{D}} \frac{1}{2} \xi_d^2 1_{\{\xi_d \geq 0\}} ; \sup_{d_1 \in \mathcal{D}_1, d_2 \in \mathcal{D}_2} \frac{1}{2} (\xi_{d_1}^2 + 1_{\{\xi_{d_2}^2 \geq 0\}}) \right\}$$

6 Test del orden de un proceso ARMA[3]

Sea un proceso X_n estacionario, gaussiano, centrado, dado por la ecuación de recurrencia

$$X_n + a_1 X_{n-1} + \dots + a_p X_{n-p} = \varepsilon_n + b_1 \varepsilon_{n-1} + \dots + b_q \varepsilon_{n-q}$$

donde ε_n es el ruido de innovación y donde los polinomios $P(z)$ y $Q(z)$, $P(z) = 1 + \sum_{\gamma=1}^p a_\gamma z^\gamma$, $Q(z) = 1 + \sum_{\gamma=1}^q b_\gamma z^\gamma$ no tienen raíces en el círculo unitario cerrado.

La densidad espectral es $f(x) = \frac{\sigma^2}{2\pi} \left| \frac{Q}{P}(e^{-ix}) \right|^2$. Suponemos que la verdadera densidad espectral está asociada a (σ_0^2, Q_0, P_0) . Llamamos (p, q) al orden del modelo y queremos contrastar (p_0, q_0) contra (p, q) , donde (p_0, q_0) son los grados de (P_0, Q_0) y $p_0 \leq p, q_0 \leq q$. El modelo no es identificable (cf [14]).

Hannan y Veres han resuelto el problema de contrastes en casos particulares [18], [20] donde $(p = p_0 + 1, q = q_0 + 1)$.

Llamamos $\frac{1}{u_i}$, $i = 1 \dots q_0$ y $\frac{1}{t_i}$, $i = 1 \dots p_0$ los ceros y los polos de f_0 , que suponemos distintos para simplificar. Ponemos $r = \min(p - p_0, q - q_0)$ y suponemos por ejemplo que $r = p - p_0$ entonces $s = q - q_0 - r$.

Definimos entonces la representación localmente cónica de las densidades espectrales de la manera siguiente. Ponemos

$$\beta = (\delta, \mu, \tau, c, \gamma, \nu)$$

donde $\delta \in \mathbb{R}, \mu \in \mathbb{C}^{q_0}, \tau \in \mathbb{C}^{q_0}, c \in \mathbb{C}^r, \gamma \in \mathbb{C}^r, \nu \in \mathbb{C}^s$ y $\|\mu\|^2 + \|\tau\|^2 + \|\gamma\|^2 + \|\nu\|^2 = 1$ y además

$$\begin{aligned} f(\theta, \beta)(e_{ix}) &= \left(\frac{\sigma_0^2 + \frac{\theta}{N(\beta)} \delta}{2\pi} \right) \left| \frac{\prod_{i=1}^{q_0} (1 - (u_i + \frac{\theta}{N(\beta)} \mu_i) z)}{\prod_{i=1}^{p_0} (1 - (t_i + \frac{\theta}{N(\beta)} t_i) z)} \right|^2 \\ &\times \left| \prod_{i=1}^r \frac{(1 - (c_i + \epsilon_i \frac{\theta}{N(\beta)} \gamma_i) z)}{(1 - (c_i + (1 - \epsilon_i) \frac{\theta}{N(\beta)} \gamma_i) z)} \right|^2 \left| \prod_{i=1}^s (1 - (\frac{\theta}{N(\beta)} \nu_i) z) \right|^2 \end{aligned}$$

donde $z = e^{ix}$, $\epsilon_i = 1$ si $|c_i - t_i| < |c_i - u_i|$ y $\epsilon_i = 0$ si $|c_i - t_i| \geq |c_i - u_i|$, $i = 1 \dots r$ y donde

$$N(\beta) = \left\| \frac{\delta}{\sigma_0^2} + \sum_{i=1}^{p_0} \left(\frac{\tau_i z}{1 - \tau_i z} + \frac{\bar{\tau}_i}{z - \bar{\tau}_i} \right) - \sum_{i=1}^{q_0} \left(\frac{\mu_i z}{1 - \nu_i z} + \frac{\bar{\mu}_i}{z - \bar{\nu}_i} \right) + \sum_{i=1}^r \left(\frac{(1 - 2\epsilon_i - \gamma_i z)}{1 - c_i z} + \frac{(1 - 2\epsilon_i)\bar{\gamma}_i}{z - \bar{c}_i} - \sum_{i=1}^s \left(\nu_i z + \frac{\bar{\nu}_i}{z} \right) \right) \right\|_H$$

con $H = L^2\left((0, 2\pi), \frac{dx}{2\pi}\right)$. Seleccionamos como en el párrafo precedente una permutación de los polos y de los ceros de $f_{\theta, \beta}$ de manera que $f_{\theta, \beta} = 0 \Leftrightarrow \theta = 0$ (con un algoritmo similar).

Tenemos $\frac{\theta}{N(\beta)} \leq 2(p_0 + q_0 + r + s)$ y entonces $N(\beta) \rightarrow 0 \Rightarrow \theta \rightarrow 0$. Definimos entonces como antes el espacio de gradientes $\mathcal{D} \subset S(H)$ y los espacios \mathcal{D}_1 y \mathcal{D}_2 , no daremos aquí los detalles (cf [3]) donde \mathcal{D}_1 y \mathcal{D}_2 son tales que $\bar{\mathcal{D}} - \mathcal{D} = \mathcal{D}_1 \oplus \mathcal{D}_2$. Nos interesa ahora el test de verosimilitudes aproximadas, más simple de implementar que el test de verosimilitudes, donde la verosimilitud aproximada de Whittle [14] está dada por su logaritmo:

$$\begin{aligned} C_n(f) &= n \log 2\pi + {}^T X^{(n)}. T_n\left(\frac{1}{f}\right). X^{(n)} + n \log \sigma^2 \\ &= n \log 2\pi + n \log \sigma^2 + I_n\left(\frac{1}{f}\right) \end{aligned}$$

$$C_n(f) = n \log 2\pi + n \log \sigma^2 + I_n\left(\frac{1}{f}\right)$$

donde

$$I_n(v) = \int_{-\pi}^{\pi} v(e^{ix}) \left| \sum_{k=1}^n X_k e^{ikx} \right|^2 \frac{dx}{2\pi}$$

Teorema 9

El estadístico del test de verosimilitud aproximada converge a

$$\sup \left\{ \sup_{d \in \mathcal{D}} \frac{1}{2} \xi_d^2 1_{\{\xi_d \geq 0\}}; \sup_{d_1 \in \mathcal{D}_1, d_2 \in \mathcal{D}_2} \frac{1}{2} (\xi_{d_1}^2 + \xi_{d_2}^2 1_{\{\xi_{d_2} \geq 0\}}) \right\}$$

cuando $p = p_0 + 1$, $q = q_0 + 1$, caso particular tratado por Hannan ($p_0 = q_0 = 0$) y Veres, la ley límite se reduce a

$$\sup_{d \in \mathcal{D}} \frac{1}{2} \xi_d^2 1_{\xi_d \geq 0}.$$

7 Estimación por máximo de verosimilitud del orden y de los parámetros del modelo.

Curiosamente es la parte clásicamente más desarrollada de la teoría. Las dificultades debidas a la degeneración de la información de Fisher que encontramos en el problema del test no existen y el problema es, en alguna medida, exactamente el mismo con grandes dificultades de cálculo e implementación que el problema tratado en **1**. Ajustar el modelo lo mejor posible nos lleva a tomar un orden tan elevado como sea posible. Debemos entonces recurrir a un método de penalización y entonces utilizar como función de objetivo para determinar el orden y los parámetros.

$$U_n(r, \alpha) - A(r)\ell(n)$$

donde $U_n(r, \alpha)$ es la log-verosimilitud del modelo de orden r para un valor α de los parámetros. Para una función de penalización $A(r)\ell(n)$ escogida como en la primera parte se obtiene un teorema de consistencia para \hat{r}_n definida por

$$\hat{r}_n = \max_p (U_n(p, \hat{\alpha}_n) - A(p)\ell(n))$$

donde $\hat{\alpha}_n$ es el estimador de máximo de verosimilitud para el modelo de orden p . Sobre el caso ARMA se puede consultar [14] y los resultados de **2** sobre las mezclas permiten extender este tipo de resultado. De nuevo, el problema de selección optimal de $\ell(n)$ está abierto.

Bibliografía sobre los modelos no-identificables y las técnicas sugeridas en este artículo.

I - Los resultados expuestos provienen esencialmente de los tres artículos siguientes:

D.Dacunha-Castelle y E. Gassiat:

- [1] *The estimation of the order of a mixture model*. Bernoulli (1997) (en prensa)
- [2] *Testing in locally conic model and application to mixture models*. ESAIM (1997) (en prensa) - Prepublication Université Paris-Sud Orsay 96-61
- [3] *Testing the order of a model using locally conic parametrization: population mixtures and stationary ARMA processes*. Prepublication Université Paris-Sud Orsay 96-72

Artículos sobre mezclas.

- [4] Celeux, G. and Diebolt, J. (1992) *A stochastic approximation type EM algorithm for the mixture problem*. Stochastics and Stochastics reports 41, 119-134
- [5] Ghosh, J.K. and Sen, P.K. (1985) *On the asymptotic performance of the log-likelihood ratio statistic for the mixture model and related results*. Proc. Berkeley Conf. in Honor of Jerzy Neyman and Jack Kiefer (Vol.II), L.M. le Cam and R.A. Olshen (eds.), Wadsworth, Monterey, 807-810.
- [6] Hartigan, J.A. (1985). *A failure of likelihood asymptotics for normal mixtures*. Proc. Berkeley Con. in Honor of Jerzy Neyman and Jack Kiefer (Vol. II), L.M. le Cam and R.A. Olshen (eds.), Wadsworth, Monterey, 789-806.
- [7] Lindsay, B.G. (1989) *Moment matrices: applications in mixtures*. Ann. Statist. 17, 722-740.
- [8] Redner, R.A. (1981) *Note on the consistency of the maximum likelihood estimate for non identifiable distributions*. Ann. Statist. 9, 225-228.
- [9] Self, S. , Liang, K. *Asymptotic properties of maximum likelihood estimators and likelihood ration tests under nonstandard conditions*. Jour. Amer. Stat. Assoc., 823 (398): 605-610, 1987
- [10] Teicher, H. *Identifiability of finite mixtures*. Annals of Math. Statist., 36:423-439.
- [11] Yakowitz, S.J., Spragins, J.D. *On the identifiability of finite mixtures*. Annals of Math. Stat., 39:209-214, 1968.

Artículos sobre series cronológicas.

- [12] Akaike H.(1970). *Statistical predictor identification*. Ann. Inst. Statis. Math., 22, 202-217.
- [13] Akaike H. (1974). *A new look at the statistical model identification*. IEEE Trans. Aut. Cont. AC-19, 716-723
- [14] Azencott R., Dacunha-Castelle D. (1986) *Series of irregular observations forecasting and model building*. Springer
- [15] Dahlhaus, R. (1988) *Empirical processes and their applications to time series analysis*. Stochastic Processes and their Applications, 30, 69-83
- [16] Gerencser L., Rissanen J. *Asymptotics of predictive stochastic complexity*. In *New Directions in Time Series Analysis*, Springer-Verlag. Brillinger D., Caines P., Geweke J., Parzen E., Rosenblat M., Taqqu M. Eds.
- [17] Gill L., Lewbel A. (1992) *Testing the rank and definiteness of estimated matrices with applications to factor, state-space and Arma models*. J. Amer. Statist. Assoc. 87, 766-776
- [18] Hannan J. (1980) *The estimation of the order of an Arma process*. Annals of Stat., 8, (5) 1071-1081

[19] Hannan E.J. (1982) *Testing for autocorrelation and Akaike's criterion*. In Essays in Statistical Science, 403-412. Gani J.M., Hannan E.J. Eds.

[20] Veres S.(1987) *Asymptotic distribution of likelihood ratios for overparametrized Arma processes*. J. Time Series Anal. 8, 345-357

Artículos sobre los cálculos de entropía y los teoremas de Donsker.

[21] Ossiander M. (1987) *A central limit theorem under metric entropy with ℓ^2 bracketing*. Annals of Prob., 15, 897-919

[22] Pollard D. (1984) *Convergence of stochastic processes*. Springer-Verlag

[23] Van der Vart A., Wellner J.A. (1996) *Empirical processes*. Springer-Verlag

Artículos sobre las técnicas de momentos.

[24] Gamboa F. and Gassiat E. (1993). *Bayesian methods and Maximum entropy for ill posed inverse problems*. Submitted.

[25] Kailath, T. (1980) *Linear systems. Minimal order of state representation*. Prentice Hall.

[26] Karlin, S. and Studden, W.J. (1966) *Tchebychev systems with applications in analysis and statistics*. Wiley

[27] Krein, M.G. and Nudel'man, A. (1977) *The Markov moment problem and extremal problems*. Am. Math. Soc.

Artículos sobre otros modelos.

[28] Gamboa F. and Gassiat E. (1994) *Separation of discrete or circular sources*. Manuscript

[29] Gassiat, E. and Gautherat, E. (1994) *Identification of noisy linear systems with discrete random input*. Submitted