

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tesisenxarxa.net) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tesisenred.net) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tesisenxarxa.net) service has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized neither its spreading and availability from a site foreign to the TDX service. Introducing its content in a window or frame foreign to the TDX service is not authorized (framing). This rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author

**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 2

FILTRE NO LINEAL PER L'ESTIMACIO DE PARAMETRES DE PROCESSOS AR(P) AMB LA VARIANCIA DEL SOROLL DESCONEGUDA.

2.1 INTRODUCCIO

Suposem que volem estimar els paràmetres i la variància del soroll d'un model AR(p) definit per l'equació:

$$y_t = a_1 y_{t-1} + \dots + a_p y_{t-p} + e_t \quad (1.6)$$

on $\{e_t\}$ és una successió de variables aleatòries independents amb distribució $N(0, \sigma^2)$ essent la variància σ^2 desconeguda. Les úniques variables observables són la successió d'observacions $\{y_t, t=1, \dots, n\}$. El nostre interès en aquest capítol és formular l'esmentat model en l'espai d'estat i a continuació, utilitzant la tècnica del filtre no lineal òptim, estimar conjuntament el vector de paràmetres $\vartheta' = (a_1, \dots, a_p)$ i la variància del soroll σ^2 que són desconeguts.

La modelització adient pel filtre no lineal òptim serà:

Denotem per x l'estat del sistema i y la successió d'observacions, llavors el model matemàtic discret és:

$$\begin{aligned} x_{t+1} &= F(\vartheta) x_t + G(\vartheta) e_{t+1} \\ y_t &= H(\vartheta) x_t \end{aligned} \quad (2.1)$$

on $\{e_t\}$ és una successió de variables aleatòries, independents distribuïdes

segons una $N(0, \Sigma_1)$ i independents de la condició inicial x_0 . Suposem que x_0 té distribució a priori coneguda amb mitjana μ_0 i matriu de covariàncies Π_0 .

Aquesta parametrització no és única [4]. En aquest cas, la parametrització no conté soroll en l'observació.

Suposem que la distribució a priori del vector de paràmetres té distribució a priori coneguda amb mitjana \emptyset_0 i matriu de covariàncies P_0 i la distribució a priori de la variància del soroll és coneguda també amb mitjana σ_0^2 i matriu de covariàncies Σ_0 .

Per determinar un estimador recursiu de \emptyset i de σ^2 definim un vector d'estat ampliat

$$X_t = \begin{bmatrix} x_t \\ \emptyset \\ \sigma^2 \end{bmatrix} \quad (2.2)$$

i considerem el problema de l'estimació de l'estat per aquest estat ampliat.

L'estat compleix les equacions:

$$\begin{aligned} X_{t+1} &= \bar{F}(X_t) + \bar{\theta}_t \\ y_t &= \bar{H}(X_t) \end{aligned} \quad (2.3)$$

on

$$\bar{F}(X_t) = \begin{bmatrix} F(\emptyset) \\ \emptyset \\ \sigma^2 \end{bmatrix} \quad (2.3a)$$

$$\bar{e}_t = \begin{bmatrix} G(\emptyset)e_t \\ 0 \end{bmatrix} \quad (2.3b)$$

$$G(\emptyset) = \begin{bmatrix} 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} \quad (2.3c)$$

$$\bar{H}(X_t) = H(\emptyset) x_t$$

Definim per $y^t = \{ y_1, \dots, y_t \}$ el conjunt de totes les observacions fins l'instant t .

El problema en (2.3) és determinar $E(X_t / y^t)$ que s'obtindrà a partir de la funció de densitat de probabilitat $p(X_t / y^t)$ calculada de forma recurrent, sempre i quan la distribució a priori inicial $p(X_0)$ sigui coneguda, on X_0 són les condicions inicials. A partir de l'esmentada distribució $p(X_t / y^t)$ podrem calcular els estimadors òptims $E(\emptyset / y^t)$ i $E(\sigma^2 / y^t)$, les variàncies dels quals són mínimes [15], [26], la qual cosa justifica el qualificatiu d'estimadors òptims.

El problema de l'estimació recursiva de θ i σ^2 està formulat com un problema d'estimació d'estat, però l'evolució de l'estat X_t és no lineal en les equacions (2.3), per tant és un problema de filtratge no lineal [26].

Poden plantejar-se també els problemes de predicció i d'interpolació, els quals tenen una formulació semblant, només que el valor que es pretén estimar és y_t , amb $t' > t$ al cas de la predicció i $t' < t$ al cas de la interpolació.

2.1.1 Model de l'espai d'estat general per un procés AR.

El model AR(p) definit en l'apartat 1.1.2.1 per l'equació

$$y_t = a_1 y_{t-1} + \dots + a_p y_{t-p} + e_t \quad (1.6)$$

es pot representar en l'espai d'estat mitjançant la segona forma canònica observable [20] com

$$\begin{bmatrix} x_{t+1}^1 \\ x_{t+1}^2 \\ \cdot \\ \cdot \\ x_{t+1}^p \end{bmatrix} = \begin{bmatrix} a_1 & \cdot \\ a_2 & I_{p-1} \\ \cdot & \cdot \\ \cdot & \cdot \\ a_p & 0 \end{bmatrix} \begin{bmatrix} x_t^1 \\ x_t^2 \\ \cdot \\ \cdot \\ x_t^p \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{bmatrix} e_{t+1}$$

$$y_t = [1 \ 0 \ \cdot \ \cdot \ \cdot \ 0] x_t$$

Definit el vector d'estat ampliat (2.2), amb la parametrització precedent, es tindran les equacions d'estat no lineals:

$$\begin{aligned} X_t &= \bar{F}(X_t) + \bar{e}_t \\ y_t &= \bar{H}(X_t) \end{aligned} \quad (2.4)$$

$$\text{on } \bar{F}(X_t) = \begin{bmatrix} A x_t \\ \emptyset \\ \sigma^2 \end{bmatrix} \quad \bar{e}_t = \begin{bmatrix} e_t \\ 0 \\ 0 \end{bmatrix} \quad \bar{H}(X_t) = H x_t ; \emptyset = (a_1, \dots, a_p)'$$

$$H = [1 \ 0 \ \dots \ 0]$$

x_t és el vector d'estat definit en el paràgraf precedent i

$$A = \begin{bmatrix} a_1 & \cdot \\ a_2 & I_{p-1} \\ \cdot & \cdot \\ a_p & 0 \end{bmatrix}$$

és a dir

$$\begin{bmatrix} x_{t+1}^1 \\ x_{t+1}^2 \\ \vdots \\ x_{t+1}^p \\ \emptyset \\ \sigma^2 \end{bmatrix} = \begin{bmatrix} a_1 x_t^1 + x_t^2 \\ a_2 x_t^1 + x_t^3 \\ \vdots \\ a_p x_t^1 \\ \emptyset \\ \sigma^2 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \end{bmatrix} e_{t+1} \quad (2.4')$$

$$y_t = [1 \ 0 \ \dots \ 0] x_t \quad (2.4'')$$

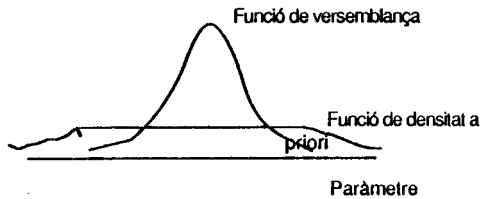
2.1.2 Discussió de l'elecció de la distribució a priori.

L'elecció de la distribució a priori dels paràmetres desconeguts del model ha de ser feta d'acord amb la idea que a priori es coneix molt poc sobre el comportament dels paràmetres i la variància del soroll. Aquesta hipòtesi es coneix com el postulat de Bayes [6]. Per aquest motiu ens situarem en un marc en el qual la informació a priori no ha d'estar subjecte a gaires restriccions i ha de ser el més general possible. Veurem més endavant que les distribucions a priori d'aquest tipus es diuen no informatives.

El nostre problema és relacionar el coneixement a priori dels esmentats paràmetres amb la informació subministrada per les dades, que no és res més que la funció de versemblança.

En primer lloc definirem el concepte de distribució a priori localment uniforme [6]:

Es defineix com distribució a priori localment uniforme aquella que és dominada per la funció de versemblança, és a dir, "una funció que no canvia molt sobre la regió en la que la funció de versemblança és apreciable i no presenta grans valors fora d'aquesta regió" [6], és a dir, de l'estil de la següent figura



Per aquestes funcions de densitat a priori podrem calcular la funció de densitat a posteriori a partir de la fórmula de Bayes:

$$p(A / B, C) = \frac{p(B / A, C) p(A / C)}{p(B / C)}$$

on $p(A / B, C)$ és la probabilitat de l'esdeveniment A, condicionada per B i C.

La funció de densitat a posteriori $p(y^1 | \varnothing, \sigma^2)$ pot ser considerada una funció de \varnothing, σ^2 en lloc d'una funció de y^1 . Quan això succeeix i d'acord amb Fisher (1922) [6], l'anomenarem funció de versemblança dels paràmetres \varnothing, σ^2 donats y^1 i la denotarem per $l(\varnothing, \sigma^2 / y^1)$.

Aplicant la fórmula de Bayes a la funció de densitat a posteriori i substituint la funció de densitat a priori per una constant s'obté:

$$p(\varnothing, \sigma^2 / y^1) = \frac{l(\varnothing, \sigma^2 / y^1) p(\varnothing, \sigma^2)}{\int l(\varnothing, \sigma^2 / y^1) p(\varnothing, \sigma^2) d\varnothing d\sigma^2} \propto \frac{l(\varnothing, \sigma^2 / y^1)}{\int l(\varnothing, \sigma^2 / y^1) d\varnothing d\sigma^2} \quad (2.5)$$

on $p(\varnothing, \sigma^2 / y^1)$ és la funció de densitat a posteriori

$l(\varnothing, \sigma^2 / y^1)$ és la funció de versemblança que és igual a $p(y^1 | \varnothing, \sigma^2)$

i $p(\varnothing, \sigma^2)$ és la funció de densitat a priori.

Així, per una densitat a priori localment uniforme, la funció de densitat a posteriori numèricament és aproximadament igual a la funció de versemblança normalitzada.

A continuació introduïrem el concepte de distribució a priori no informativa:

Sigui $\varphi(\varnothing, \sigma^2)$ una transformació bijectiva de \varnothing i σ^2 que compleix les condicions de mètrica i tal que la funció de versemblança $l(\varnothing, \sigma^2/y^1)$ és invariant per translació, és a dir

$$l(\varnothing, \sigma^2/y^1) = g[\varphi(\varnothing, \sigma^2) - f(y^1)] \quad (2.6)$$

on $g(x)$ és una funció coneguda independent de les dades y^1 i $f(y^1)$ és una funció de les dades y^1 .

Direm que una distribució a priori de \varnothing i σ^2 , amb funció de densitat $p(\varnothing, \sigma^2)$, és no informativa pels paràmetres \varnothing i σ^2 si és localment uniforme sobre $\varphi(\varnothing, \sigma^2)$ (Box i Tiao [6]). Notem que això implica que la funció de densitat a priori $p(\varnothing, \sigma^2)$ és proporcional localment a $|J|$ on $|J|$ és el valor absolut del Jacobià de la transformació.

El concepte de distribució a priori no informativa ens proporciona un argument per triar una mètrica adient, és a dir, escollirem una mètrica pels paràmetres $\varphi(\varnothing, \sigma^2)$ sota la qual la funció de versemblança sigui invariant per translació i la distribució a priori en aquesta mètrica sigui localment uniforme.

2.2 ALGORISME DE FILTRATGE NO LINEAL.

2.2.1 Elecció de la distribució a priori

A partir de l'equació (1.6):

$$y_t = a_1 y_{t-1} + \dots + a_p y_{t-p} + e_t \quad (1.6)$$

es pot veure que

$$\begin{aligned} y_0 &= e_0 \\ y_1 &= a_1 y_0 + e_1 = a_1 e_0 + e_1 \\ &\vdots \\ y_p &= a_1 y_{p-1} + \dots + a_p e_0 + e_p = a_1^p e_0 + \dots + e_p \end{aligned}$$

i així successivament. Per tant, quan es coneixen els paràmetres, les observacions són combinacions lineals dels diferents sorolls blancs gaussians, i llavors

$$p(y^t | \theta, \sigma^2) = \frac{1}{(2\pi)^{t/2} (\det R_t)^{1/2}} \exp\{-1/2 (y^t)' R_t^{-1} y^t\} \quad (2.7)$$

on R_t és la matriu de covariàncies de les observacions. L'expressió de l'esmentada matriu de covariàncies la trobarem a partir de les següents equacions:

Definim, pel cas estacionari,

$$\gamma_k = E [y_{t+k} y_t] \quad k \geq 0 \quad (2.8)$$

$$\gamma_{ey}(j) = E [e_t y_{t+j}] \quad j > 0$$

ja que el soroll en l'instant t , e_t , és independent de les observacions anteriors i [5] trobem la recurrència

$$\gamma_k = a_1 \gamma_{k-1} + \dots + a_p \gamma_{k-p} \quad k \geq 1 \quad (2.9)$$

$$\gamma_0 = a_1 \gamma_1 + \dots + a_p \gamma_p + \sigma^2$$

que ens permet obtenir els $p+1$ primers elements de $R_\gamma \{ \gamma_k, k=0, \dots, p \}$ resolent un sistema d'equacions lineal de $p+1$ equacions amb $p+1$ incògnites que és:

$$\begin{aligned} \gamma_0 - a_1 \gamma_1 - \dots - a_p \gamma_p &= \sigma^2 \\ \gamma_1 - a_1 \gamma_0 - \dots - a_p \gamma_{p-1} &= 0 \\ &\vdots \\ \gamma_p - a_1 \gamma_{p-1} - \dots - a_p \gamma_0 &= 0 \end{aligned}$$

és a dir:

$$\begin{bmatrix} 1 & -a_1 & \dots & -a_p \\ -a_1 & (1-a_2) & \dots & 0 \\ & \cdot & \cdot & \cdot \\ & \cdot & \cdot & \cdot \\ -a_p & -a_{p-1} & \dots & 1 \end{bmatrix} \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \cdot \\ \cdot \\ \gamma_p \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{bmatrix}$$

Un cop coneguts aquests $p+1$ elements $\{\gamma_k, k=0, \dots, p\}$, els altres elements de R_1 es poden calcular fàcilment utilitzant la primera equació de (2.9).

Per tant

$$R_1 = E[y^t (y^t)'] = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \dots & \dots \\ \cdot & \gamma_0 & \gamma_1 & \dots & \dots \\ \cdot & \cdot & \cdot & \dots & \dots \\ \cdot & \cdot & \cdot & \dots & \dots \end{bmatrix} =$$

$$= \sigma^2 \begin{bmatrix} f_0(a_1, \dots, a_p) & f_1(a_1, \dots, a_p) & \dots & \dots \\ \cdot & f_0(a_1, \dots, a_p) & \dots & \dots \\ \cdot & \cdot & \dots & \dots \\ \cdot & \cdot & \cdot & \dots \end{bmatrix} \quad (2.10)$$

és simètrica i té estructura de Toeplitz. Els diferents $\{\sigma^2 f_i(a_1, \dots, a_p), i=0, \dots, p\}$ són les solucions del sistema d'equacions ja esmentat en (2.9).

Si fem el canvi de notació $R_1 = \sigma^2 T_1$ on T_1 és la matriu definida a (2.10):

$$T_1 = \begin{bmatrix} f_0(a_1, \dots, a_p) & f_1(a_1, \dots, a_p) & \dots & \dots \\ \cdot & f_0(a_1, \dots, a_p) & \dots & \dots \\ \cdot & \cdot & \dots & \dots \\ \cdot & \cdot & \cdot & \dots \end{bmatrix}$$

arribem a la següent expressió per la funció de versemblança:

$$\begin{aligned}
 l(\emptyset, \sigma^2 / y^1) &\propto (\sigma^2)^{-t/2} (\det T_1)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2} (y^1)' T_1^{-1} (y^1)\right\} = \\
 &= (\det T_1)^{-1/2} \exp\left\{-\frac{t}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (y^1)' T_1^{-1} y^1\right\}
 \end{aligned}$$

(on \propto indica proporcional a) que està d'acord amb la definició donada per l'equació (2.6), és a dir, aquesta funció de versemblança és invariant per translació en el sentit donat en l'apartat 2.1.1 i en aquest cas Box i Tiao ([6], pàg. 28), aconsellen prendre la mètrica $\varphi(\emptyset, \sigma^2) = (\emptyset, \log \sigma^2)$, ja que la funció de versemblança en aquesta mètrica compleix la propietat de ser invariant per translació.

La justificació de l'elecció de la mètrica que donen Box i Tiao és la següent:

Volem trobar una mètrica pels paràmetres $\varphi(\emptyset)$ de manera que la funció de versemblança $l(\emptyset / \sigma^2, y^1)$ sigui invariant per translació. Per ser l'esmentada funció de versemblança gaussiana serà coneguda a priori excepte la seva localització que dependrà de y^1 . Es a dir la funció de versemblança $l(\emptyset / \sigma^2, y^1)$ és invariant per translació en la mètrica original $\varphi(\emptyset) = \emptyset$ i per tant, la distribució a priori no informativa és localment uniforme en \emptyset , és a dir:

$$p(\emptyset) \propto c.$$

A continuació veurem com triar una mètrica per la variància del soroll σ^2 , de manera que la distribució a priori per σ^2 sigui no informativa.

Si en comptes de treballar en la pròpia mètrica de σ^2 , treballem en la mètrica de $\log \sigma^2$, la funció de versemblança de $\log \sigma^2$ donats \emptyset i y^t serà:

$$l(\log \sigma^2 / y^t, \emptyset) \propto \exp \left\{ -\frac{t}{2} \log \sigma^2 - \frac{1}{2} \log \det T_t \right. \\ \left. - \frac{1}{2} \exp \left[-\log \sigma^2 - \log (y^t)' T_t^{-1} y^t \right] \right\}$$

i aquesta funció de versemblança compleix la definició donada per l'equació (2.6), és a dir, les dades $y^t = \{y_1, \dots, y_t\}$ només serveixen per traslladar la funció de versemblança. Per tant, la distribució a priori no informativa haurà de ser localment uniforme en $\log \sigma^2$. Si volem treballar en la mètrica original $\varphi(\sigma^2) = \sigma^2$, la densitat a priori no informativa haurà de ser proporcional localment a σ^{-2} , ja que

$$p(\sigma^2 / \emptyset) \propto \left[\frac{d \log \sigma^2}{d \sigma^2} \right] = \sigma^{-2}$$

Intuitivament, això vol dir que localment a priori les variàncies del soroll σ^2 petites tenen gran probabilitat i les variàncies del soroll σ^2 grans tenen poca probabilitat. Aquest raonament està d'acord amb el fet que el quocient ns^2 / σ^2 on s^2 és la variància mostral del soroll, segueix una χ^2 "Chi-quadrat".

El cas que nosaltres tractarem en tot aquest capítol és aquell en que els paràmetres \emptyset i la variància del soroll σ^2 siguin desconeguts. En aquest cas, suposarem que a priori \emptyset i σ^2 són independents i tindrem

$$p(\emptyset / \sigma^2) = p(\emptyset)$$

$$p(\sigma^2 / \emptyset) = p(\sigma^2)$$

i llavors la densitat a priori no informativa per \emptyset i σ^2 serà

$$p(\emptyset, \sigma^2) = p(\emptyset) p(\sigma^2)$$

Per tant, pel que fa als paràmetres, prendrem com a funció de densitat a priori una uniforme en l'interval adient i pel que fa a la variància del soroll, proporcional a $1/\sigma^2$, de manera que

$$p(\emptyset, \sigma^2) \propto k / \sigma^2 \tag{2.11}$$

on k és la constant de normalització.

A més aquesta densitat a priori $p(\emptyset, \sigma^2)$ sota la mètrica $\varphi(\emptyset, \sigma^2) = (\emptyset, \log \sigma^2)$ és no informativa d'acord amb l'apartat 2.1.2 ja que és localment proporcional al jacobià de la transformació que és:

$$|J| = \begin{vmatrix} \frac{\partial \varphi_1}{\partial \emptyset} & \frac{\partial \varphi_1}{\partial \sigma^2} \\ \frac{\partial \varphi_2}{\partial \emptyset} & \frac{\partial \varphi_2}{\partial \sigma^2} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 0 & \frac{1}{\sigma^2} \end{vmatrix} = \frac{1}{\sigma^2}$$

2.2.2 Predicció i filtratge

Per abordar el problema de l'estimació dels paràmetres del model AR(p) es determina l'evolució de la funció de densitat de probabilitat a posteriori $p(X_t / y^t)$, $t=1, \dots, n$. Suposem que en un instant t és coneguda $p(X_t / y^t)$ i que volem calcular $p(X_{t+1} / y^{t+1})$ en l'instant $t+1$. Aquest càlcul es farà en dues etapes:

La primera etapa anomenada de predicció ens permet calcular $p(X_{t+1}/y^t)$ a partir de $p(X_t / y^t)$. La segona etapa anomenada de filtratge ens ha de permetre calcular $p(X_{t+1} / y^{t+1})$ a partir de $p(X_{t+1} / y^t)$, quan es coneix la nova observació y_{t+1} .

L'etapa de predicció es resol mitjançant les equacions de Chapman-Kolmogorov per processos de Markov [26]

$$\begin{aligned} p_{X_{t+1}/y^t}(\zeta) &= \int_{X_t} p_{X_{t+1}/X_t, y^t}(\zeta/\eta) p_{X_t/y^t}(\eta) d\eta = \\ &= \int_{X_t} p_{X_{t+1}/X_t}(\zeta/\eta) p_{X_t/y^t}(\eta) d\eta \end{aligned} \tag{2.12}$$

La funció $p_{X_{t+1}/X_t, y^t}(\zeta/\eta)$ és igual a $p_{X_{t+1}/X_t}(\zeta/\eta)$ per ser el model dinàmic definit en (2.4) un procés de Markov [22], és a dir, conegut X_t , les observacions que formen y^t no aporten cap informació addicional sobre el valor de X_{t+1} .

La funció de densitat $p_{X_{t+1}/X_t}(\zeta/\eta)$ està relacionada amb la densitat de probabilitat de e_t .

L'etapa de filtratge es resol mitjançant el teorema de Bayes, tenint en compte la informació subministrada en l'etapa de predicció.

$$\begin{aligned}
 p_{X_{t+1}/y^{t+1}}(\zeta) &= \frac{p_{y_{t+1}/X_{t+1}, y^t}(y/\zeta) p_{X_{t+1}/y^t}(\zeta)}{\int_{X_{t+1}} p_{y_{t+1}/X_{t+1}, y^t}(y/\zeta) p_{X_{t+1}/y^t}(\zeta) d\zeta} = \\
 &= \frac{p_{y_{t+1}/X_{t+1}}(y/\zeta) p_{X_{t+1}/y^t}(\zeta)}{\int_{X_{t+1}} p_{y_{t+1}/X_{t+1}}(y/\zeta) p_{X_{t+1}/y^t}(\zeta) d\zeta}
 \end{aligned}
 \tag{2.13}$$

on y és el valor numèric que pren l'observació y_{t+1} ,

$p_{y_{t+1}/X_{t+1}, y^t}(y) = p_{y_{t+1}/X_{t+1}}(y)$ com es pot deduir de l'equació d'observació (2.4") ja que conegut X_{t+1} , les observacions que formen y^t no aporten informació addicional sobre el valor de y_{t+1} .

$p_{y_{t+1}/X_{t+1}}(y)$ s'obté de l'equació de l'observació (2.4"). Aquesta probabilitat serà una delta de Dirac, ja que un cop conegut l'estat X_{t+1} , l'esmentada equació (2.4") ens determina y_{t+1} .

Substituint en (2.13) p_{X_{t+1}/y^t} obtinguda en (2.12), aconseguim la recurrència:

rència:

$$p_{X_{t+1}/y^{t+1}}(\zeta) = \frac{p_{y_{t+1}/X_{t+1}}(y) \int_{X_t} p_{X_{t+1}/X_t}(\zeta/\eta) p_{X_t/y^t}(\eta) d\eta}{\int_{X_{t+1}} \delta(y - \zeta_t) \int_{X_t} p_{X_{t+1}/X_t}(\zeta/\eta) p_{X_t/y^t}(\eta) d\eta d\zeta} \quad (2.14)$$

D'aquesta forma, es va obtenir, coneguda la distribució a priori com a condició inicial, l'evolució temporal de la densitat de probabilitat de les variables a estimar condicionada per les observacions i a partir d'aquí calcularem $E(\theta / y^t)$ i $E(\sigma^2 / y^t)$ que no són res més que els primers moments de les distribucions marginals. També podem calcular les variàncies corresponents.

El conjunt constitueix un algorisme recursiu que té com a principal dificultat la representació i emmagatzemat de la densitat de probabilitat.

2.3 ESTIMACIÓ D'UN MODEL AR(1) AMB σ^2 DESCONEGUT.

2.3.1 Formulació

El model AR(1)

$$y_{t+1} = a y_t + e_{t+1}$$

format en notació d'espai d'estat és

$$x_{t+1} = a x_t + e_{t+1}$$

$$y_t = x_t$$

Ampliem l'estat x_t a l'estat

$$X_t = \begin{bmatrix} x_t \\ a \\ \sigma^2 \end{bmatrix}$$

i les equacions d'estat no lineals són:

$$\begin{bmatrix} X_{t+1}^1 \\ X_{t+1}^2 \\ X_{t+1}^3 \end{bmatrix} = \begin{bmatrix} X_t^2 & X_t^1 \\ X_t^2 \\ X_t^3 \end{bmatrix} + \begin{bmatrix} e_{t+1} \\ 0 \\ 0 \end{bmatrix}$$

$$y_t = X_t^1$$

amb les condicions inicials

$$\begin{bmatrix} X_0^1 \\ X_0^2 \\ X_0^3 \end{bmatrix} = \begin{bmatrix} x_0 \\ a \\ \sigma^2 \end{bmatrix}$$

de les quals introduïrem en el filtre la distribució a priori d'acord amb l'apartat 2.2.1.

2.3.2 Implementació de l'algorisme de filtratge no lineal

Donada la dificultat d'obtenir solucions analítiques de l'equació recurrent plantejada en (2.14), cal implementar mitjançant un ordinador una solució numèrica general. Per fer-ho, es representen les funcions de densitat de probabilitat per un nombre finit de valors, és a dir es discretitza la densitat de probabilitat i s'avalua la funció en el conjunt de punts de discretització, anomenat malla.

En aquest cas concret s'ha realitzat una discretització inicial a les variables que componen l'espai d'estat $[x_0, a, \sigma^2]$; per tant, s'obté una malla cúbica.

L'equació recurrent (2.12) s'implementa calculant per cada punt de la malla, la integral del producte de dues funcions; el resultat de la integració cal multiplicar-ho per una altra funció coneguda i normalitzar-ho per a que es pugui interpretar com una densitat de probabilitat (2.13). Les integrals de (2.12) i (2.13) s'aproximen per sumes.

La malla és mòbil, és a dir, es va adaptant a la densitat de probabilitat a mesura que la variància d'aquesta canvia. Per l'adaptació de la malla seguirem les següents etapes:

1) Inicialitzacions, adopció d'una malla inicial.

2) Càlcul de la probabilitat a cada punt de la malla segons l'algorisme descrit en l'apartat 2.2.2 i representat per la fórmula (2.14).

3) Comprovar que tots els punts de la malla tenen associat un valor de la probabilitat superior a una certa tolerància. Els punts que no compleixin aquesta condició són eliminats.

4) Comprovar que el nombre total de punts en cada dimensió és superior a un valor prefixat. Si resulta ser més petit que l'esmentat valor, es canvia l'interval de discretització i s'interpolava la funció de densitat de probabilitat pels nous punts.

5) Tornar a 2)

2.3.2.1 Inicialització de la recursió.

Hem construït la següent malla inicial

$X_0^1 \in [-R, R]$ i es discretitza aquest interval en k_1 punts.

$X_0^2 \in (-1, 1)$ i es discretitza aquest interval en k_2 punts.

$X_0^3 \in [S, T]$ i es discretitza aquest interval en k_3 punts.

Per tant, es té una malla inicial de $k_1 \times k_2 \times k_3$ punts sobre la qual s'inicialitza la densitat de probabilitat a priori.

La densitat a priori de les condicions inicials està d'acord amb l'equació (2.11):

$$p_{X_0^1, X_0^2, X_0^3}(\zeta_1, \zeta_2, \zeta_3) \propto 1 / (k_1 \times k_2 \times \zeta_3^2) \quad \text{si} \quad \begin{cases} \zeta_1 \in [-R, R] \\ \zeta_2 \in (-1, 1) \\ \zeta_3 \in [S, T] \end{cases}$$

L'interval $[-R, R]$ per a X_0^1 es pren inicialment a partir d'una primera aproximació de l'ordre de magnitud dels valors, i a partir d'aquí es dona la distribució a priori de X_0^1 en el moment d'iniciar el funcionament del filtre. Per a X_0^2 es pren l'interval $(-1, 1)$ perquè el pol d'un model AR(1) estacionari pertany sempre a l'esmentat interval. Finalment es pren l'interval $[S, T]$ per a X_0^3 en funció de l'ordre de magnitud de la variància del soroll a estimar.

En la primera iteració, al fer la predicció s'obté la probabilitat definida damunt tota la malla, al fer el filtratge aquesta probabilitat queda concentrada sobre la recta $X_0^1 = y_1$ i la probabilitat damunt tots els altres punts de la malla passa a ser zero. Per tant, a partir de la segona iteració, la predicció i el filtratge es poden fer en un únic pas. En conseqüència, si es pren una malla inicial de $k_1 \times k_2 \times k_3$, a partir de la segona iteració quedarà reduïda a $k_2 \times k_3$ i l'algorisme té un guany notable en rapidesa i capacitat de memòria.

2.3.2.2 Càlcul de la predicció.

Tal com ja s'ha esmentat, aproximem les integrals de les equacions de Chapman-Kolmogorov (3.12) per sumes

$$p_{x_{t+1}^1, x_{t+1}^2, x_{t+1}^3}(\zeta_1, \zeta_2, \zeta_3 / y^t) = \sum_{\eta_1} \sum_{\eta_2} \sum_{\eta_3} p_{x_{t+1}^1, x_{t+1}^2, x_{t+1}^3 / x_t^1, x_t^2, x_t^3}(\zeta_1, \zeta_2, \zeta_3 / \eta_1, \eta_2, \eta_3) p_{x_t^1, x_t^2, x_t^3}(\eta_1, \eta_2, \eta_3 / y^t)$$

Ara bé

$$p_{x_{t+1}^1, x_{t+1}^2, x_{t+1}^3 / x_t^1, x_t^2, x_t^3}(\zeta_1, \zeta_2, \zeta_3 / \eta_1, \eta_2, \eta_3) = p_{a_{t+1}}(\zeta_1 - \eta_1, \eta_2) \delta(\zeta_2 - \eta_2) \delta(\zeta_3 - \eta_3) =$$

$$\begin{cases} k \frac{1}{\eta_3} \exp \left\{ -\frac{1}{2} \frac{(\zeta_1 - \eta_1 \eta_2)^2}{\eta_3^2} \right\} & \text{si } \zeta_2 = \eta_2, \zeta_3 = \eta_3 \\ 0 & \text{altrament} \end{cases}$$

on k és la constant de normalització.

Per tant

$$p_{x_{t+1}^1, x_{t+1}^2, x_{t+1}^3 / y^t}(\zeta_1, \zeta_2, \zeta_3) = \begin{cases} k \sum_{\eta_1} p_{x_t^1, x_t^2, x_t^3 / y^t}(\eta_1, \eta_2, \eta_3) \exp \left\{ -\frac{1}{2} \left\{ \frac{\zeta_1 - \eta_1 \eta_2}{\eta_3} \right\}^2 \right\} & \text{si } \zeta_2 = \eta_2, \zeta_3 = \eta_3 \\ 0 & \text{altrament} \end{cases}$$

2.3.2.3 Càlcul del filtratge

Anàlogament es calcula (2.13)

$$\begin{aligned} p_{x_{t+1}^1, x_{t+1}^2, x_{t+1}^3 / y^{t+1}}(\zeta_1, \zeta_2, \zeta_3) &= \\ &= k p_{x_{t+1}^1, x_{t+1}^2, x_{t+1}^3 / y^t}(\zeta_1, \zeta_2, \zeta_3) p_{y_{t+1} / x_{t+1}^1, x_{t+1}^2, x_{t+1}^3}(y) \end{aligned}$$

on k és la nova constant de normalització ja que el denominador de (2.13) és una constant respecte a les variables ζ .

$$p_{y_{t+1} / x_{t+1}^1, x_{t+1}^2, x_{t+1}^3}(y) = \begin{cases} 1 & \text{si } y = \zeta_1 \\ 0 & \text{altrament} \end{cases}$$

Per tant

$$p_{x_{t+1}^1, x_{t+1}^2, x_{t+1}^3 / y^{t+1}}(\zeta_1, \zeta_2, \zeta_3) = \begin{cases} k p_{x_{t+1}^1, x_{t+1}^2, x_{t+1}^3 / y^t}(\zeta_1, \zeta_2, \zeta_3) & \text{si } y = \zeta_1 \\ 0 & \text{altrament} \end{cases}$$

2.3.2.4 Càlcul de la predicció i filtratge conjuntament.

Tal com ja s'ha explicat en l'apartat 2.3.2.1, a partir de la segona iteració es calcula la predicció i el filtratge conjuntament, el que ens permet obtenir la recurrència:

$$p_{x_{t+1}^2, x_{t+1}^3 / y^{t+1}}(\zeta_2, \zeta_3) = k \exp \left\{ -\frac{1}{2} \left\{ \frac{y_{t+1} - y_t \zeta_2}{\zeta_3} \right\}^2 \right\} p_{x_t^2, x_t^3 / y^t}(\zeta_2, \zeta_3) \quad (2.15)$$

2.4 **RESULTATS**

2.4.1 **Resultats**

S'han simulat vàries successions d'observacions segons un model AR(1) utilitzant el generador de nombres aleatoris gaussians NRANDOM existent en el paquet estadístic MINITAB [29]. En totes les simulacions s'ha pres la mitjana del soroll igual a zero i la variància igual a la unitat.

Es presenten aquí dues successions d'observacions que es consideren les més representatives. Aquestes dues successions corresponen als models en que el pol és 0.8 i 0.05 respectivament. Les funcions de densitat conjunta a posteriori del pol i de la variància del soroll són obtingudes implementant la recurrència (3.15), i a partir d'ella s'han calculat les funcions de densitat a posteriori marginals pel pol i per la variància del soroll respectivament. En aquest cas la malla inicial és $k_1=121$, $k_2=21$, $k_3=21$. A partir de la segona iteració la malla té 21x21 punts.

Les figures 1 i 2 són les funcions de densitat de probabilitat a posteriori després de N observacions ($N=100$, 1000, 10000) en les que els vertaders valors del pol són 0.8 i 0.05 respectivament. La malla inicial utilitzada pels càlculs va des de $(-1+1.e-37)$ fins a $(+1-1.e-37)$, el pas inicial és 0.1, per tant l'esmentada malla inicial té en total 21 punts.

Noti's que en les figures 1 i 2, les escales són diferents

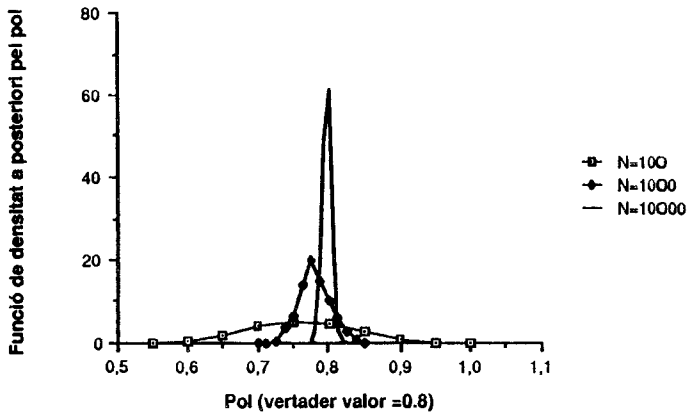


Figura 1. Densitat de probabilitat a posteriori del pol a. (El vertader valor és 0.8)

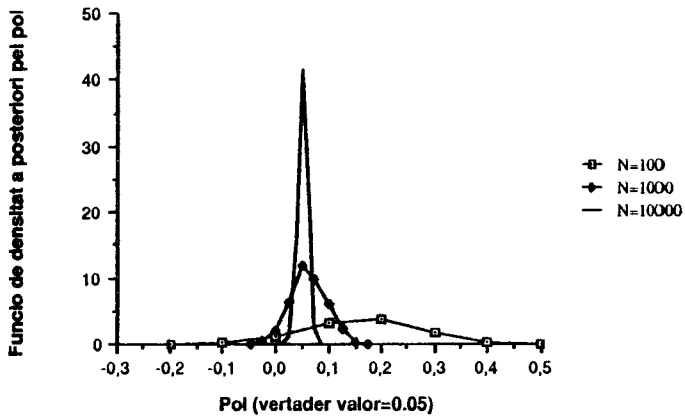


Figura 2. Igual que figura 1 però el vertader valor del pol és 0.05

En primer lloc ens ocuparem de l'evolució de l'estimació del pol.

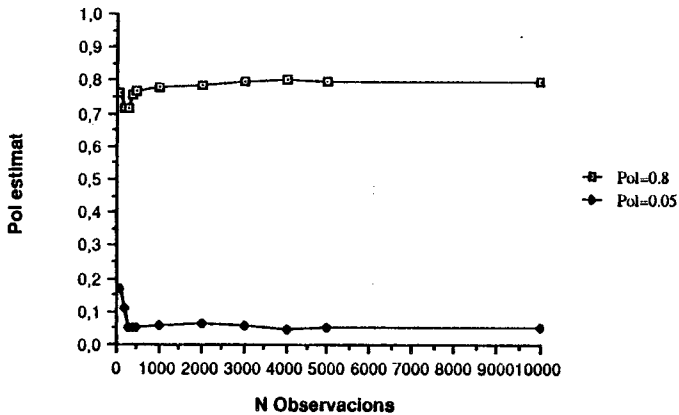


Figura 3. Estimació del pol pels diferents valors del vertader pol 0.8 i 0.05 respectivament.

Com es pot veure en aquesta figura, la convergència cap el vertader valor és més ràpida quan el pol és 0.8 que quan el pol és 0.05. Per altra banda també es posa en evidència d'una manera clara que a partir de 5000 observacions no hi ha gaire diferència entre les dues gràfiques, però quan el nombre d'observacions és més petit de 1000, les diferències entre ambdues gràfiques són substancials.

A continuació ens ocuparem de l'evolució de la desviació tipus de l'estimador del pol.

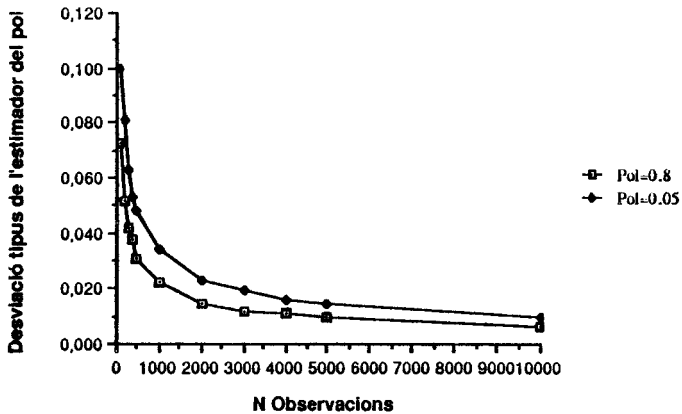


Figura 4. Desviació tipus de l'estimador del pol quan els vertaders valors són 0.8 i 0.05 respectivament.

S'observa en la figura 4 que la desviació tipus de l'estimador del pol és més gran quan el pol és a prop de zero que quan el pol és a prop del cercle unitat.

La densitat de probabilitat a posteriori de la variància del soroll després de N observacions ($N=100, 1000, 10000$) quan el vertader valor del pol és 0.8 i 0.05 és a les figures 5 i 6 respectivament. La malla inicial pels càlculs va des de 0.1 fins a 10, el pas inicial és 0.5, per tant en total hi ha 21 punts.

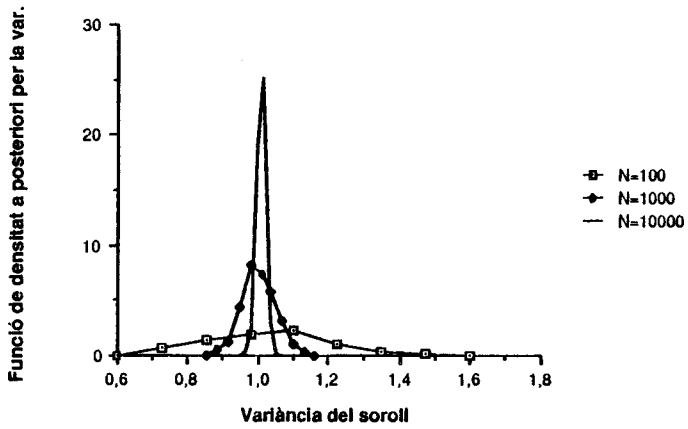


Figura 5. Densitat de probabilitat a posteriori de la variància del soroll. El vertader valor del pol és 0.8

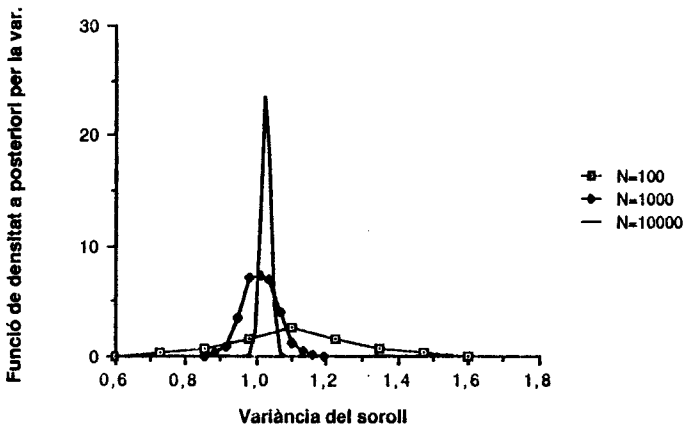


Figura 6. Igual que figura 5 però el vertader valor del pol és 0.05

A la figura 7 s'observa l'evolució de l'estimació de la variància del soroll pels diferents valors del pol 0.8 i 0.05. Com ja s'ha dit, el vertader valor del pol és desconegut a priori i s'estima simultàniament mitjançant el filtre no lineal.

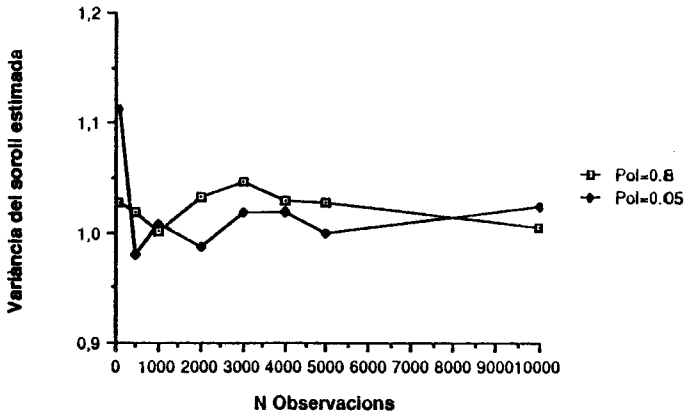


Figura 7. Estimació de la variància del soroll segons els diferents valors del pol: 0.8 i 0.05 respectivament.

Es pot observar en aquesta figura que l'estimació de la variància del soroll, en funció de la localització en el cercle unitat del vertader valor del pol, difereix molt poc quan el nombre d'observacions disponibles és petit ja que la diferència màxima entre ambdues estimacions és d'una dècima, i a partir de 1000 observacions és pràcticament la mateixa.

Finalment a la figura 8 hi ha l'evolució de la desviació tipus de l'estimació de la variància del soroll. Es pot veure que les dues gràfiques obtingudes segons els diferents valors del pol, gairebé coincideixen, resultat que està d'acord amb les afirmacions fetes en el paràgraf anterior.

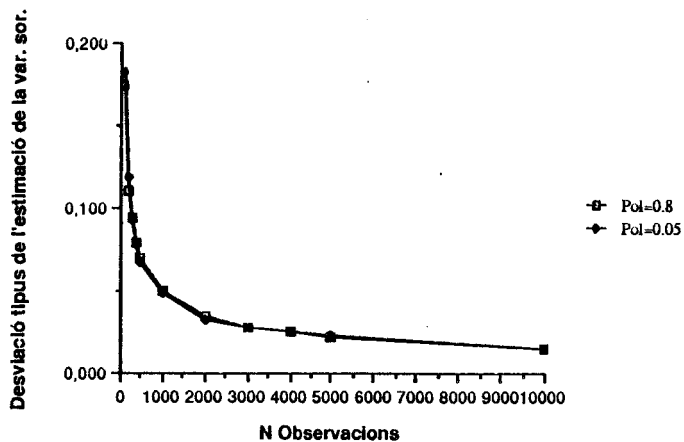


Figura 8. Desviació tipus de l'estimació de la variància del soroll segons els diferents valors del pol: 0.8 i 0.05 respectivament.

Resumint les figures 5, 6, 7 i 8, es pot afirmar que l'estimació de la variància del soroll no sembla estar influenciada pel valor del pol.

2.4.2 Comparació amb el comportament asimptòtic.

Segons [5], [18], la desviació tipus de la distribució asimptòtica del paràmetre estimat del model AR(1) és

$$\text{des } \hat{a}_N \approx \sqrt{\frac{1}{N} (1 - a^2)} \quad (2.9)$$

on N és la grandària de la mostra, \hat{a} és el paràmetre estimat i a és el paràmetre del model generat.

S'han calculat les desviacions tipus de la distribució assimptòtica segons els diferents tamanys de mostra d'acord amb (2.9) per les dues successions ja esmentades i s'han comparat amb les desviacions tipus de l'estimador calculades segons el filtre no lineal. Els resultats són posats de manifest a les figures 9 i 10.

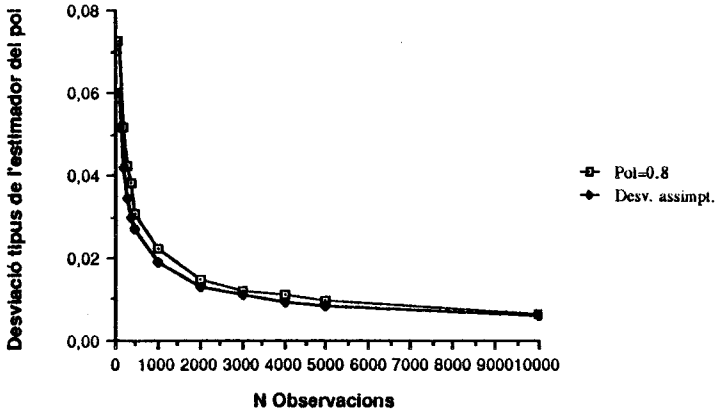


Figura 9.
 ■ Desviació tipus de l'estimador calculada segons el filtre no lineal.
 ▲ Desviació tipus assimptòtica de l'estimador
 El vertader valor del pol és 0.8

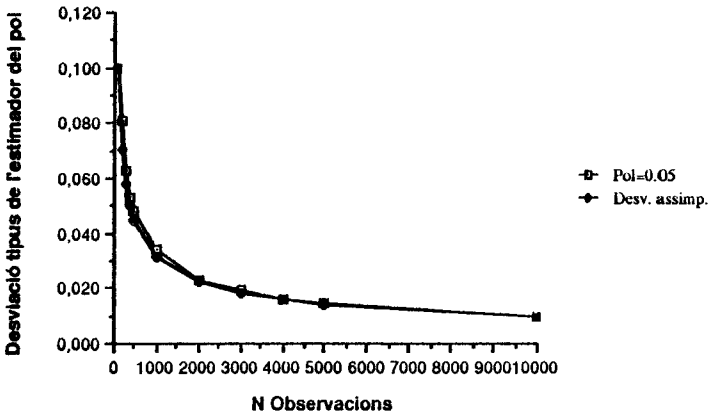


Figura 10. Igual que figura 9 però el vertader valor del pol és 0.05

2.5. CONCLUSIONS

Les conclusions són de dos tipus ben diferenciats. Per una banda treurem les conclusions numèriques de la implementació del filtre no lineal i d'altra banda ens ocuparem dels aspectes estadístics de l'estimació del model AR(1) amb σ^2 desconegut.

El filtre no lineal és una tècnica útil per a investigar la dependència de la precisió respecte del nombre d'observacions. Els problemes més grans són el temps de càlcul i la memòria necessària que fan difícil de plantejar el cas general del model ARMA(p,q). Si a la densitat de probabilitat es volen guardar N_n punts en cada dimensió, cal emmagatzemar $N_n^{(2p+q+1)}$ valors de la densitat de probabilitat ja que d'acord amb les equacions (2.4') i (2.4'') calen p variables d'estat més (p+q) variables corresponents als paràmetres del model ARMA(p,q) més una dimensió més corresponent a la variància del soroll. D'altra banda, s'han de calcular $p \cdot N_n^{(2p+q+1)}$ operacions (sumes i productes), és a dir $N_n^{(3p+q+1)}$ operacions a cada pas. Per tant, si tenim t observacions, el temps de càlcul necessari serà de l'ordre de $t \cdot N_n^{(3p+q+1)}$. Això fa que aquest filtre no es pugui utilitzar exceptuant si p i q són petits.

L'algorisme que nosaltres hem implementat per l'estimació d'un model AR(1) amb σ^2 desconegut té un guany notable en capacitat de memòria i temps de càlcul. Tal com ja s'ha dit en l'apartat 2.3.2.4, al fer la predicció i filtratge conjuntament, ja no calen emmagatzemar $N_n^{(2+1)}$ valors de la densitat de probabilitat sinó que només és necessari tenir memòria disponible de dimensió $N_n^{(1+1)}$. Per altra banda, d'acord amb l'equació (2.15), el nombre d'operacions a cada pas és de l'ordre de $N_n^{(1+1)}$ en

comptes de $N_n^{(3+1)}$ operacions necessàries si no es fes la predicció i el filtratge conjuntament.

Respecte a l'estimació dels paràmetres del model ja esmentat, els resultats recolzen el fet que aquesta estimació depèn de la grandària de la mostra i de la localització del pol en el cercle unitat, ja que com es pot veure en l'apartat anterior el model simulat amb el pol en el 0.05 necessita més observacions per la correcta estimació del paràmetre que el model simulat amb el pol en el 0.8.

També es pot observar que l'estimació de la variància del soroll ve poc influenciada per la situació del pol.

La desviació tipus de la distribució asimptòtica del paràmetre és superior en el primer cas que en el segon, per tant calen més observacions per a estimar el paràmetre amb la mateixa precisió en el cas en que el pol és 0.05 que en el cas en que el pol és 0.8. A més la variabilitat dels resultats observats en el primer és major que en el segon cas, com era de preveure. Per tant és obvi que caldran més observacions en el primer cas que en el segon per la correcta estimació del paràmetre. Així doncs, l'interval de confiança de les estimacions reals coincideix pràcticament amb l'obtingut utilitzant els valors asimptòtics.

