ESTIMACIO DELS PARAMETRES DE MODELS ARMA (P, Q) MITJANÇANT ALGORISMES DE FILTRATGE OPTIM

Tesi elaborada per:

Maria Pilar MUÑOZ i GRACIA

Sota la direcció del

Dr. Jaume PAGES i FITA

I presentada per a l'obtenció del

Grau de Doctora en Informàtica

Juny 1988

FACULTAT D'INFORMATICA DE BARCELONA

UNIVERSITAT POLITECNICA DE CATALUNYA

CAPITOL 1

INTRODUCCIO A L'ESTIMACIO DE MODELS ARMA UNIVARIANTS.

1.1 INTRODUCCIO

A la segona meitat del segle XIX apareix la preocupació per estudiar el comportament d'una sèrie temporal [28]. En aquesta època s'aplica l'anàlisi harmònica, utilitzant funcions sinusoïdals, per estudiar i representar moviments cíclics d'una sèrie temporal. El períodegrama és introduït l'any 1898 per Schuster; posteriorment, l'any 1934 Slustsky desenvolupa les propietats teòriques sota la hipòtesi de normalitat. Entre els anys 1920 i 1940 es desenvolupà la teoria dels processos estocàstics estacionaris, demostrant-se que tot procés estacionari es pot representar com una superposició d'oscil.lacions harmòniques. Al voltant de la tercera dècada del nostre segle, s'introdueix el concepte d'espectre d'un procés. La funció d'autocorrelació i la funció de densitat espectral es relacionen mitjançant la transformada de Fourier, essent una la transformada de l'altra. Entre 1940 i 1960 apareixen gran quantitat de treballs sobre l'estimació de l'espectre.

Wiener [26] en els anys quaranta introduí les bases de la teoria del filtratge. En els seus treballs va utilitzar l'anàlisi espectral per investigar les relacions de dependència entre sèries. Els treballs de Kalman i Bucy als anys seixanta van donar un canvi substancial a la teoria del filtratge. El seu plantejament ja no és freqüencial sinó temporal i es parla per primera vegada de formulació de models en l'espai d'estat. Es basa en les equacions diferencials que governen l'evolució de l'estat d'un sistema dinàmic. Les pertorbacions aleatòries converteixen els sistemes deterministes en estocàstics. Els processos generats són markovians, la qual cosa permet utilitzar gran quantitat de treballs realitzats anteriorment.

El canvi qualitatiu més substancial en el tractament de sèries temporals es produeix en la segona meitat dels anys seixanta gràcies a les investigacions realitzades per Box i Jenkins [5]. Seguint la pauta marcada per ells la millor manera d'abordar una anàlisi de sèries temporals és seguint un procediment iteratiu en tres etapes: -Identificació, Estimació i Verificació-. Box i Jenkins enfoquen l'estudi de les sèries temporals des de la perspectiva de l'espai del temps. En aquest marc estudien les propietats i comportament dels models lineals estacionaris ARMA (Autorregressive Moving Average) i dels models lineals no estacionaris ARIMA (Autorregressive Integrated Moving Average).

1.1.1 Anàlisi i Modelat de sèries temporals

L'estudi d'una sèrie temporal $y_1, \ldots, y_T$, comporta dos aspectes fonamentals: d'una banda l'anàlisi i d'altra, el modelat [13]. L'objectiu de l'anàlisi és resumir les propietats de les sèries i destacar-ne la modelització de les principals característiques. Les sèries es poden representar en l'espai del temps o en l'espai de les freqüències. En l'espai del temps, s'estudien les relacions entre les observacions obtingudes en diferents instants de temps, mentre que si es treballa en l'espai de les freqüències s'estudien moviments cíclics per analitzar el comportament de les sèries. Ambdós enfocaments són més aviat complementaris que oposats.

La principal raó del modelat d'una sèrie temporal és aconseguir previsions de futurs valors de les observacions. Les previsions són fetes per extrapolació dels valors ja coneguts.

El model matemàtic adequat per una sèrie temporal és el procés estocàstic [27]. Suposarem que el valor observat de la sèrie en l'instant t és una extracció a l'atzar d'una variable aleatòria definida en aquest instant, que està relacionada amb les d'altres instants a través del model que defineix el procés. En conseqüència, una sèrie de n observacions serà una mostra de n variables aleatòries ordenades en el temps ($y_1, \ldots, y_t, \ldots, y_n$). S'anomena procés estocàstic al conjunt d'aquestes variables $\{y_t\}$, t=1,...,n i la sèrie observada es considera una realització o trajectoria del procés.

En la literatura hi ha un gran nombre de procediments estadístics pel modelat d'una sèrie temporal. Aquest modelat va des de mètodes senzills com poden ser l'allisament exponencial, regressió, ... a tècniques més sofisticades com poden ser els models ARMA i ARIMA.

1.1.1.1 Models explícits del temps i descomposició de sèries temporals.

Un dels principals problemes que plantegen les sèries temporals és que la seva dependència del temps pot adoptar diverses formes. Una forma habitual de modelar les series temporals consisteix en considerar que l'observació té tres components: tendència, estacionalitat i error, totes elles depenents del temps

$$y_t = T_t + S_t + e_t \tag{1.1}$$

on y_t és l'observació a l'instant t, T_t és la tendència, S_t és l'estacionalitat i e_t és la pertorbació aleatòria.

La manera més senzilla d'estimar la tendència secular és la regressió, efectuant simultàniament un ajust estacional introduint variables fictícies; llavors el model a estimar és:

$$S_t = \beta_0 + \beta_1 T_t + \beta_2 S_t Q + e_t \tag{1.2}$$

on Q és una variable fictícia que prendrà el valor 1 quan hi hagi component estacional i zero en els altres casos.

1.1.1.2 Models recurrents i allisament exponencial

Un procés molt utilitzat per fer previsions és el següent:

$$y_t = y_{t-1} + e_t - c\, e_{t-1} \tag{1.3}$$

Si definim per B l'operador de retard i per Δ l'operador diferència, és a dir $\Delta = 1 - B$, aquest procés el podrem escriure de manera equivalent:

$$\Delta\, y_t = (1 - cB)\, e_t \tag{1.4}$$

on y_t és l'observació i e_t és la pertorbació aleatòria, ambdues a l'instant t.

Notem que, en particular, l'elecció c=1 ens porta al model (1.1). Per tant, els models recurrents com el (1.4) ens permeten també modelar tendències i estacionalitats.

Si $|c| < 1$, podem invertir la part (1-cB) de l'equació (1.4) i el procés s'escriurà com

$$y_t = (1-c)y_{t-1} + c\,(1-c)y_{t-2} + c^2(1-c)y_{t-3} + \ldots + e_t \tag{1.5}$$

i per tant, representa una situació en la que el pes de les observacions anteriors decreix en progressió geomètrica. Aquest procediment es coneix amb el nom d'allisament exponencial simple i va ser estudiat per Holt i Winters [27].

1.1.2 Processos estocàstics estacionaris ARMA.

1.1.2.1 Representació en l'espai del temps

Els treballs realitzats per Box-Jenkins [5] són particularment útils per fer previsions. Per comprendre aquests models és necessari en primer lloc, examinar amb tot detall les seves dues components, l'autorregressiva i la de mitjanes mòbils.

Considerem una sèrie temporal que representem per y_t i suposem que depèn de p valors previs, segons una regressió lineal anomenada autorregressió (AR) d'ordre p:

$$AR(p): y_t = a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + e_t \tag{1.6}$$

on e_t són pertorbacions aleatòries no correlacionades. En particular, per tot t,

e_t té mitjana constant zero

e_t té variància constant σ^2

e_t i e_{t+h} són no correlacionades ($h \neq 0$)

Per operar amb aquests processos es defineix l'operador de retard, B, per

$$B\, y_t = y_{t-1}$$

$$B^k y_t = y_{t-k}$$

i llavors l'equació (1.6) serà:

$$(1 - a_1 B - a_2 B^2 - \dots - a_p B^p)\, y_t = e_t$$

i si definim

$$A(B) = 1 - a_1 B - \dots - a_p B^p$$

obtenim l'expressió compacta

$$A(B)\, y_t = e_t$$

Anomenem equació característica del procés a l'equació $A(B)=0$. Aquesta equació tindrà p arrels $G_1^{-1}, \dots, G_p^{-1}$ i podrem escriure

$$A(B) = \prod_{i=1}^{p} (1 - G_i B)$$

El procés AR(p) és estacionari si $|G_i| < 1$ per tot $i=1,\dots,p$.

Una altra forma interesant de modelar una sèrie temporal és la de mitjanes mòbils. En general, es defineix un model de mitjanes mòbils d'ordre q com:

$$MA(q):\ y_t = e_t - c_1 e_{t-1} - c_2 e_{t-2} - \dots - c_q e_{t-q} \qquad (1.7)$$

on e_t compleix les propietats enunciades al paràgraf anterior.

Anàlogament, fent servir la mateixa notació utilitzada pels models autorregressius, podrem escriure el model MA(q) com:

$$y_t = (1 - c_1 B - \dots - c_q B^q)\, e_t$$

que és equivalent a definir

$$y_t = C(B)\, e_t$$

on en aquest cas

$$C(B) = 1 - c_1 B - \dots - c_q B^q$$

Denotem per M_j, $1 \leq j \leq q$ les q arrels de l'equació $C(B)=0$. El model MA(q) és invertible si $|M_j| > 1$ per tot $j=1,\dots,q$. Direm que un procés és invertible quan l'efecte del passat decreix amb el temps.

La sèrie temporal que combina els dos aspectes (1.6) i (1.7) s'anomena model ARMA(p,q):

$$\text{ARMA(p,q)}: y_t = a_1 y_{t-1} + \dots + a_p y_{t-p} + e_t - c_1 e_{t-1} - \dots - c_q e_{t-q} \qquad (1.8)$$

que podrem escriure de manera compacta com

$$A(B)\, y_t = C(B)\, e_t,$$

el que és equivalent a expressar aquest model com:

$$y_t = \frac{C(B)}{A(B)} e_t \qquad (1.9)$$

El procés es estacionari si les arrels del polinomi $A(B)=0$ estan fora del cercle unitat i invertible si hi són les del $C(B)=0$. El quocient $C(B)/A(B)$ s'anomena funció de transferència.

En l'espai del temps, les propietats dels models AR, MA i ARMA són caracteritzades per la seva funció d'autocovariància o bé la funció d'autocorrelació. No obstant, mentre que la funció d'autocovariància pot ressaltar algunes característiques del procés, hi ha d'altres aspectes que queden menys destacats. Quan les propietats cícliques de les sèries no queden caracteritzades per la seva funció d'autocovariància, serà desitjable examinar l'espectre de potència del procés [13].

1.1.2.2 Representació en l'espai d'estat

Els models lineals d'espai d'estat van ser desenvolupats originalment per l'enginyeria de control. En les aplicacions típiques es concentra l'atenció en un conjunt de m variables d'estat que van canviant al llarg del temps. La immensa majoria de vegades aquestes variables no seran directament observables, estan subjectes a una evolució sistemàtica emmascarada per soroll [13].

Encara que el vector d'estat, x_t, és no observable directament, el seu moviment vindrà caracteritzat per l'equació de la dinàmica:

$$x_{t+1} = F_t x_t + G_t e_{t+1} \tag{1.10}$$

on F_t i G_t són matrius de constants de dimensions mxm i mxg respectivament, i e_{t+1} és el vector de pertorbacions aleatòries de dimensió gx1 que segueix una distribució $N(0,\Sigma_1)$.

Les N variables que es poden observar estan contigudes en un vector Nx1, y_t, i estan relacionades amb les variables d'estat x_t per l'equació de l'observació. Aquesta equació la podrem escriure per:

$$y_t = H_t x_t + w_t \tag{1.11}$$

on H_t és una matriu de constants de dimensió Nxm i el vector de pertorbacions aleatòries w_t de dimensió nx1 es distribueix segons $N(0,\Sigma_2)$.

A més, els vectors de pertorbacions aleatòries d'ambdues equacions són no correlacionats i també estan no correlacionats amb el vector d'estat inicial x_0.

Qualsevol model ARMA(p,q) es pot representar en espai d'estat:

$$x_{t+1} = \left[\begin{array}{c|c} a_1 & . \\ a_2 & . \\ . & . \\ . & I_{m-1} \\ . & . \\ \hline a_m & 0' \end{array}\right] x_t + \begin{bmatrix} 1 \\ c_1 \\ . \\ . \\ . \\ c_{m-1} \end{bmatrix} e_t \qquad (1.12)$$

on m= max(p,q+1). Aquesta és una representació markoviana del model ARMA(p,q) definit a (1.8), encara que aquesta parametrització no és única [21]. L'equació (1.12) és l'equació de la dinàmica ja definida a (1.10) on

$$F_t = \left[\begin{array}{c|c} a_1 & . \\ a_2 & . \\ . & . \\ . & I_{m-1} \\ . & . \\ \hline a_m & 0' \end{array}\right] ; \qquad G_t = \begin{bmatrix} 1 \\ c_1 \\ . \\ . \\ . \\ c_{m-1} \end{bmatrix}$$

on F_t i G_t són matrius de constants i $\Sigma_1 = \sigma^2 I$.

Pel que fa a l'equació de l'observació:

$$y_t = \begin{bmatrix} 1 & 0'_{m-1} \end{bmatrix} x_t \qquad (1.13)$$

on $H_t = [\,1 \quad 0'_{m-1}]$ i en aquest cas no hi ha soroll en l'equació de l'observació.

1.1.3 Problemes resolts i per resoldre

L'estudi dels models estacionaris comporta tres aspectes diferents: Identificació, Estimació, i Detecció.

En primer lloc, la identificació del model. Aquest aspecte està resolt pels models AR(p) o MA(q). D'entre els criteris utilitzats per la identificació del model cal destacar el criteri d'Akaike [1] o la metodologia de Box-Jenkins [5].

D'altra banda, hi ha diverses formes d'identificar els models mixtes ARMA (p,q) però els criteris no han estat encara unificats. Per una exposició més completa i detallada es pot consultar l'article de Gooijer, Abraham, Gould i Robinson [12].

En segon lloc, l'estimació dels paràmetres del model. Quan es disposa d'un nombre de dades suficientment gran, l'estimació dels paràmetres d'un model ARMA (p,q) ha estat resolta mitjançant diverses tècniques.

No obstant, quan es disposa d'una única realització amb un nombre de dades limitat, inferior a una cinquantena, les tècniques desenvolupades amb la finalitat d'estimar aquests paràmetres no donen resultats sufientment satisfactoris.

En aquesta tesi desenvolupem un mètode d'estimació òptim mitjançant un filtre no lineal. Veure capítols 2 i 3.

Per últim, cal citar un aspecte encara no resolt, el problema de detecció. Aquest problema s'origina quan al llarg del temps es produeixen canvis en els paràmetres i canvis de model. Aquesta qüestió és bastant freqüent en models de tipus econòmic.

1.2 ESTIMACIO DE PARAMETRES D'UN MODEL ARMA EN L'ESPAI DEL TEMPS.

Una formulació adequada per una sèrie temporal va ser donada per Yule [28], l'any 1927. Yule representà una sèrie temporal y_t com la sortida d'un filtre lineal excitat mitjançant un soroll blanc; és a dir,

$$y_t = \sum_s h_s e_{t-s}$$

on e_t representa un soroll blanc i h_s la resposta impulsional del sistema. Yule va introduir formalment els models autoregressius (AR). La demostració que tota sèrie temporal estocàstica i estacionària es pot representar mitjançant un model de mitjanes mòbils es deu a Wold i és de l'any 1938. Durant els vint anys següents es van estudiar teòricament els models autorregressius i els models de mitjanes mòbils destacant els treballs de Walker, Barlett, Hannan i Quenouille entre molts altres.

Cap a finals dels anys 50 es van desenvolupar tota una sèrie de mètodes "ad hoc" per tractar sèries no estacionàries, i que es coneixen amb el nom de mètodes d'allisament exponencial. Aquests models són casos particulars de models ARIMA i bàsicament destaquen els treballs de Holt i Winters [28].

L'estimació dels paràmetres de sèries univariants ARMA ha progressat notablement des de la publicació del llibre de Box i Jenkins, l'any 1970. Box i Jenkins aproximen la funció de versemblança de dues maneres. La primera aproximació consisteix en minimitzar la suma dels quadrats dels residus condicionat a certs valors inicials. La segona utilitza el mètode de "back-forecasting".

El càlcul de la funció de versemblança d'un model ARMA(p,q) ha estat abordat per diversos autors. En primer lloc, Newbold [25] generalitza els algorismes proposats per Box i Jenkins [5] per l'estimació de processos MA al cas de processos ARMA. D'altra banda, en Ansley [3] utilitza la descomposició de Cholesky de la funció de covariàncies. Per últim, en Gardner, Harvey i Phillips [11] basen el càlcul exacte de la funció de versemblança en el filtre de Kalman.

Els algorismes proposats per Box i Jenkins, Newbold i Ansley poden ser classificats com algorismes globals, ja que tracten tota la informació conjuntament i troben els estimadors dels paràmetres fent servir el conjunt de totes les observacions, mentre que l'algorisme proposat per Gardner i al. és un algorisme recursiu que obté l'estimació dels paràmetres pas a pas. Cal destacar també l'aportació feta per Ljung i Söderstrom [19] a la teoria i pràctica de l'estimació recursiva dissenyant i demostrant entre altres algorismes, l'algorisme recursiu de l'error de predicció.

A continuació donem un breu repàs d'alguns algorismes d'estimació.

1.2.1 Algorismes globals

Cal destacar que aquests algorismes resulten ser subòptims degut a que estimen les pertorbacions aleatòries del model les a partir dels residus. D'altra banda, com que aquests algorismes es basen en resultats assimptòtics, no són apropiats quan el nombre de dades disponibles és petit.

1.2.1.1. Enfocament de Box i Jenkins

Seguint la notació donada per Box i Jenkins (1976), suposarem que el procés estacionari $\{y_t\}$ és generat per un model autorregressiu i de mijtanes mòbils d'ordre (p,q) definit per:

$$y_t = a_1 y_{t-1} + \dots + a_p y_{t-p} + e_t - c_1 e_{t-1} - \dots - c_q e_{t-q} \qquad (1.8)$$

és a dir $A(B)y_t = C(B)e_t$, on B és l'operador de retard. Donades n observacions de la sèrie $y' = (y_1, \dots, y_n)$, volem trobar els estimadors dels paràmetres $a' = (a_1, \dots, a_p)$ i $c' = (c_1, \dots, c_q)$ que maximitzin la funció de versemblança. La funció de versemblança per les observacions és:

$$l(a,c,\sigma^2/y) = (2\pi)^{-\frac{1}{2}n} |R|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} y' R^{-1} y\right) \qquad (1.14)$$

on σ^2 és la variància de les pertorbacions aleatòries e_t i R és la matriu de covariàncies de y de dimensió nxn.

Box i Jenkins [5] proposen no tenir en compte el determinant. D'aquesta manera, l'estimació es pot resoldre utilitzant el mètode dels mínims quadrats. No obstant, es pot comprovar que aquesta eliminació no dóna bons estimadors quan el nombre d'observacions a l'abast no és més gran d'una centena [21].

Box i Jenkins també demostren que en la funció de versemblança no condicional (1.14), es compleix

$$y' R^{-1} y = \sum_{t=-\infty}^{n} [e_t]^2$$

on $[e_t]=E(e_t \;/\; y,a,c)$. El límit inferior d'aquesta suma pot ser aproximat prenent $t=-t_0$ per un t_0 suficientment gran i obtenir els valors de $[e_t]$ fent servir el procediment iteratiu de "backforecasting". Newbold [25] remarca el fet que si hi ha un zero o un pol a prop del cercle unitat, serà necessari triar un t_0 bastant gran i/o iterar moltes vegades fins a aconseguir la convergència.

L'aproximació sobre les condicions inicials que proposen és la següent:

Es defineix per

$$S(a,c) = \sum_{t=1-q}^{n} [e_t]^2$$

El càlcul de la suma condicional S(a,c) necessita el coneixement de les condicions inicials e_i. L'aproximació de Box i Jenkins consisteix en calcular-la de manera iterativa fent primer una predicció i després una retropredicció. En síntesi, aquest algorisme funciona de la manera següent:

(1) En un primer pas es consideren els residus igual a zero en la finestra $[e_{1-q}, e_0]$.

(2) A continuació s'obté un valor inicial $(a_1^*, \ldots, a_p^*, c_1^*, \ldots, c_q^*)_0 = (a^*, c^*)_0 = Ø_0^*$ dels paràmetres. Aquest valor es calcula per mínims quadrats i és, per tant, l'estimador condicionat.

(3) El conjunt de les dades observades permet estimar els residus en la finestra [1,n]. Aquests residus es calculen recursivament:

$$e_t^* = y_t - a_{10}^*\, y_{t-1} - \ldots - a_{p0}^*\, y_{t-p} + c_{10}^*\, e_{t-1}^* + \ldots + c_{q0}^*\, e_{t-q}^* \qquad t=p+1,\ldots,n$$

(4) Els residus en la finestra temporal $[e_{n-q}, e_n]$ serveixen de condicions inicials per la retropredicció dels residus $[e_{1-q}, e_0]$ a partir de les observacions $(y_1, \ldots, y_n)$ i es torna al pas (2) calculant de nou un estimador dels paràmetres, prenent $Ø_1^*$ en comptes de $Ø_0^*$.

Es repeteixen les iteracions (2)-(4) fins a obtenir la convergència dels estimadors.

L'estimació màxim-versemblant de σ^2, s^2, proposada per Box i Jenkins és:

$$s^2 = \frac{S(Ø^*)}{n}$$

Es pot demostrar [5] que les propietats assimptòtiques del mètode de màxima versemblança són vàlides, en condicions molt generals, pels estimadors màxim-versemblants dels models ARMA. Per tant [27], la matriu de les segones derivades del logaritme de la funció de versemblança, evaluada en el seu màxim, proporciona la matriu de variàncies i covariàncies assimptòtiques dels estimadors:

$$\mathrm{Var}\,(\varnothing^*) = -\left[\frac{\partial^2 L(\varnothing^*)}{\partial \varnothing_i\, \partial \varnothing_j}\right]^{-1} \qquad (1.14a)$$

on L($\varnothing^*$) és el logarisme de la funció de versemblança.

Aquest resultat es basa en que la funció de versemblança és aproximadament quadràtica en el màxim, per tant l'hessià serà constant en un entorn del màxim. Per comprovar-ho cal, sempre que sigui possible, estudiar la forma del logaritme de la funció de versemblança al voltant de l'estimador màxim-versemblant, calculant numèricament i dibuixant les seves corbes de nivell.

Es pot demostrar que la matriu hessiana de les segones derivades augmenta quan n augmenta i aplicant la desigualtat de Cramer-Rao s'arriba a la conclusió de que efectivament la matriu de variàncies i covariànces assimptòtiques dels estimadors es poden obtenir a partir de l'equació (1.14a)

Un gran nombre d'autors han donat expressions exactes per (1.14) útils computacionalment. Box i Jenkins donen una expressió pel cas d'un procés de mitjanes mòbils pur i aquesta idea l'ha desenvolupat Newbold [25] per un model mixte ARMA.

1.2.1.2 Enfocament de Newbold

Ens proposem examinar l'enfocament fet per Newbold [25], que permet obtenir la funció de versemblança exacta. Per fer-ho es defineixen tres vectors:

$$\varepsilon = (e_1, \ldots, e_n)'$$

$$\varepsilon^* = (e_{1-q}, \ldots, e_0, y_{1-p}, \ldots, y_0)'$$

$$y = (y_1, \ldots, y_n)'$$

ε^* representa el vector de les condicions inicial. Es necessària la seva estimació per obtenir l'estimador màxim-versemblant dels paràmetres del model ARMA.

Noti's també, que la matriu de correlacions de ε^* té una forma simple ja que conté les autocorrelacions de les observacions, les dels residus i les autocorrelacions creuades de les observacions i dels residus.

L'equació (1.8) es pot escriure com

$$e_t = y_t - a_1 y_{t-1} - \ldots - a_p y_{t-p} + c_1 e_{t-1} + \ldots + c_q e_{t-q} \quad (1 \le t \le n) \qquad (1.15)$$

i aquesta equació (1.15) proporciona un conjunt d'equacions que relacionen les components de ε, ε^* i y i que es poden posar en forma matricial:

$$\begin{bmatrix} \varepsilon^* \\ \varepsilon \end{bmatrix} = \begin{bmatrix} 0 \\ L \end{bmatrix} y + \begin{bmatrix} I \\ X \end{bmatrix} \varepsilon^* \qquad (1.16)$$

on L i X són respectivament les matrius nxn i nx(p+q) que inclouen només els coeficients $a_1,\ldots,a_p$ i $c_1,\ldots,c_q$.

Sigui

$$E[\ \varepsilon^* \varepsilon^{*\prime}\ ] = \sigma^2 \Omega$$

on Ω és una matriu de dimensió (p+q)x(p+q).

Sigui T una matriu quadrada no singular tal que

$$T\Omega T' = I_{p+q}$$

la matriu T es pot obtenir per triangularització de Ω

$$\Omega = T^{-1}(T')^{-1}$$

L'equació (1.16) s'escriu com:

$$\begin{bmatrix} u^* \\ u \end{bmatrix} = \begin{bmatrix} 0 \\ L \end{bmatrix} y + \begin{bmatrix} I \\ X\,T^{-1} \end{bmatrix} u^* \qquad (1.17)$$

on $u^* = T\,\varepsilon^*$ i $u = \varepsilon$.

Es compleix que:

$$E\left[u^*\ u^{*\prime}\right] = T\,E\left[\varepsilon^*\ \varepsilon^{*\prime}\right] T' = \sigma^2\ T\Omega T' = \sigma^2\ I$$

A més u^* és independent de u, de manera que:

$$U = \begin{bmatrix} u^* \\ u \end{bmatrix}$$

és soroll blanc.

Si, a més

$$\begin{bmatrix} u^* \\ u \end{bmatrix}$$

és gaussià, llavors la seva densitat de probabilitat s'escriu com:

$$p(U/\sigma) = (2\pi\sigma^2)^{-1/2(n+p+q)} \exp\{\frac{1}{2}\frac{U'U}{\sigma^2}\}$$

i si posem l'equació (1.17) en la forma reduïda $U = \Lambda y + \chi u^*$

$$\text{on } \Lambda = \begin{bmatrix} 0 \\ L \end{bmatrix} \text{ i } \chi = \begin{bmatrix} T \\ XT^{-1} \end{bmatrix}$$

s'obté:

$$p(y,u^*/a,c,\sigma) = (2\pi\sigma^2)^{-\frac{1}{2}(n+p+q)} \exp\{-\frac{1}{2} S(a,c,u^*)\} \tag{1.18}$$

on $S(a,c,u^*) = (\Lambda y + \chi u^*)'(\Lambda y + \chi u^*)$.

L'estimador $\hat{u}^*$ de u^* que dóna el màxim de $p(y,u^*/a,c,\sigma)$ ve donat per

$$\hat{u}^* = -(\chi'\chi)^{-1}\chi'\Lambda y$$

de manera que

$$S(a,c,u^*) = S(a,c) + (u^* - \hat{u}^*)'\chi'\chi(u^* - \hat{u}^*)$$

amb

$$S(a,c) = (\Lambda y + \chi\hat{u}^*)'(\Lambda y + \chi\hat{u}^*)$$

i reescrivint (1.18) de la forma:

$$p(y,u^*/a,c,\sigma) = p(y/a,c,\sigma)p(u^*/y,a,c,\sigma)$$

l'expressió de la funció de versemblança exacta per qualsevol model ARMA(p,q) és:

$$p(y/a,c,\sigma) = (2\pi\sigma^2)^{-1/2n} |\chi'\chi|^{-1/2} \exp\{-\frac{1}{2} S(a,c)/\sigma^2\}$$

1.2.1.3 Enfocament de Ansley

Ansley [3] proposa el següent algorisme pel càlcul exacte de la funció de versemblança d'un procés autorregressiu i de mitjanes mòbils:

Es considera la transformació

$$z_t = \begin{cases} A(B)y_t & (t=1,\dots,m) \\ y_t & (t=m+1,\dots,n) \end{cases} \tag{1.20}$$

on m=max(p,q). Es obvi que $cov(z_t, z_{t+s})=0$ per $|s|>m$, i per $|s|>q$ si $min(t,t+s)>m$. La matriu de covariàncies R^z de $z'=(z_1,\dots,z_n)$ és una matriu en banda amb amplada màxima de banda m per les m primeres columnes i q per les altres.

La matriu R^z admet la descomposició de Cholesky $R^z=LL'$ on L és una matriu en banda triangular inferior amb les amplades de bandes corresponents a les de R^z. Definint $b=L^{-1}z$ s'obté una successió de variables aleatòries independents $b'=(b_1,\dots,b_n)$ amb distribució normal de mitjana zero i variància σ^2. El jacobià de la transformació (1.20) és la unitat i tenim

$$p(a,c,\sigma^2/y) = (2\pi\sigma^2)^{-\frac{1}{2}n} |L|^{-1} \exp\left(-\sum_{t=1}^{n} \frac{1}{2} b_t^2/\sigma^2\right) \qquad (1.21)$$

Les files de L i les solucions b_t de l'equació Lb=z es poden calcular recursivament simultàniament. El determinant |L| és simplement el producte dels elements de la diagonal principal.

La forma adient de calcular computacionalment els estimadors màxim versemblants és maximitzar el logaritme neperià de (1.21) respecte de σ^2 obtenint L_{max} = const -1/2nlogS, on

$$S = \sum_{t=1}^{n} u_t^2 \ , \quad u_t = |L|^{1/n} b_t$$

El màxim es trobarà aplicant un algorisme no lineal de mínims quadrats a S.

1.2.2 Algorismes recursius

1.2.2.1 Enfocament de Gardner, Harvey i Phillips

L'algorisme que proposen Gadner, Harvey i Philips [11] per trobar la funció de versemblança exacta d'un model ARMA(p,q) utilitzant el filtre de Kalman és el següent:

Tal com ja hem esmentat en l'apartat 1.1 d'aquest capítol, tot model ARMA(p,q) definit per (1.8)

$$y_t = a_1 y_{t-1} + \dots + a_p y_{t-p} + e_t - c_1 e_{t-1} - \dots - c_q e_{t-q} \qquad (1.8)$$

es pot representar en espai d'estat, definint l'equació de la dinàmica com

$$x_{t+1} = \left[\begin{array}{c|c} a_1 & \cdot \\ a_2 & \cdot \\ \cdot & \cdot \\ \cdot & I_{m-1} \\ \vdots & \cdot \\ \hline a_m & 0' \end{array}\right] x_t + \begin{bmatrix} 1 \\ c_1 \\ \cdot \\ \cdot \\ \cdot \\ c_{m-1} \end{bmatrix} e_t \qquad (1.12)$$

on m= max(p,q+1) i

$$F_t = \left[\begin{array}{c|c} a_1 & \cdot \\ a_2 & \cdot \\ \cdot & \cdot \\ \cdot & I_{m-1} \\ \cdot & \cdot \\ \hline a_m & 0' \end{array}\right]; \qquad G_t = \begin{bmatrix} 1 \\ c_1 \\ \cdot \\ \cdot \\ \cdot \\ c_{m-1} \end{bmatrix}$$

Tal com ja hem dit també en l'apartat 1.1 l'equació de l'observació és

$$y_t = \begin{bmatrix} 1 & 0^{'}_{m-1} \end{bmatrix} x_t \tag{1.13}$$

Donat un estimador de l'estat x^*_{t-1} a l'instant t-1 juntament amb la matriu P_{t-1} definida per

$$E[\,(x^*_{t-1} - x_{t-1})(x^*_{t-1} - x_{t-1})'] = \sigma^2 P_{t-1}$$

es pot fer una predicció de x_t, $x^*_{t|t-1}$. Aquesta pot ser actualitzada quan l'observació t-èssima y_t estigui disponible. La predicció i l'actualització és a dir el filtrat seran portades a terme utilitzant el conjunt recursiu d'equacions conegut amb el nom de "Filtre de Kalman".

Per iniciar la recursió cal un estimador de l'estat inicial x_0 i de la matriu associada a aquest estat P_0. Gardner i al. [11] proposen agafar com el millor estimador de x_0 per un model ARMA $x^*_0=0$ i la matriu P_0 és calculada a partir de $\sigma^{-2}E[x_0\, x_0']$. El mètode proposat per calcular-la és el següent :

La matriu inicial, P_0, compleix l'equació

$$P_0 = F_0 P_0 F_0\,' + G_0\, G_0\,' \tag{1.22}$$

Si $V=G_0G_0'$ i p_{ij}, f_{ij} i v_{ij} denoten l'element de la i–èssima fila i j-èssima columna de les matrius P_0, F_0 i V respectivament, llavors

$$p_{ij} = \sum_k \sum_l f_{ik}\, p_{kl}\, f_{jl} + v_{ij} \tag{1.23}$$

és a dir,

$$v_{ij} = p_{ij} - \sum_k \sum_l f_{ik}\, p_{kl}\, f_{jl} \tag{1.24}$$

Així cada element de V és una combinació lineal de elements de P_0. Es a dir, podem escriure

$$\mathrm{vec}(V) = S\ \mathrm{vec}(P_0) \tag{1.25}$$

on S és una matriu quadrada adient, la forma de la qual depèn de la definició de vec(V). Per les posibles eleccions de vec(V) veure Gardner i al. [11]. L'expressió (1.25) és un conjunt d'equacions lineals a partir de les quals es pot obtenir P_0.

L'aplicació de les fòrmules recursives del "filtre de Kalman" dóna un conjunt de n residus estandartitzats, denotats per u_t, t=1,...,n juntament amb un conjunt de n quantitats b_t t=1,...,n proporcionals als errors quadràtics mitjans de predicció de cada pas. El logaritme de la funció de versemblança pot ser maximitzat respecte de (a,c) minimitzant

$$L^* (a,c) = n \log S(a,c) + \sum_{t=1}^{n} \log b_t \tag{1.26}$$

on $S(a,c)=\sum u_t^2$.

1.2.2.2 Enfocament de Ljung i Söderström

Finalment, Ljung i Söderström [19] proposen l'algorisme recursiu de l'error de predicció, que denotem per RPEM:

Es denota per $yp_{t/\varnothing}$ la predicció en l'instant t basada en les observacions y_s $0 \leq s \leq t-1$ i en la hipòtesi que les observacions segueixen un model ARMA(p,q) definit per (1.8) amb:

$$\varnothing' = (a_1,\ldots,a_p,c_1,\ldots,c_q)$$

Es defineix l'error de predicció per

$$\varepsilon_t(\varnothing) = y_t - yp_{t/\varnothing} \qquad (1.27)$$

i el vector

$$\phi_t = (-y_{t-1}, \ldots, -y_{t-p})'$$

El vector de paràmetres desconegut $\varnothing$ serà estimat minimitzant una funció de cost quadràtica que inclogui el vector de les observacions ϕ_t i els errors de predicció. El criteri quadràtic a minimitzar proposat és:

$$V_t = \sum_{s=1}^{t} \frac{\varepsilon_s^2(\varnothing)}{\lambda_s + \phi_s' P_{s-1} \phi_s} \qquad (1.29)$$

i P_t és la matriu de covariàncies de l'estimador.

Suposem conegut en l'instant t un estimador $\emptyset_{t-1}$ dels paràmetres. Segons Ljung i Söderström [19], desenvolupant en sèrie de Taylor V_t al voltant de $\emptyset_{t-1}$, l'estimador de $\emptyset$ pot ser actualitzat utilitzant les següents fòrmules:

CALCUL DE LA PREDICCIO

$$yp_{V\emptyset} = \emptyset_{t-1}'\phi_t \qquad (1.30)$$

ACTUALITZACIO DELS PARAMETRES ESTIMATS

$$\emptyset_t = \emptyset_{t-1} + L_t\,\varepsilon_t(\emptyset) \qquad (1.31)$$

CALCUL DEL VECTOR DE GUANY

$$L_t = \gamma_t\, R_t^{-1}\psi_t\, P_t^{-1} \qquad (1.32)$$

on R_t és l'aproximació de l'hessià, és a dir $R_t \approx V_t''$ i ψ_t és el gradient de les prediccions, és a dir

$$\psi_t = -\frac{d\varepsilon_t(\emptyset)}{d\emptyset}$$

ACTUALITZACIO DE LA MATRIU DE COVARIANCIES DELS ESTIMADORS:

$$P_t = P_{t-1} + \gamma_t\,[\,\varepsilon_t(\emptyset)\,\varepsilon_t'(\emptyset) - P_{t-1}] \qquad (1.33)$$

i γ_t ha de ser tal que:

$$\lambda_t = \frac{\gamma_{t-1}}{\gamma_t} [1 - \gamma_t]$$

$$\lambda_t = 0.99 \lambda_{t-1} + 0.01$$

$$\lambda_0 = 0.95$$

S'ha simulat una bateria de sèries estacionàries i els paràmetres s'han estimat mitjançant el paquet estadístic BMDP2T[9] que utilitza la metodologia Box i Jenkins[5] i mitjançant l'algorisme recursiu RPEM. Aquest algorisme ha estat implementat, pel cas univariant, a l'ordinador de la F.I.B. (Cluster VAX 8600-785) obtenint-se un tempss promig de CPU de 1.75 segons pel cas de 500 observacions.

Aquest treball empíric posa de manifest que la dimensió mínima de mostra per a estimar els coeficients correctament depèn de la distància entre pols i zeros i de la distància d'aquests a l'origen, com es pot veure a l'annex 1.

En general, per tots els models estudiats, les estimacions dels paràmetres realitzades amb 500 observacions donen resultats equivalents segons s'hagi utilitzat BMDP2T o RPEM, no obstant l'algorisme RPEM permet l'estimació amb tamany de mostra sense limitacions mentre que BMDP2T presenta limitacions.

L'estudi d'aquests algorismes subòptims ens du a la necessitat d'una estimació òptima, en el sentit de trobar un filtre òptim que no es basi en aproximacions. Al mateix temps, la variància dels estimadors en els algorismes abans exposats no queda gaire definida, en uns perquè en la bibliografia no apareix cap referència sobre aquest punt, per exemple

Gardner i al [11], Newbold [25], Ansley [3]; en l'algorisme de Box i Jenkins [5], es fan tantes aproximacions que quan el nombre de dades disponibles sigui al voltant d'una centena, difícilment s'aconseguirà trobar una bona estimació de la variància dels estimadors. L'únic algorisme que va calculant la variància dels estimadors pas a pas és l'algorisme recursiu de l'error de predicció proposat per Ljung i Söderström [19]. Tots aquests algorismes subòptims donen lloc a que les estimacions dels paràmetres convergeixin cap als vertaders valors a partir d'un nombre d'observacions.

En aquesta tesi desenvolupem un filtre no lineal òptim on tota la informació sobre els paràmetres pot ser obtinguda a partir de la funció de densitat a posteriori: estimadors òptims, precisió d'aquests estimadors, estimació de la variància del soroll, nombre de dades necessaries par aconseguir una precisió determinada, ...

1.3 ESTIMACIO OPTIMA: EL NOSTRE ENFOCAMENT

L'estimació òptima dels paràmetres del model es plantejarà sota un punt de vista baiesià, ja que els paràmetres són considerats com variables aleatòries. Podrem inferir informació sobre els valors d'aquests paràmetres si considerem els esmentats paràmetres com una variable aleatòria i tenim en compte les observacions d'altres variables aleatòries que estiguin correlacionades amb aquests paràmetres. El filtre de Kalman [22] està desenvolupat des d'aquest punt de vista bayesià. El vector d'estat del sistema no observat està correlacionat amb les dades observables de manera que es pot fer una estimació del vector d'estat del sistema basada en aquestes dades observables.

Suposem que la dinàmica del sistema pot ser descrita en termes d'un vector de paràmetres $\emptyset$. Sota l'òptica baiesiana, podem considerar que aquest vector d'estat $\varnothing$ és una variable aleatòria amb una certa distribució a priori. Obviament, les observacions y^t (on el superíndex t indica el conjunt d'observacions fins l'instant t) estan correlacionades amb aquest vector $\varnothing$. Volem determinar a l'instant t la funció de densitat de probabilitat a posteriori de $\varnothing$, i.e., $p(\varnothing / y^t)$.

Hi ha diversos mètodes per trobar un estimador $\varnothing_t^*$ a partir de la seva distribució a posteriori. El mètode més utilitzat és triar com estimador la mitjana de la distribució a posteriori, és a dir l'esperança condicional:

$$\varnothing_t^* = E[\varnothing / y^t] \tag{1.34}$$

Un altre possible estimador $\varnothing_t^*$ és el màxim de la distribució a posteriori. Aquest estimador és conegut amb les sigles de MAP i coincideix amb l'estimador obtingut en (1.34) per algunes distribucions simètriques, en particular per les distribucions gaussianes. L'estimador $\varnothing_t^*$ definit a (1.34) és també el que minimitza l'error quadràtic mitjà [15], és a dir minimitza:

$$E\{ (\varnothing - \varnothing_t^*)^2 / y^t \}$$

El nostre problema serà determinar com l'estimador definit en (1.34) o bé la densitat de probabilitat a posteriori dels paràmetres, evoluciona al llarg del temps. Aquest és un problema complicat i no resoluble moltes vegades.