



Universitat de les  
Illes Balears



Treball Final de Grau

GRAU DE MATEMÀTIQUES

## Anàlisi de sèries temporals

ANTÒNIA HORRACH MORELL

**Tutor**

Jaume Suñer Llabrés

Escola Politècnica Superior  
Universitat de les Illes Balears  
Palma, 8 de setembre de 2016



# SUMARI

<b>Sumari</b>	<b>i</b>
<b>Resum</b>	<b>iii</b>
<b>1 Introducció</b>	<b>1</b>
<b>2 Modelització per components</b>	<b>5</b>
2.1 Tendència: mitjanes mòbils . . . . .	8
2.2 Estacionalitat . . . . .	11
2.3 Autocorrelació . . . . .	17
2.3.1 Correlograma . . . . .	18
<b>3 Models probabilístics</b>	<b>21</b>
3.1 Procés estocàstic . . . . .	22
3.1.1 Estimació de la correlació . . . . .	23
3.2 Estacionarietat . . . . .	24
3.3 Models . . . . .	28
3.3.1 Soroll blanc . . . . .	29
3.3.2 Passejada aleatòria . . . . .	30
3.3.3 Mitjanes Mòbils (MA) . . . . .	31
3.3.4 Processos Autoregressius (AR) . . . . .	34
3.3.5 La funció d'autocorrelació parcial (PACF) . . . . .	38
3.3.6 Processos ARMA . . . . .	45
3.3.7 Processos ARIMA . . . . .	47
3.4 Validació del model . . . . .	48
3.5 Exemple . . . . .	50
3.6 Prediccions . . . . .	60
3.7 Altres models . . . . .	61
<b>4 Conclusions</b>	<b>63</b>
<b>5 Apèndix</b>	<b>65</b>
<b>Bibliografia</b>	<b>71</b>



## RESUM

L'objectiu d'aquest treball és conèixer algunes de les tècniques utilitzades en l'anàlisi de **sèries temporals**. El treball s'ha dividit en dues parts; a la primera part, s'estudien les tècniques clàssiques, que consisteixen en descompondre la sèrie temporal en les seves components: la tendència, que descriu el comportament mitjà de la variable a llarg plaç; la component estacional, que són moviments repetitius que es produeixen al llarg de la trajectòria de la sèrie; la component cíclica, que són oscil·lacions periòdiques que es produeixen a llarg termini; i la component aleatòria, que recull l'aportació aleatòria que s'atribueix a l'atzar.

Una vegada s'ha descompost la sèrie, s'aconsegueix un model per a la sèrie temporal amb la modelització de la tendència i la component estacional i cíclica. En el cas additiu, obtenim el model amb la suma de les modelitzacions anteriors, i en el cas multiplicatiu amb el producte de les modelitzacions i dividint per 100. A través del model es poden fer prediccions dels valors futurs de la sèrie, tot i que els resultats obtinguts no seran exactes ja que el model no té en compte la part aleatòria. Per millorar aquesta modelització de la sèrie temporal tenim la segona part.

La segona part estudia la tècnica coneguda com el mètode de Box-Jenkins, que consisteix en modelitzar la sèrie mitjançant un procés estocàstic que intenta recollir el comportament aleatori de la sèrie. Així, s'han estudiat els processos: soroll blanc, passejada aleatòria, **processos autoregressius** i **processos de mitjanes mòbils**, i la unió d'aquests dos, els anomenats processos ARMA i **ARIMA**.

Per tal de trobar el model que millor s'ajusti a la sèrie temporal s'ha de seguir el següent procés: primer, s'ha de tenir la sèrie estacionària, s'aconsegueix diferenciant  $d$  vegades. Amb el test de Dickey-Fuller augmentat es pot comprovar l'**estacionarietat** de la sèrie temporal. El següent pas és seleccionar el model ARIMA apropiat, és a dir, trobar els valors de  $p$  i  $q$  més adequats per a un ARIMA( $p,d,q$ ). Una vegada tinguem un conjunt de models candidats a modelitzar la sèrie temporal, s'haurà de triar el millor. Per això existeixen diferents criteris, per exemple, el criteri d'Akaike (AIC) i el criteri d'Informació Bayesiana (BIC). Una vegada s'ha triat el millor model, s'ha de validar si és adequat per modelitzar la sèrie temporal donada. Concretament, una sèrie estarà ben identificada si quan s'ajusta el model els residus es comporten com un soroll blanc. Un test per comprovar si els residus són un soroll blanc és el test de Ljung-Box. Finalment, podem fer prediccions dels valors futurs de la sèrie temporal amb el model obtingut.

L'estudi ha estat completat amb exemples reals, als que hem aplicat les tècniques estudiades.



## INTRODUCCIÓ

Una sèrie temporal és una seqüència d'observacions mesurades en diferents moments de temps, ordenades cronològicament i espaiades entre sí de manera uniforme [1].

Les sèries temporals ja van jugar un paper important en les primeres ciències naturals. L'astronomia babilònica va utilitzar sèries temporals de les posicions relatives de les estrelles i planetes per predir esdeveniments astronòmics.

L'anàlisi de sèries temporals ajuda a detectar regularitats en les observacions d'una variable i derivar "lleis" d'elles, i/o explotar tota la informació inclosa en aquesta variable per predir millor l'evolució futura. La idea metodològica bàsica darrera aquests procediments, que també eren vàlids per als babilonis, és que és possible descompondre sèries de temps en un nombre finit de components independents però no directament observables que es desenvolupen amb regularitat i per tant es pot calcular en endavant. Per aquest procediment, cal que hi hagi diferents factors independents que tenen impacte en la variable.

A mitjans del segle XIX, aquest enfocament metodològic per a l'astronomia va ser considerat pels economistes Charles Babbage i William Stanley Jevons. La descomposició en components no observables que depenen de diferents factors causals, com es fa servir en l'anàlisi clàssica de sèries temporals, va ser desenvolupat per Warren M. Persons (1919). Es van distingir quatre components diferents:

- La tendència.
- Els cicles.
- L'estacionalitat.
- Els residus.

L'anàlisi clàssica de sèries temporals suposa que els components sistemàtics, és a dir, la tendència, els cicles i l'estacionalitat, no es veuen afectats per perturbacions estocàstiques i per tant poden ser representats per funcions deterministes de temps.

L'impacte estocàstic es limita als residus, que d'altra banda, no contenen cap moviment sistemàtic. Per tant, es modela com una sèrie de variables independents o no correlacionades amb esperança zero i variància constant, és a dir, com un procés aleatori pur.

No obstant això, des de la dècada de 1970, cada vegada més es va anar aplicant un enfocament totalment diferent a l'anàlisi de sèries de temps. Els procediments purament descriptius d'anàlisi clàssica de sèries temporals van ser abandonats i, en canvi, es van emprar els resultats i mètodes de la teoria de la probabilitat i l'estadística. Mentre que l'enfocament clàssic considera que els residus no tenen cap significat per a l'estructura de la sèrie de temps, l'enfocament modern assumeix que hi ha impactes estocàstics sobre tots els components d'una sèrie temporal.

Els primers passos en aquesta direcció van ser presos per l'estadístic de Rússia Evgenij Evgenievich Slutsky i l'estadístic britànic George Udny Yule a principis del segle passat. Tots dos van desenvolupar els processos de mitjanes mòbils i autoregressius com a models per a representar sèries temporals. Herman Wold (1938) va sistematitzar i generalitzar aquests enfocaments en la seva tesi doctoral. El seu ús pràctic generalitzat es deu a George E. P. Box i Gwilym M. Jenkins (1970), que van desenvolupar mètodes per implementar aquests models empíricament. Van assumir que hi havia un model estocàstic comú per a tot procés de generació de sèries temporals.

Box i Jenkins van publicar un llibre de text sobre anàlisi de sèries temporals que va rebre una considerable atenció. En primer lloc, s'introdueixen models univariants de sèries temporals de manera que simplement feia ús sistemàtic de la informació inclosa en els valors observats de la sèrie temporal. Això ofereix una manera fàcil de predir el desenvolupament futur d'aquesta variable. Avui en dia, el procediment es coneix com anàlisi de Box-Jenkins i s'aplica àmpliament.

D'altra banda, des de la dècada de 1980 la possible no estacionarietat de les sèries cada vegada es va prendre amb més consideració. La no estacionarietat no només podia ser causada de forma determinista sinó també per tendències estocàstiques i, a més, l'estacionarietat de la sèrie temporal ja no s'eliminava simplement a través de l'aplicació de filtres per tal de seguir en el marc de models estacionaris. En conseqüència, després d'aquesta introducció dels principis bàsics, primer es tracten els models de sèries temporals estacionàries i després la modelització de sèries temporals no estacionàries [2].

Es presentaran diferents criteris metodològics que permeten l'estudi de fenòmens, i la previsió de l'evolució futura.

Dedicarem el segon capítol a la modelització de sèries temporals per components, que s'anomena l'anàlisi clàssica de les sèries temporals. Ens centrarem en com extreure les diferents components: tendència, estacionalitat i residus, de la sèrie temporal. I com trobar un model que finalment ens permetrà fer prediccions futures d'aquesta sèrie.

Al tercer capítol ens centrarem en la modelització probabilística de les sèries temporals, que s'anomena l'anàlisi de les sèries temporals a través de la metodologia Box-Jenkins. Explicarem quins processos s'han de seguir per obtenir una sèrie de models candidats a modelitzar la nostra sèrie, llavors quins criteris s'han de seguir per trobar quin és el millor d'aquests candidats. I finalment, veure si el model que s'ha



---

elegit s'ajusta correctament a la nostra sèrie. Per això es definiran diferents tipus de models, cada un amb unes característiques pròpies.

En cada capítol es mostraran diferents exemples per posar en pràctica el que s'ha exposat al llarg de tot el capítol.

Acabarem amb una secció que recollirà les conclusions més importants d'aquest treball.



## MODELITZACIÓ PER COMPONENTS

**Definició 2.0.1** *Una sèrie temporal és una seqüència d'observacions mesurades en diferents moments de temps, ordenades cronològicament i espaiades entre sí de manera uniforme [1].*

El mètode clàssic d'anàlisi de sèries temporals consisteix en intentar separar la sèrie en diferents components i modelitzar cada una per separat. S'estableixen quatre components teòriques, que són:

- **Tendència:** es pot definir com un canvi a llarg termini de la sèrie que es produeix en la mitjana. Descriu el comportament mitjà de la variable a llarg plaç. S'expressa com una funció del temps de tipus polinòmic o logarítmic.
- **Variacions estacionals:** moviments repetitius que es produeixen al llarg de la trajectòria de la sèrie i generalment representen les fluctuacions que es registren de forma constant en períodes de temps generalment inferiors a l'any.
- **Variacions cícliques:** són oscil·lacions periòdiques que es produeixen a llarg termini i sovint van lligades a etapes de prosperitat o depressió econòmica. Solen ser més difícils d'identificar quant més llarg és el seu període, degut fonamentalment a que el temps de recollida d'informació no aporta suficients dades. Per això, a vegades quedaran confoses amb les altres components.
- **Residus:** recull l'aportació aleatòria de qualsevol fenomen subjecte a l'atzar. És un component que no segueix cap norma ni tendència, s'atribueix a l'atzar.

La figura 2.1 mostra una sèrie temporal i les diferents components que la formen.

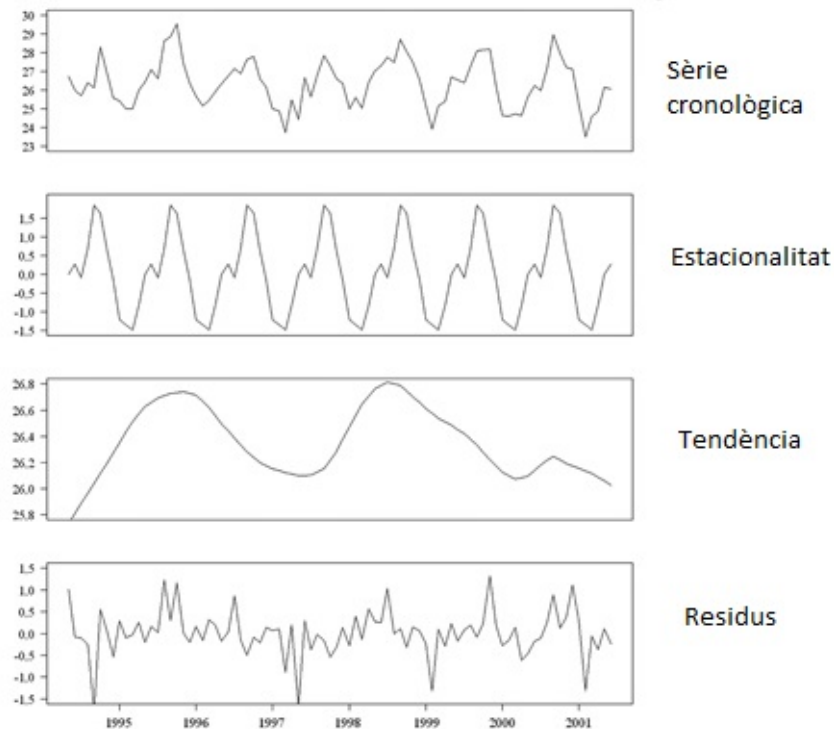


Figura 2.1: Components d'una sèrie cronològica

Els mètodes que descomponen la sèrie original en les diferents components de variació de dades consideren que aquestes es generen a partir d'un model de la forma

$$X_t = f(T_t, E_t, R_t)$$

essent  $X_t$  el valor de la sèrie en el temps  $t$ ,  $T_t$ ,  $E_t$  i  $R_t$  les components de tendència-cicle, estacional i aleatòria en el temps  $t$  respectivament. L'objectiu d'aquests mètodes és identificar i aïllar, de la forma més precisa possible, cada una de les components de variació presents a la sèrie a partir de les dades observades  $X_1, \dots, X_N$ .

Les formes més habituals de la funció  $f$  són la suma i/o el producte de les components. Així, els models de descomposició més freqüents són:

- Model additiu:  $X_t = T_t + E_t + R_t$ ,
- Model multiplicatiu:  $X_t = T_t \times E_t \times R_t$ ,
- Model mixt:  $X_t = T_t \times E_t + R_t$

Alguns autors consideren que el model multiplicatiu és el que nosaltres anomenem mixt. En qualsevol cas un model multiplicatiu sempre pot convertir-se en additiu a través d'una transformació logarítmica, és a dir, si

$$X_t = T_t \times E_t \times R_t$$

prenent logaritmes obtenim que

$$\log X_t = \log T_t + \log E_t + \log R_t,$$

---

és dir, un model additiu per a la sèrie  $\log X_t$ .

Per tal de decidir si la sèrie a estudiar mostra un comportament additiu o multiplicatiu, un primer pas consisteix en una identificació visual: si el patró estacional es manté amb amplitud constant, és a dir, la magnitud de les fluctuacions estacionals no varien amb el temps, es tractarà d'un model additiu (figura 2.2); i si el patró es va amplificant amb el temps, el model serà multiplicatiu (figura 2.3).

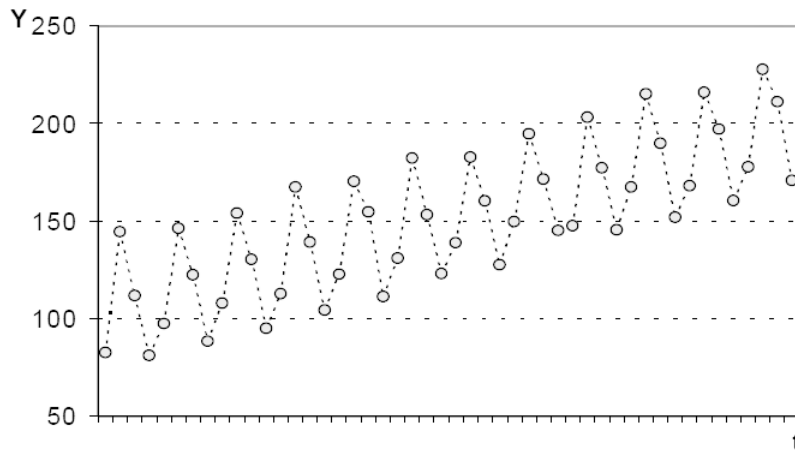


Figura 2.2: Sèrie additiva

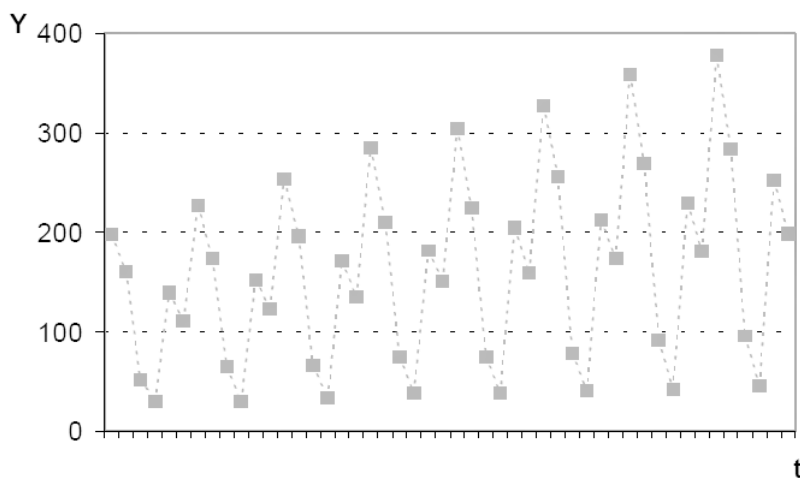


Figura 2.3: Sèrie multiplicativa

A continuació veurem els procediments que s'han de seguir per separar la tendència i l'estacionalitat seguint un model additiu o multiplicatiu. Els canvis cíclics no els tindrèm en compte ja que es tracta de períodes repetitius a llarg termini i es necessiten moltes dades per poder apreciar-los.

## 2.1 Tendència: mitjanes mòbils

Amb aquest mètode s'aconsegueixen suavitzar tant les oscil·lacions periòdiques d'una sèrie com les aleatòries. La seva aplicació, però, requereix decidir prèviament el període en què es repeteix cert patró de comportament. L'observació de l'evolució de la gràfica de la sèrie pot ajudar a prendre la decisió, no obstant, més endavant veurem mètodes més efectius per determinar aquest període.

Una vegada fixat el període  $p$ , es calculen les mitjanes mòbils de la sèrie preses de  $p$  en  $p$ , successivament, des de l'inici. Vegem la definició de les mitjanes mòbils.

**Definició 2.1.1 (Mitjanes mòbils)** Sigui  $\{X_t\}_{t=1}^{t=N}$  una successió de valors i sigui  $p$  un nombre senar. Aleshores direm procés de mitjanes mòbils d'ordre  $p$  a la col·lecció de valors  $\{\bar{X}_t\}_{t=\frac{p+1}{2}}^{t=N-\frac{p+1}{2}}$  donats per

$$\bar{X}_t = \frac{\sum_{i=-(p-1)/2}^{(p-1)/2} X_{t+i}}{p}.$$

Vegem alguns casos particulars. Suposem que  $p = 5$ , aleshores

$$\bar{X}_t = \frac{X_{t-2} + X_{t-1} + X_t + X_{t+1} + X_{t+2}}{5}$$

i si donam valors a  $t$ , tenint en compte que  $t \in \{3, \dots, N-3\}$ , obtenim

$$\begin{aligned} \bar{X}_3 &= \frac{X_1 + X_2 + X_3 + X_4 + X_5}{5} \\ \bar{X}_4 &= \frac{X_2 + X_3 + X_4 + X_5 + X_6}{5} \\ &\vdots \\ \bar{X}_{N-3} &= \frac{X_{N-5} + X_{N-4} + X_{N-3} + X_{N-2} + X_{N-1}}{5} \end{aligned}$$

Observem que, en el cas que  $p$  sigui parell, el centre del grup de cada  $p$  valors dels que feim la mitjana correspon a un valor no observat del temps. Per arreglar-ho, la nova sèrie queda constituïda per les mitjanes aritmètiques de les mitjanes mòbils preses de dos en dos,

$$\begin{aligned} \bar{X}_{(p+2)/2} &= \frac{\bar{X}_{(p+1)/2} + \bar{X}_{(p+3)/2}}{2}, \\ \bar{X}_{(p+4)/2} &= \frac{\bar{X}_{(p+3)/2} + \bar{X}_{(p+5)/2}}{2} \\ &\vdots \end{aligned}$$

Per exemple, si  $p = 4$ ,

$$\bar{X}_3 = \frac{\bar{X}_{(5/2)} + \bar{X}_{(7/2)}}{2} \quad \text{i} \quad \bar{X}_4 = \frac{\bar{X}_{(7/2)} + \bar{X}_{(9/2)}}{2},$$

on  $\bar{X}_{(5/2)} = \frac{X_1 + X_2 + X_3 + X_4}{2}$ ,  $\bar{X}_{(7/2)} = \frac{X_2 + X_3 + X_4 + X_5}{2}$  i  $\bar{X}_{(9/2)} = \frac{X_3 + X_4 + X_5 + X_6}{2}$ .

Amb aquest procés s'obté una nova sèrie de valors molt més estables degut, per una part, a la reducció de la variabilitat ocasionada en fer les mitjanes i, per altra, a que si el període elegit és el correcte, al passar d'una mitjana mòbil a la següent, la nova dada incorporada té el mateix comportament que la dada sortint.

Són freqüents les situacions en que és difícil modelitzar la tendència a través d'un simple model polinòmic de menor o major grau.

Per resoldre aquest problema, hi ha altres tècniques, com l'allisat exponencial [3]. Aquests procediments d'obtenció de la tendència són semblants a les mitjanes mòbils. En els dos casos la tendència és el resultat de fer una mitjana de valors, bé de forma ponderada o bé sense ponderar. En els procediments d'allisat exponencial, s'utilitzen ponderacions, i a més, els valors que s'obtenen són una combinació lineal de totes les observacions de la sèrie, però amb la particularitat que la ponderació decreix quan ens allunyam de l'origen. Aquests procediments estan especialment dissenyats per a la predicció, però no els analitzarem en el present treball.

Es poden distingir diferents tipus d'allisat exponencial.

- L'allisat exponencial simple, s'utilitza per a sèries sense tendència ni estacionalitat.
- L'allisat exponencial lineal de Holt, s'utilitza per a sèries amb tendència lineal i sense estacionalitat.
- L'allisat exponencial estacional de Holt-Winters, s'utilitza per a sèries amb tendència i estacionalitat.

Vegem un exemple de dades reals on posarem en pràctica el càlcul de les mitjanes mòbils.

### Exemple

La sèrie temporal següent recull les temperatures mitjanes mensuals registrades en Palma de Mallorca, Son Sant Joan, en el període de temps des de gener de 2006 a desembre de 2015 [4]. En total es tracta de 120 observacions i estan recopilades a la taula 2.1.

Dels 120 valors disponibles en aquest exemple es fan dos grups, els 108 primers, corresponents a 9 anys, seran utilitzats per modelitzar la sèrie i fer prediccions, i les darreres 12 dades, un any, s'utilitzaran per validar les previsions realitzades.

A la figura 2.4 trobam la representació gràfica de les 108 primeres dades. Observant la gràfica podem detectar que el patró estacional es manté amb amplitud constant, és a dir, que la magnitud de les fluctuacions estacionals no varia amb el temps, per tant es tracta d'un model additiu.

## 2. MODELITZACIÓ PER COMPONENTS

t	$X_t$	t	$X_t$	t	$X_t$	t	$X_t$	t	$X_t$	t	$X_t$	t	$X_t$	t	$X_t$
1	8.8	16	16.1	31	24.9	46	18.3	61	9.2	76	15.1	91	26.3	106	20.6
2	9.4	17	19.3	32	25.1	47	14.7	62	10.4	77	18.9	92	25.8	107	16.3
3	12.7	18	22.4	33	22.1	48	12.0	63	12.0	78	24.0	93	22.8	108	11.0
4	16.1	19	24.6	34	18.8	49	10.1	64	15.9	79	25.2	94	20.9	109	9.3
5	19.0	20	24.5	35	12.9	50	10.8	65	19.3	80	26.9	95	13.5	110	9.2
6	22.8	21	21.7	36	9.9	51	10.9	66	21.6	81	22.6	96	10.5	111	12.1
7	26.4	22	17.7	37	10.0	52	14.1	67	24.7	82	19.7	97	11.7	112	14.8
8	25.1	23	12.6	38	9.9	53	14.1	68	25.7	83	15.2	98	11.4	113	19.2
9	23.3	24	10.0	39	11.2	54	21.2	69	23.5	84	12.0	99	12.0	114	23.4
10	20.9	25	10.9	40	13.6	55	25.5	70	19.1	85	10.7	100	16.2	115	27.9
11	16.6	26	10.7	41	19.3	56	24.8	71	15.8	86	9.8	101	18.0	116	26.2
12	12.1	27	12.5	42	23.0	57	21.9	72	12.1	87	13.7	102	23.1	117	22.1
13	11.2	28	15.1	43	25.7	58	17.6	73	9.3	88	14.8	103	24.8	118	18.6
14	12.4	29	17.8	44	25.6	59	13.5	74	7.0	89	16.9	104	25.6	119	13.9
15	12.5	30	21.6	45	21.8	60	10.0	75	12.2	90	21.8	105	24.2	120	11.5

Taula 2.1: Temperatures mitjanes mensuals

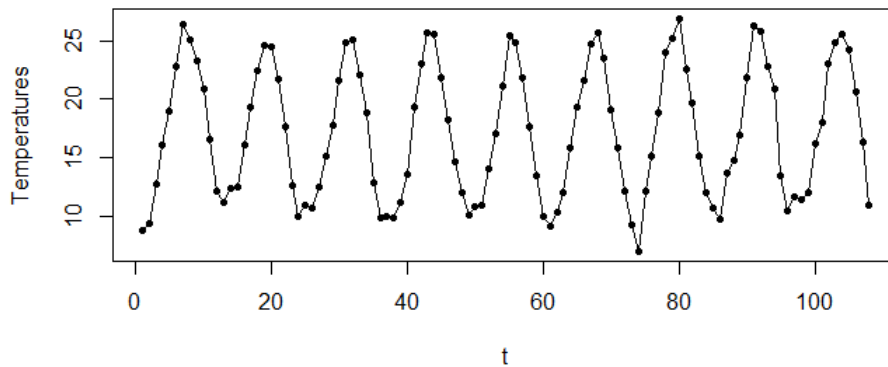


Figura 2.4: Evolució cronològica de les dades dels primers 9 anys

Es calculen les mitjanes mòbils amb  $p = 12$ , més endavant explicarem la justificació d'elegir  $p = 12$ . La seva representació gràfica (figura 2.5) confirma l'estacionalitat per l'estabilització de la sèrie. S'ha provat d'ajustar el model lineal i quadràtic de la tendència sobre les mitjanes mòbils però el coeficient de determinació  $R^2$  no es podia considerar prou bo, ja que era 0.01153 i 0.5655. Per tant, s'ha pres el model cúbic. El resultat és

$$\bar{X}_t = 1.845 \times 10 - 1.151 \times 10^{-1}t + 2.265 \times 10^{-3}t^2 - 1.210 \times 10^{-5}t^3,$$

amb un coeficient de determinació de  $R^2 = 0.7547$ , que és bastant significatiu.



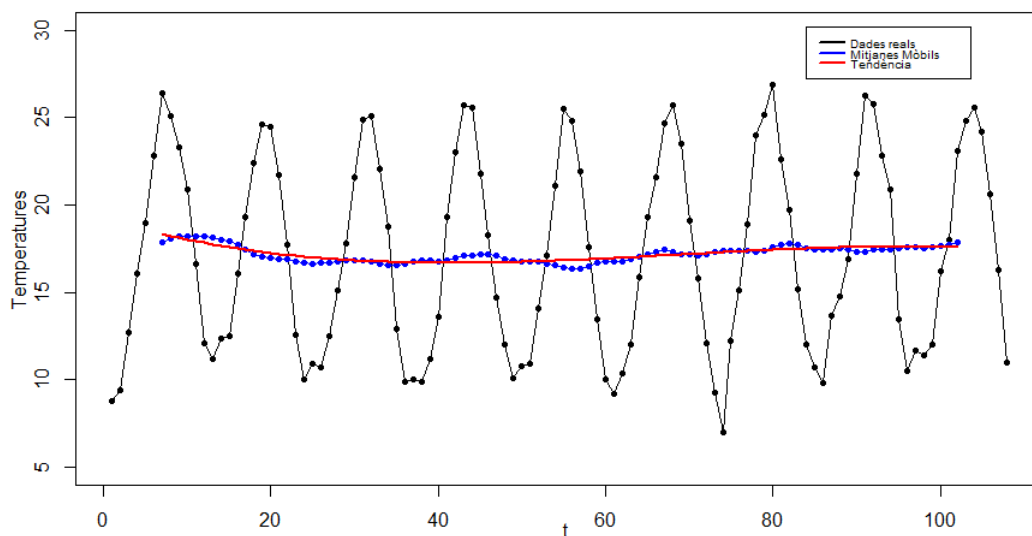


Figura 2.5: Dades reals, mitjanes mòbils i la línia de la tendència de la sèrie

## 2.2 Estacionalitat

La component estacional provoca una oscil·lació sistemàtica de període curt, generalment no superior a l'any i, per tant, pot emmascarar l'evolució a llarg plaç, la tendència, si no l'aïllem convenientment.

L'anàlisi de l'estacionalitat queda lligada al mètode que es decideixi emprar per a modelitzar la tendència. Com que la tendència l'hem calculat a partir de les mitjanes mòbils també calcularem l'estacionalitat a partir d'elles.

Per calcular els valors dels índexs estacionals hem de seguir el següent procés:

Una vegada calculades les mitjanes mòbils,  $\bar{X}_t$ , de la sèrie original  $X_t$  prenent el període d'agrupació  $p$  que es consideri oportú, s'ha de separar la part de la tendència explicada a l'apartat anterior. Segons el model s'ha de procedir d'una forma o d'una altra,

- Model additiu:  $W_t = X_t - \bar{X}_t$ .
- Model multiplicatiu:  $W_t = X_t / \bar{X}_t$ .

Cal destacar però, que en  $W_t$  estan incloses les components associades a l'estacionalitat, els cicles i residus.

Si consideram que els residus són variables aleatòries de mitjana nul·la i que la component cíclica, en cas d'existir, és de període suficientment llarg com per no ser recollida per les dades, podrem avaluar l'estacionalitat associada a cada component del període. Per fer-ho s'han de calcular les mitjanes dels  $W_t$  de la mateixa estació,

$$E_s^* = \frac{\sum_{t=s+\hat{p}} W_t}{n_s} \quad s = 1, \dots, p$$

on  $s$  representa l'índex estacional i  $n_s$  el nombre de valors associats a aquest índex dels que es fa la mitjana. Com que els índexs estacionals mesuren les discrepàncies respecte de la mitjana, necessitam aquest valor com a referència,

$$\bar{E} = \frac{\sum_{s=1}^p E_s^*}{p}$$

Ara ja podem calcular els índexs estacionals; distingirem dos casos segons el tipus de model:

- Model additiu: Els índexs estacionals són les diferències entre les mitjanes de les  $W_t$  de cada estació i la mitjana general que acabam de definir, és a dir,

$$E_s = E_s^* - \bar{E}$$

Cal destacar que la suma d'aquests índexs és zero,  $\sum_{s=1}^p E_s = 0$ , ja que

$$\sum_{s=1}^p E_s = \sum_{s=1}^p (E_s^* - \bar{E}) = \sum_{s=1}^p E_s^* - \sum_{s=1}^p \bar{E} = \sum_{s=1}^p E_s^* - p\bar{E} = \sum_{s=1}^p E_s^* - \sum_{s=1}^p E_s^* = 0.$$

- Model multiplicatiu: En aquest cas els índexs estacionals són el quocient entre les mitjanes de les  $W_t$  de cada estació i la mitjana general, és a dir,

$$E_s = \frac{E_s^*}{\bar{E}}$$

Ara la suma d'aquests índexs és igual al període,  $\sum_{s=1}^p E_s = p$ , ja que

$$\sum_{s=1}^p E_s = \sum_{s=1}^p \frac{E_s^*}{\bar{E}} = \frac{1}{\bar{E}} \sum_{s=1}^p E_s^* = \frac{1}{\bar{E}} (p\bar{E}) = p$$

Finalment, els residus es calcularan com  $R = X - T - E$  (la sèrie lliure de la tendència i l'estacionalitat). Els residus ens poden confirmar la bona modelització de la sèrie.

Continuarem amb l'exemple anterior, el de les Temperatures de Palma de Mallorca, per estudiar l'estacionalitat.

### Exemple

Anem a estudiar l'estacionalitat de l'exemple de les Temperatures de Palma de Mallorca. Per avaluar l'estacionalitat és necessari calcular els índexs estacionals.

En la taula 2.2 es detallen els càlculs fets fins obtenir els índexs estacionals.

## 2.2. Estacionalitat

t	$X_t$	$\bar{X}_t$	$W_t$	Estació:s
1	8.8	...	...	1
2	9.4	...	...	2
3	12.7	...	...	3
4	16.1	...	...	4
5	19.0	...	...	5
6	22.8	...	...	6
7	26.4	17.86667	8.53333	7
8	25.1	18.09167	6.89167	8
9	23.3	18.20833	5.09167	9
10	20.9	18.20000	2.70000	10
11	16.6	18.21250	-1.61250	11
12	12.1	18.20833	-6.10833	12
13	11.2	18.11667	-6.91667	1
14	12.4	18.01667	-5.61667	2
15	12.5	17.92500	-5.42500	3
16	16.1	17.72500	-1.62500	4
17	19.3	17.42500	1.87500	5
18	22.4	17.17083	5.22917	6
19	24.6	17.07083	7.52917	7
20	24.5	16.98750	7.51250	8
21	21.7	16.91667	4.78333	9
22	17.7	16.87500	0.82500	10
23	12.6	16.77083	-4.17083	11
24	10.0	16.67500	-6.67500	12
25	10.9	16.65417	-5.75417	1
26	10.7	16.69167	-5.99167	2
27	12.5	16.73333	-4.23330	3
28	15.1	16.79583	-1.69583	4
29	17.8	16.85417	0.94583	5
30	21.6	16.86250	4.73750	6
31	24.9	16.82083	8.07917	7
32	25.1	16.75000	8.35000	8
33	22.1	16.66250	5.43750	9
34	18.8	16.54583	2.25417	10
35	12.9	16.54583	-3.64583	11
36	9.9	16.66667	-6.76670	12
37	10.0	16.75833	-6.75833	1
38	9.9	16.81250	-6.91250	2
39	11.2	16.82083	-5.62083	3
40	13.6	16.78750	-3.18750	4
41	19.3	16.84167	2.45833	5
42	23.0	17.00417	5.99583	6
43	25.7	17.09583	8.60417	7
44	25.6	17.13750	8.46250	8
45	21.8	17.16250	4.63750	9
46	18.3	17.17083	1.12917	10
47	14.7	17.10000	-2.40000	11
48	12.0	16.92917	-4.92917	12
49	10.1	16.84167	-6.74167	1
50	10.8	16.80000	-6.00000	2
51	10.9	16.77083	-5.87083	3
52	14.1	16.74583	-2.64583	4
53	17.1	16.66667	0.43333	5
54	21.1	16.53333	4.56667	6
55	25.5	16.41250	9.08750	7
56	24.8	16.35833	8.44167	8
57	21.9	16.38750	5.51250	9
58	17.6	16.50833	1.09167	10
59	13.5	16.67500	-3.175	11
60	10.0	16.78750	-6.78750	12
61	9.2	16.77500	-7.57500	1
62	10.4	16.77917	-6.37917	2
63	12.0	16.88333	-4.88333	3
64	15.9	17.01250	-1.11250	4
65	19.3	17.17083	2.12917	5
66	21.6	17.35417	4.25483	6
67	24.7	17.44583	7.25417	7
68	25.7	17.30833	8.39167	8
69	23.5	17.17500	6.32500	9
70	19.1	17.15000	1.95000	10
71	15.8	17.10000	-1.30000	11
72	12.1	17.18333	-5.08333	12
73	9.3	17.30417	-8.00417	1
74	7.0	17.37500	-10.37500	2
75	12.2	17.38750	-5.18750	3
76	15.1	17.37500	-2.27500	4
77	18.9	17.37500	1.52500	5
78	24.0	17.34583	6.65417	6
79	25.2	17.40000	7.80000	7
80	26.9	17.57500	9.32500	8
81	22.6	17.75417	4.84583	9
82	19.7	17.80417	1.89583	10
83	15.2	17.70833	-2.50833	11
84	12.0	17.53333	-5.53333	12
85	10.7	17.48750	-6.78750	1
86	9.8	17.48750	-7.68750	2
87	13.7	17.45000	-3.75000	3
88	14.8	17.50833	-2.70833	4
89	16.9	17.48750	-0.58750	5
90	21.8	17.35417	4.44583	6
91	26.3	17.33333	8.96667	7
92	25.8	17.44167	8.35833	8
93	22.8	17.43750	5.36250	9
94	20.9	17.42500	3.47500	10
95	13.5	17.52917	-4.02917	11
96	10.5	17.62917	-7.12917	12
97	11.7	17.62083	-5.92083	1
98	11.4	17.55000	-6.15000	2
99	12.0	17.60000	-5.60000	3
100	16.2	17.64583	-1.44583	4
101	18.0	17.75000	0.25000	5
102	23.1	17.88750	5.21250	6
103	24.8	...	...	7
104	25.6	...	...	8
105	24.2	...	...	9
106	20.6	...	...	10
107	16.3	...	...	11
108	11.0	...	...	12

Taula 2.2: Evolució de l'estacionalitat

## 2. MODELITZACIÓ PER COMPONENTS

Per exemple per al setèim semestre ( $s = 7$ ), la mitjana de les  $W_t$ , amb valors de temps corresponents al setèim semestre, per ser múltiples de 12 més 7 ( $t=7, 19, 31, 43, 55, 67, 79, 91$ ), seria:

$$E_7^* = \frac{8.53333 + 7.52917 + 8.07917 + 8.60417 + 9.08750 + 7.25417 + 7.80000 + 8.96667}{8}$$

Anàlogament, per a cada trimestre, s'obté:

$$\begin{array}{llll} E_1^* = -6.807293 & E_2^* = -6.889064 & E_3^* = -5.071349 & E_4^* = -2.086978 \\ E_5^* = 1.128645 & E_6^* = 5.137063 & E_7^* = 8.231773 & E_8^* = 8.216668 \\ E_9^* = 5.249479 & E_{10}^* = 1.915105 & E_{11}^* = -2.855208 & E_{12}^* = -6.126566 \end{array}$$

La mitjana general és:  $\bar{E} = 0.003523$ .

Per tant, els índex estacionals són:

Mes	s	$E_s$	Mes	s	$E_s$
Gener	1	-6.811936	Juliol	7	8.227127
Febrer	2	-6.893707	Agost	8	8.226606
Març	3	-5.075998	Setembre	9	5.244835
Abril	4	-2.091623	Octubre	10	1.910460
Maig	5	1.124002	Novembre	11	-2.859852
Juny	6	5.131293	Desembre	12	-6.131207

Taula 2.3: Índexs estacionals

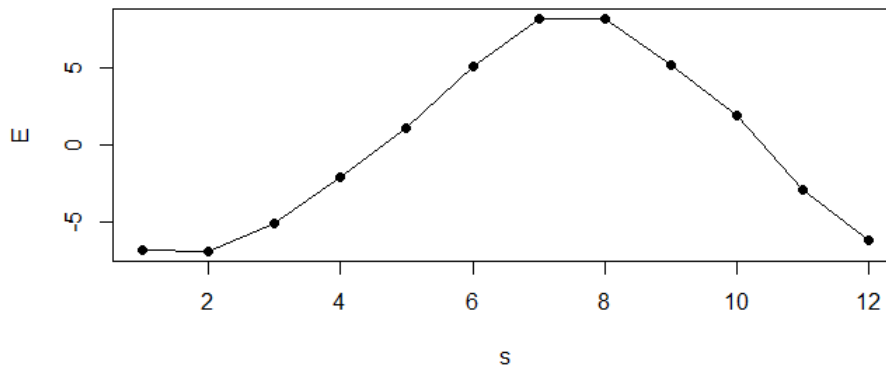


Figura 2.6: Component estacional: índexs

La interpretació dels índexs és molt senzilla, de maig fins a octubre la temperatura mitjana mensual està per sobre de la mitjana anual mentre que de novembre a abril

està per sota, essent juliol i agost els mesos on es registren les temperatures més altes, i gener i febrer les més baixes.

D'aquesta manera, el model proposat per a aquesta sèrie temporal és

$$\tilde{X}_t = 1.845 \times 10 - 1.151 \times 10^{-1}t + 2.265 \times 10^{-3}t^2 - 1.210 \times 10^{-5}t^3 + E_s,$$

amb  $t = 12 + s$ ,  $s = 1, \dots, 12$ .

L'evolució del model, juntament amb les dades reals, es presenten a la figura 2.7. Cal destacar la bona concordança que hi ha entre la sèrie i el model, tot i que en alguns punts presenten discrepàncies.

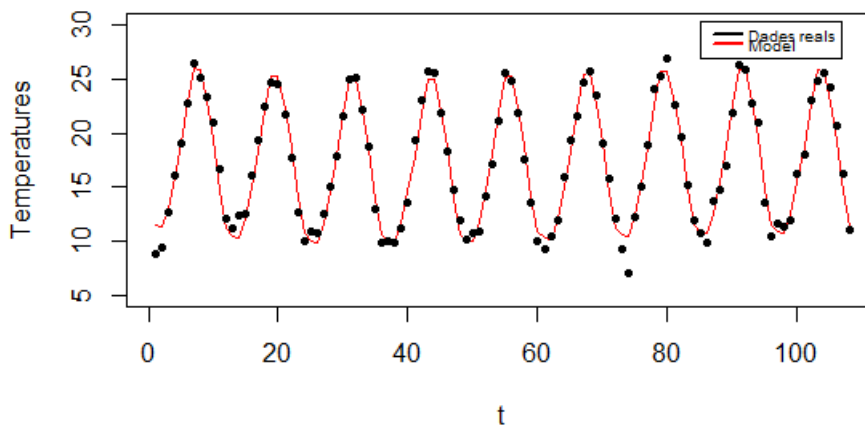


Figura 2.7: Dades reals i model

Això s'aprecia millor a la figura 2.8, on es representen els residus, és a dir, la diferència entre el valor real i el modelitzat. Podem observar un bon ajustament a tots els punts en general.

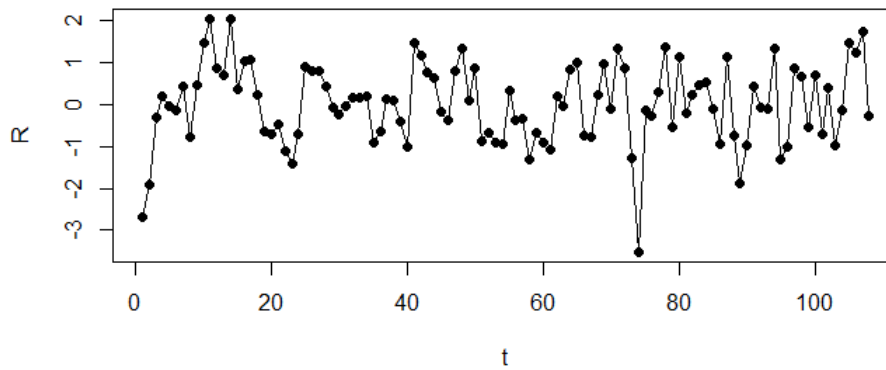


Figura 2.8: Residus

Amb el model disponible es poden fer previsions, per exemple per als pròxims dotze mesos. En la taula 2.4 es presenten els valors prevists junt als que realment es varen obtenir en aquests mesos i que han estat separats i guardats a l'hora de fer la modelització anterior. Cal destacar que les petites discrepàncies entre les previsions i el valor real, en part, han de atribuir-se al component aleatori, que és present en cada moment. A la figura 2.9 es mostren les dades reals i el model, a més de les previsions fetes. Cal comentar la gran concordança entre les previsions i la realitat, llevat d'alguns valors extrems.

t	X	Previsió	Diferència
109	9.3	10.56160	-1.26160
110	9.2	10.43994	-1.23994
111	12.1	12.21481	-0.11481
112	14.8	15.15331	-0.35331
113	19.2	18.31997	0.88003
114	23.4	22.27514	1.12486
115	27.9	25.31563	2.58437
116	26.2	25.25648	0.94352
117	22.1	22.21273	-0.11273
118	18.6	18.81296	-0.21296
119	13.9	13.97377	-0.07377
120	11.5	10.62999	0.87001

Taula 2.4: Previsions i valors reals obtinguts al mateix període

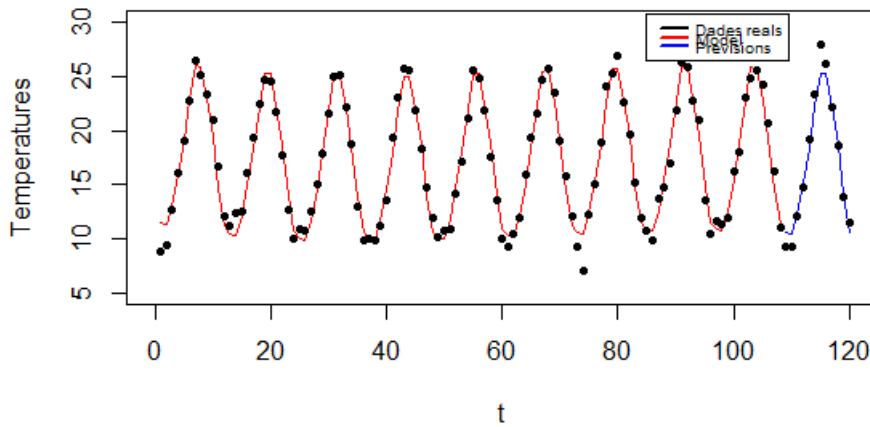


Figura 2.9: Dades de 2006 a 2014 i previsions del 2015

Com hem comentat al principi, hi ha altres mètodes més específics per a realitzar prediccions. Així, podem parlar dels allisats exponencials; l'allisat exponencial simple, l'allisat exponencial lineal de Holt i l'allisat exponencial de Holt-Winters.

## 2.3 Autocorrelació

Com s'ha explicat a la secció anterior, per poder aplicar els mètodes de descomposició el primer que s'ha de determinar és la periodicitat de la sèrie. L'autocorrelació i la seva representació gràfica, el correlograma, ens ajudaran a fer-ho.

La funció d'autocorrelació té una doble utilitat, per una banda serveix per a confirmar la presència d'estacionalitat i determinar el seu període; per l'altra, indica quantes previsions són admissibles a partir del darrer temps que s'ha recollit informació.

**Definició 2.3.1 (Coeficient d'autocorrelació)** Donada una sèrie temporal  $\{X_i\}_{i=1}^N$  i  $k$  un valor enter menor que  $N$ , consideram les sèries  $\{X_i\}$  i  $\{X_{i+k}\}$  amb  $1 \leq i \leq N-k$ . Definim el coeficient d'autocorrelació d'ordre  $k$ ,  $\rho_k$ , com el coeficient de correlació d'ambdues sèries.

Un valor no nul de  $\rho_k$  indica que existeix correlació entre informacions separades  $k$  unitats de temps, és a dir, que la informació es transmet  $k$  unitats de temps més enllà. D'aquesta forma, si el darrer valor de temps del qual es disposa de dades és  $T$ , serà admissible fer previsions per a un temps igual a  $T+k$ . Evidentment, si  $\rho_k$  fos nul no seria admissible fer una predicció per a  $T+k$ , ja que les dades disponibles no transmeten cap informació rellevant a una distància com la considerada.

La sèrie  $\{X_{i+k}\}$  és la sèrie desplaçada de  $\{X_i\}$  en  $k$  unitats de temps.

De l'estructura del plantejament es dedueix que  $\rho_k = \rho_{-k}$ .

### 2.3.1 Correlograma

**Definició 2.3.2** Direm correlograma a la representació gràfica de la funció d'autocorrelació, és a dir, de  $\rho_k$  en funció del retard  $k$ .

Si disposam d'una sèrie temporal de dades  $x_1, x_2, \dots, x_N$ , per elaborar el correlograma s'han d'estimar les característiques següents:

- Mitjana:  $\bar{x} = \frac{\sum_{i=1}^N x_i}{N}$

- Autocovariància:  $\gamma_k = \frac{\sum_{i=1}^{N-k} (x_i - \bar{x})(x_{i+k} - \bar{x})}{N} \quad k = 0, 1, \dots, N-1$

- Autocorrelació:  $\tilde{\rho}_k = r_k = \frac{\gamma_k}{\gamma_0}$

Per poder estimar l'autocovariància,  $\gamma_k$ , el nombre de components de la sèrie ha de ser tal que  $N > k + 1$ , i habitualment es considera recomanable que  $N \geq 50$  i  $k \leq N/4$  [5].

Per a identificar els coeficients d'autocorrelació que siguin significativament diferents de zero és necessari estudiar el comportament estadístic dels estimadors. Bartlett [6] ha estudiat l'estadístic  $r_k$  i, sota la suposició que  $\rho_k = 0$  per a tot  $k \geq K$ , obté que

$$Var(r_k) \cong \frac{1}{N} \sum_{i=-(K-1)}^{K-1} \rho_i^2 \quad \forall k \geq K$$

En substituir  $\rho_i$  pel seu estimador,  $r_i$ , i, ja que  $\rho_0 = 1$  i  $\rho_i = \rho_{-i}$ , resulta que l'estimació de la variància de  $r_k$  és igual a

$$\tilde{Var}(r_k) \cong \frac{1}{N} \sum_{i=-(K-1)}^{K-1} r_i^2 \Rightarrow \begin{cases} Var(r_k) \cong \frac{1}{N} & k \geq K \quad K = 1, \\ Var(r_k) \cong \frac{1}{N} \left( 1 + 2 \sum_{i=1}^{K-1} r_i^2 \right) & k \geq K \quad K > 1. \end{cases}$$

Anderson [7] indica que per a valors de  $k$  tals que  $\rho_k = 0$  i  $N$  suficientment gran,  $r_k$  es distribueix aproximadament com una  $\mathcal{N}(0, Var(r_k))$ . Així, amb un nivell de significació de  $\alpha = 0.05$ , si  $\rho_k = 0$ , el seu estimador  $r_k$  es trobarà a l'interval  $\pm 2S(r_k)$ , on  $S(r_k)$  representa la desviació típica estimada de  $r_k$ , és a dir,  $S(r_k) = \sqrt{\tilde{Var}(r_k)}$ .

L'interval  $\pm 2S(r_k)$  es denomina l'interval de no significació de  $\rho_k$ , i és el conjunt de valors que pot prendre  $r_k$  perquè, amb un risc del 5% es pugui admetre l'absència de correlació entre valors de la sèrie desplaçats  $k$  unitats de temps. D'aquesta manera, al calcular la funció d'autocorrelació d'una sèrie, és bo representar-la gràficament juntament amb l'interval  $\pm 2S(r_k)$ , amb l'objectiu de considerar únicament com a coeficients d'autocorrelació no nuls aquells que tinguin l'estimació fora d'aquest interval.

Posem en pràctica el vist en aquest apartat amb l'exemple de les temperatures.



**Exemple**

En l'exemple de les Temperatures de Palma de Mallorca, tenim que el correlograma de la sèrie ve donat a la figura 2.10. Podem observar l'existència d'una estacionalitat de període 12 (comprovant el supòsit fet en el càlcul de les mitjanes mòbils), naturalment lligada a l'evolució mensual de les dades, ja que cada 12 barres d'autocorrelació es repeteix la mateixa estructura de comportament.

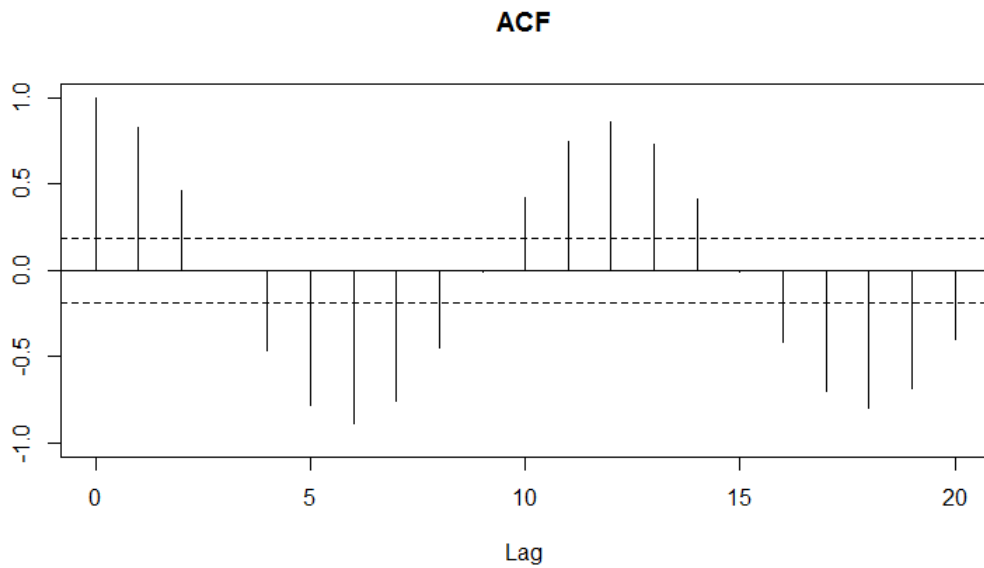


Figura 2.10: Correlograma de la sèrie



## MODELS PROBABILÍSTICS

La regressió clàssica és sovint insuficient per explicar totes les dinàmiques d'una sèrie temporal. En lloc d'això, es proposen els models autoregressius (AR) i autoregressius de mitjana mòbil (ARMA) amb la introducció de la correlació, generada a través de les relacions lineals entre els retards. L'addició dels models no estacionaris condueix al model autoregressiu integrat de mitjanes mòbils (ARIMA) [8].

A principis de la dècada dels anys 1970, G.E.P. Box, professor d'Estadística de la Universitat de Wisconsin, i G.M. Jenkins, professor d'Enginyeria de Sistemes de la Universitat de Lancaster, realitzen un estudi sobre la contaminació en la badia de San Francisco, Califòrnia. Les dades disponibles eren sèries temporals conformades per l'observació dels nivells de contaminació. Per establir un mecanisme de pronòstic i control, desenvolupen un nou enfocament per al tractament de sèries temporals discretes, la modelització de sèries amb el mètode d'autoregressió i mitjanes mòbils, anomenat "ARIMA", per les seves sigles angleses, conegut actualment com el "Mètode de Box-Jenkins".

A Espanya es realitzaren algunes aplicacions d'aquest tipus, com són els treballs de D. Peña sobre projeccions de consum de gasolina, i l'estudi d'A.B. Treadway sobre els efectes en l'economia espanyola de les devaluacions de la pesseta.

A partir de mitjans de la dècada de 1980 es comencen a realitzar múltiples aplicacions del mètode ARIMA a la modelització de sèries temporals discretes sobre diversos aspectes relacionats amb el camp de la salut [9].

L'anàlisi de les sèries temporals a través de la metodologia de Box-Jenkins consisteix en determinar quin és el model probabilístic que regeix el comportament del fenomen al llarg del temps. És a dir, partint de la premissa que no sempre és possible identificar les components de la sèrie, es tracta d'estudiar el component aleatori pur, reflectit en els residus [5].

La metodologia utilitzada en l'estudi d'una sèrie temporal per aquest sistema segueix els passos següents:

1. Identificació del model ARIMA.

Representar gràficament la sèrie, a més de la seva funció d'autocorrelació (ACF) i funció d'autocorrelació parcial (PACF). La gràfica de la sèrie ens indica si és estacionària o no. Segons els motius pels quals la sèrie no és estacionària, haurem d'aplicar els procediments següents fins fer-la estacionària.

- Si té tendència: prendrem diferències regulars fins que desapareixi. Normalment l'ordre de la diferència és 1, i rarament major a 3.
- Si la sèrie té estacionalitat: prendrem diferències estacionals fins que desapareixi el patró estacional. En la pràctica és molt estrany haver d'aplicar més d'una diferència estacional.
- Si és heterocedàstica (no té variància constant): haurem de transformar la sèrie. En molts de casos és suficient prendre logaritmes, encara que hi ha transformacions més sofisticades, com la de Box-Cox.

Una vegada que el gràfic de la nova sèrie (transformació de l'original) indica que és estacionària, podem intentar deduir l'estructura de la sèrie observant les seves ACF i PACF.

#### 2. Estimació i validació del model.

Observant els dos gràfics de l'ACF i PACF de la sèrie transformada podem fer-nos una idea del model que descriu la nostra sèrie, o almenys de quins són els primers candidats que hem de provar. Per comprovar analíticament un model freqüentment s'ajusten diferents models candidats ARIMA(p,d,q) i a partir d'una sèrie de criteris, com l'AIC (Criteri d'Informació d'Akaike) i el BIC (Criteri d'Informació Bayesiana), seleccionam un dels models. Posteriorment, el validam amb una sèrie de proves per veure si té els residus semblants als d'un soroll blanc.

#### 3. Predicció.

Una de les raons de la popularitat del procés de construcció dels models ARIMA és el seu èxit en la predicció. Els models ARIMA són bons per realitzar prediccions a curt termini [1].

## 3.1 Procés estocàstic

Començam donant les definicions bàsiques.

**Definició 3.1.1** *Un procés estocàstic és una col·lecció de variables aleatòries  $\{X_t : t \in T\}$  sobre un espai de probabilitat  $(\Omega, \mathfrak{F}, p)$ , on el conjunt de paràmetres  $T$  indica el temps. Si  $T$  és un conjunt discret, el procés estocàstic s'anomena a temps discret, i direm que és un procés estocàstic a temps continu en cas contrari.*

**Definició 3.1.2** *S'anomena realització d'un procés estocàstic a la col·lecció  $\{X_t(\omega) : t \in T\}$  per a cada  $\omega \in \Omega$  fix.*

**Definició 3.1.3** *Una sèrie temporal és una realització d'un procés estocàstic a temps discret, on els elements de  $T$  estan ordenats i corresponen a instants equidistants del temps.*

Una manera de descriure un procés estocàstic és especificar la distribució de probabilitat conjunta de  $X_{t_1}, \dots, X_{t_n}$  per a qualsevol conjunt de temps  $t_1 < \dots < t_n$  i qualsevol valor de  $n \geq 1$ . Són les anomenades distribucions en dimensió finita del procés.

Una manera d'especificar parcialment un procés estocàstic és mitjançant alguns dels seus moments. Així, tenim:

**Definició 3.1.4** *La funció mitjana  $\mu(t)$ :*

$$\mu(t) = E[X_t]$$

**Definició 3.1.5** *La funció variància  $\sigma^2(t)$ :*

$$\sigma^2(t) = Var[X_t]$$

**Definició 3.1.6** *La funció autocovariància  $\gamma(t_1, t_2)$ :*

$$\gamma(t_1, t_2) = Cov(X_{t_1}, X_{t_2}) = E\{[X_{t_1} - \mu(t_1)][X_{t_2} - \mu(t_2)]\}$$

Cal notar que la funció variància és un cas especial de la funció autocovariància quant  $t_1 = t_2 = t$ .

**Definició 3.1.7** *La funció d'autocorrelació (ACF) es defineix com*

$$\rho(s, t) = \frac{\gamma(s, t)}{\sqrt{\gamma(s, s)\gamma(t, t)}}.$$

L'ACF mesura la previsibilitat lineal de la sèrie en el temps  $t$ , diguem  $X_t$ , utilitzant només el valor  $X_s$ . Es pot demostrar fàcilment que  $-1 \leq \rho(s, t) \leq 1$  utilitzant la desigualtat de Cauchy-Schwarz. Si podem predir  $X_t$  perfectament des de la  $X_s$  a través d'una relació lineal,  $X_t = \beta_0 + \beta_1 X_s$ , llavors la correlació serà 1 quan  $\beta_1 > 0$ , i -1 quan  $\beta_1 < 0$ . D'altra banda, quan  $\rho(s, t) = 0$ , no hi ha cap relació lineal entre  $X_t$  i  $X_s$ . Per tant, tenim una mesura aproximada de la capacitat de preveure la sèrie en el temps  $t$  des del valor en el moment  $s$ .

### 3.1.1 Estimació de la correlació

Per determinar un model adequat per a un conjunt de dades de sèries temporals donat, cal dur a terme l'anàlisi de l'ACF. Aquesta funció mostra com estan relacionades entre si les observacions en la sèrie temporal.

Tot i que la funció d'autocorrelació teòrica és útil per descriure les propietats de certs models hipotètics, la majoria de les anàlisis s'han de fer mitjançant dades de la mostra. Aquesta limitació significa que només estan disponibles els punts mostrejats  $x_1, x_2, \dots, x_N$  per a l'estimació de les funcions de mitjana, autocovariància i autocorrelació. Des del punt de vista de l'estadística clàssica, això planteja un problema, ja que normalment no tindrem còpies iid<sup>1</sup> de  $x_n$  disponibles per a l'estimació de les funcions de covariància i correlació. En la situació habitual d'una sola realització, la suposició d'estacionarietat es torna crítica. D'alguna manera, hem de fer servir mitjanes

<sup>1</sup>Independents i idènticament distribuïdes.

sobre aquesta única realització per estimar les mitjanes de la població i les funcions de covariància.

En conseqüència, si una sèrie temporal és estacionària, la funció mitjana,  $\mu_t = \mu$  és constant, de manera que podem estimar la mitjana de la mostra mitjançant

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i.$$

La funció d'autocovariància teòrica s'estima mitjançant la funció d'autocovariància de la mostra, que es defineix així:

**Definició 3.1.8** *La funció d'autocovariància mostral és*

$$\gamma_k = \frac{1}{N} \sum_{i=k+1}^N (x_i - \bar{x})(x_{i-k} - \bar{x})$$

per  $k = 0, 1, \dots, N-1$ .

**Definició 3.1.9** *La funció d'autocorrelació mostral es defineix com*

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\sum_{i=k+1}^N (x_i - \bar{x})(x_{i-k} - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2}.$$

La funció d'autocorrelació de la mostra té una distribució mostral que ens permet avaluar si les dades provenen d'una sèrie completament a l'atzar o si les correlacions són estadísticament significatives en alguns retards [8].

Com van explicar Box i Jenkins [10], la gràfica de l'ACF i la PACF mostrals (que definirem posteriorment) és útil per a determinar el tipus de model que s'adapta a una sèrie de temps de longitud  $N$ . Com que l'ACF mostral és simètrica al voltant de zero, és suficient representar l'ACF mostral per a retards positius, i amb un màxim de  $N/4$  retards de desfasament [11].

## 3.2 Estacionarietat

Molts de processos tenen la propietat que la seva naturalesa aleatòria no varia amb el temps. Així, una observació del procés en un interval  $(t_1, t_2)$  mostra el mateix tipus de comportament aleatori que en un altre interval  $(t_1 + h, t_2 + h)$ . Això condueix al concepte de procés estacionari. Distingirem dos tipus d'estacionarietat, donats a les dues definicions següents:

**Definició 3.2.1** *Un procés estocàstic és estrictament estacionari si la distribució de qual-sevol col·lecció de variables aleatòries  $\{X_{t_1}, \dots, X_{t_n}\}$  no depèn dels instants  $t_1, \dots, t_n$  sinó només dels intervals entre ells. És a dir la distribució conjunta de  $X_{t_1}, \dots, X_{t_n}$  és la mateixa que la de  $X_{t_1+h}, \dots, X_{t_n+h}$  per a tots  $t_1 < \dots < t_n$  i  $h$ .*

En altres paraules, desplaçant l'origen de temps una quantitat  $h$  no té cap efecte sobre la distribució conjunta, per tant, ha de dependre només dels intervals entre  $t_1, t_2, \dots, t_n$ , per qualsevol valor de  $n$  [12].

**Definició 3.2.2** *Un procés estocàstic és estacionari de segon ordre (o dèbilment estacionari) si*

- (i) *la funció mitjana,  $\mu(t)$ , és constant (no depèn del temps  $t$ ), i*
- (ii) *la funció autocovariància,  $\gamma(s, t)$  depèn de  $s$  i  $t$  només a través de la seva diferència  $|s - t|$  [8].*

En particular, totes les variables aleatòries del procés tenen la mateixa mitjana i la mateixa variància. També, totes les variables aleatòries separades  $k$  períodes,  $(X_t, X_{t-k})$  tendran la mateixa covariància  $\gamma_k$ .

Observem que tot procés estocàstic estacionari és estacionari de segon ordre. El recíproc no és cert en general, però sí per a una família de processos estocàstics molt importants, els gaussians, que són aquells processos tals que totes les seves distribucions en dimensió finita són lleis normals multidimensionals.

L'estacionarietat estricta és una condició molt forta i difícil de comprovar a la pràctica. Per això, s'acostuma a considerar només l'estacionarietat de segon ordre, i, així, direm procés estacionari al que ho sigui de segon ordre.

En general, cal que les dades de les sèries temporals siguin estacionàries, una forma apropiada de fer-ho és amb una mitjana de productes endarrerits amb el temps. Amb dades de sèries temporals, és important mesurar la dependència entre els valors de la sèrie; hem de ser capaços almenys d'estimar amb precisió les autocorrelacions. Seria difícil de mesurar aquesta dependència si l'estructura de dependència no és regular o està canviant en cada punt de temps. Per tant, per aconseguir una anàlisi estadística significativa de les dades de sèries temporals, serà crucial que les funcions de mitjana i autocovariància satisfacin les condicions d'estacionarietat (almenys durant algun període de temps raonable) [8].

Una altra característica de les sèries és la seva variabilitat.

**Definició 3.2.3** *Deim que una sèrie és homocedàstica si la variabilitat es manté constant al llarg del temps. En cas contrari (la variabilitat augmenta o disminueix) deim que és heterocedàstica.*

Si la sèrie és heterocedàstica haurem de transformar-la en homocedàstica. En molts de casos és suficient prendre logaritmes, encara que hi ha transformacions més sofisticades, com la de Box-Cox [8].

Es defineix una sèrie com estacionària si compleix les característiques següents:

- No té tendència.
- És homocedàstica.
- No té cicles estacionals.

### 3. MODELS PROBABILÍSTICS

---

- L'estructura de dependència es manté constant, és a dir, una observació influeix sobre la següent. Aquesta condició és important per modelitzar la sèrie, ja que si la dependència de la sèrie canvia, és impossible que puguem preveure l'evolució de la sèrie.
- L'influència de les observacions sobre les posteriors decreix amb el temps [13].

#### Exemples

En la figura 3.1 podem observar una sèrie estacionària i en la figura 3.2 una no estacionària.

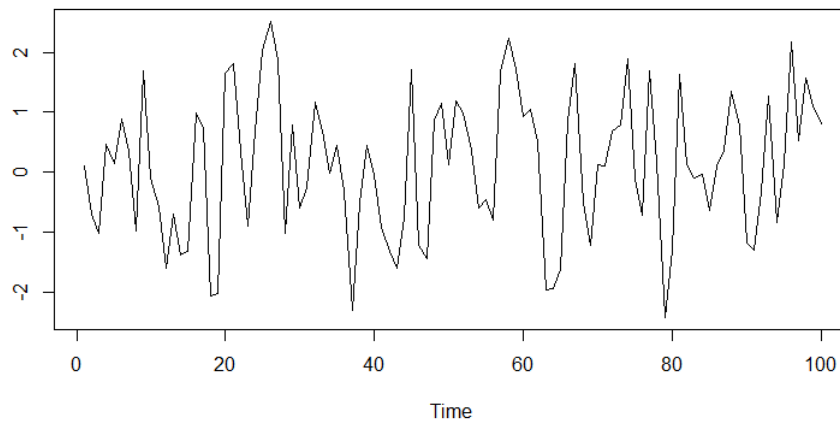


Figura 3.1: Sèrie estacionària

La sèrie 3.1 [1] no presenta tendència, té variància constant, i a més no presenta variacions estacionals.



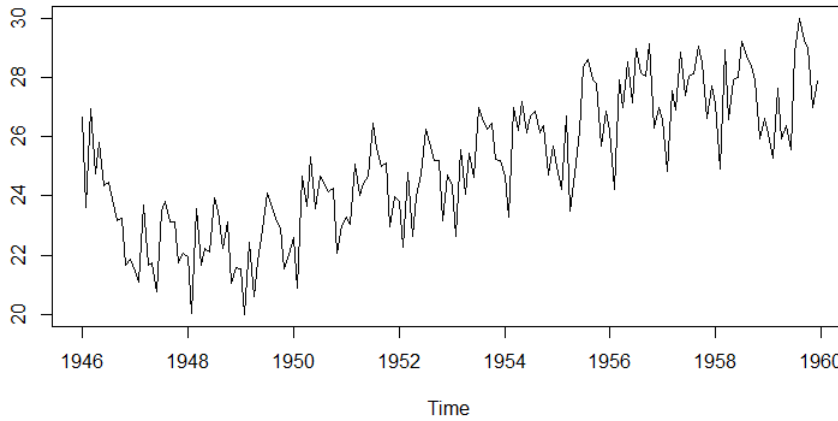


Figura 3.2: Sèrie no estacionària.

En canvi, en la sèrie 3.2 [14] podem veure que hi ha variacions estacionals. A més les fluctuacions estacionals són més o menys constants en grandària amb el temps, i les fluctuacions aleatòries també semblen ser més o menys constants en grandària amb el temps. I la sèrie clarament presenta tendència.

Si  $\{X_t : t \in T\}$  és una sèrie temporal estacionària, aleshores la funció mitjana i la funció variància són constants,  $\mu(t) = \mu, \sigma^2(t) = \sigma^2 \forall t \in T$ , i la funció d'autocovariància només depèn de la diferència entre els instants de temps. Així, podem escriure

$$\gamma(h) = \text{Cov}(X_{t+h}, X_t) = \text{Cov}(X_t, X_{t-h}) = \gamma(-h)$$

i la funció d'autocorrelació serà

$$\rho(h) = \frac{\gamma(h)}{\sqrt{\sigma^2 \cdot \sigma^2}} = \frac{\gamma(h)}{\sigma^2} = \frac{\gamma(h)}{\gamma(0)}$$

i també  $\rho(h) = \rho(-h)$ .

Com hem comentat anteriorment, el procés estocàstic que regula una sèrie temporal és desconegut i per tant, no és possible determinar els valors reals o teòrics de l'ACF. Per això, s'estima a partir dels valors mostrals de la sèrie:  $x_1, \dots, x_N$ .

En el correlograma d'una sèrie estacionària, la dependència temporal s'ha d'acostar ràpidament a zero. En canvi, un decreixement lent de les autocorrelacions indica que la sèrie no és estacionària. I un decreixement lent de la correlació dels retards estacionals indica que la sèrie té un component estacional. Moltes vegades la no estacionarietat s'observa fàcilment en el gràfic de la sèrie, altres vegades haurem de recórrer al correlograma per ajudar-nos a decidir si la sèrie és o no estacionària.

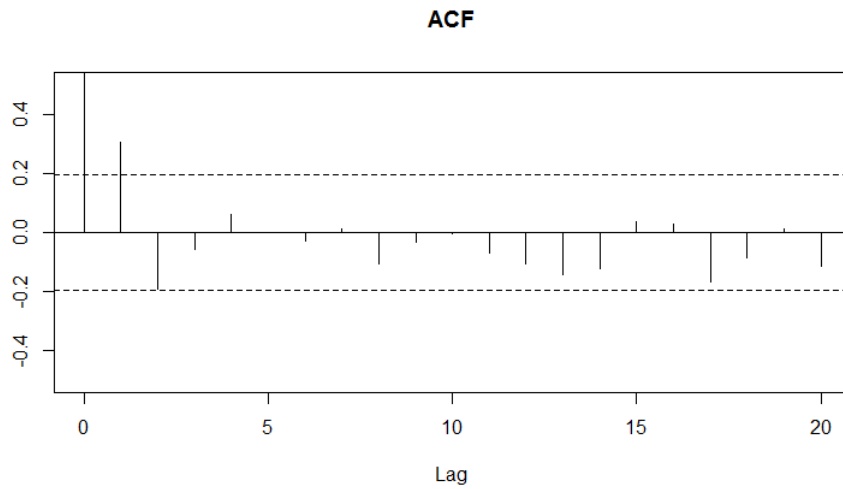


Figura 3.3: ACF de la sèrie estacionària de l'exemple anterior.

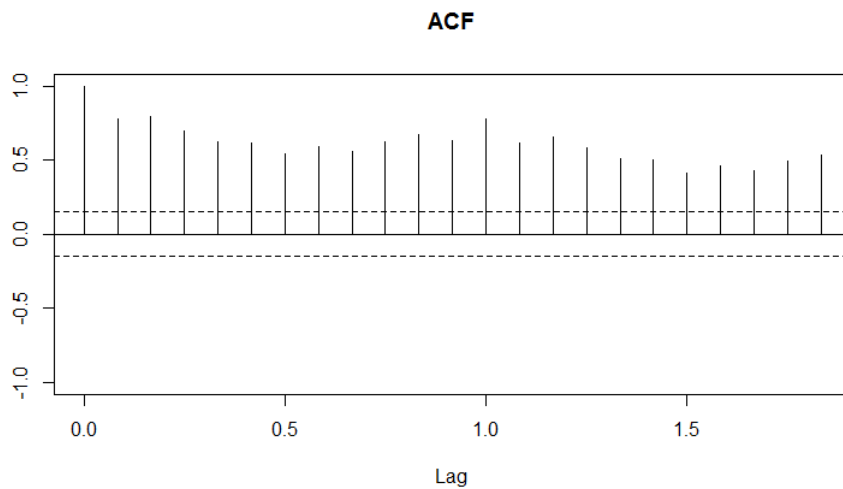


Figura 3.4: ACF de la sèrie no estacionària de l'exemple anterior.

En les dues figures anteriors podem observar clarament el que havíem comentat, que l'ACF de la sèrie estacionària decreix ràpidament cap a zero, i en canvi, l'ACF de la sèrie no estacionària decreix molt lentament cap a zero.

### 3.3 Models

En aquesta secció descriurem diversos tipus de processos estocàstics que són útils en la creació d'un model per a una sèrie temporal. Ens restringirem al cas a temps discret.

### 3.3.1 Soroll blanc

Un procés a temps discret  $\{Z_t : t \in T\}$  s'anomena un procés purament aleatori o soroll blanc si consisteix en una seqüència de variables aleatòries que són mútuament independents i idènticament distribuïdes. De la definició es dedueix que el procés té mitjana i variància constants i que

$$\gamma(k) = \text{Cov}(Z_t, Z_{t+k}) = 0 \quad \forall k = \pm 1, 2, \dots$$

Com que la mitjana i la funció d'autocovariància no depenen del temps, el procés és estacionari de segon ordre. De fet, és evident que el procés també és estrictament estacionari [12]. La funció d'autocorrelació ve donada per

$$\rho(k) = \begin{cases} 1 & k = 0 \\ 0 & k = \pm 1, \pm 2, \dots \end{cases}$$

Vegem-ne un exemple simulat en R (figura 3.5) amb el codi següent [1], on hem agafat les variables aleatòries amb distribució normal:

```
g=rnorm(100,0,1)
par(mfrow=c(1,2))
ts.plot(g,main="Nii(0,1)")
acf(g, main="ACF",ylim=c(-1,1),ci.col="black",
ylab="")
```

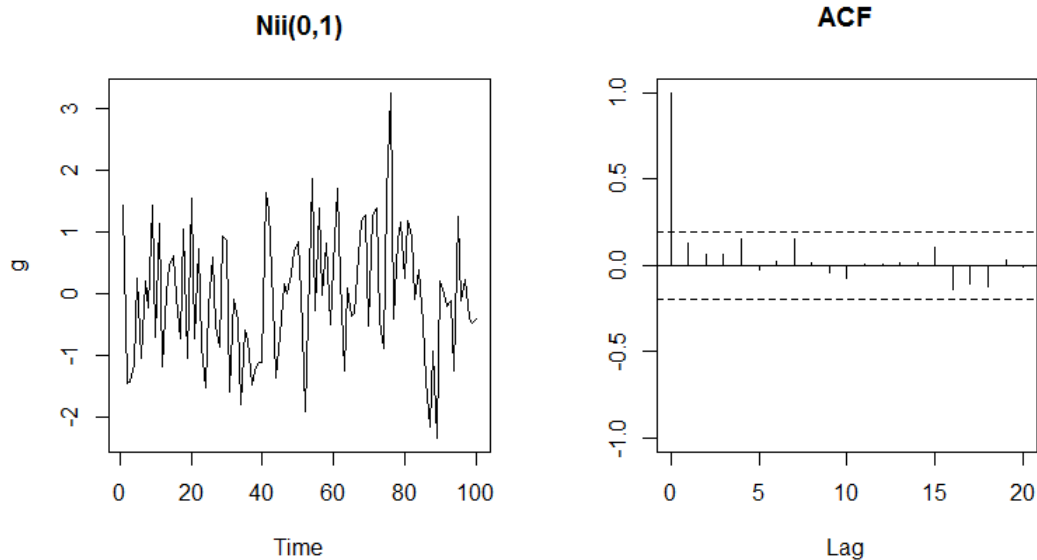


Figura 3.5: Soroll blanc gaussià amb mitjana 0 i variància constant igual a 1.

Com hem dit abans l'ACF es redueix a un únic segment de longitud unitària sobre l'eix d'ordenades.

### 3.3.2 Passejada aleatòria

Suposem que  $\{Z_t\}$  és un procés discret, purament aleatori de mitjana  $\mu$  i variància  $\sigma_Z^2$ . Un procés  $\{X_t\}$  es diu que és una passejada aleatòria si

$$X_t = X_{t-1} + Z_t$$

on s'acostuma a agafar  $X_0 = 0$ , de manera que

$$X_1 = Z_1$$

i

$$X_t = \sum_{i=1}^t Z_i \quad \forall t \geq 1.$$

Llavors resulta que  $E(X_t) = t\mu$  i  $Var(X_t) = t\sigma_Z^2$ , de manera que el procés no és estacionari.

És interessant observar que les primeres diferències d'una passejada aleatòria, donades per

$$\nabla X_t = X_t - X_{t-1} = Z_t$$

formen un procés que és un soroll blanc, que és per tant estacionari.

Vegem un exemple simulat en R (figura 3.6) amb el codi següent:

```
w=rnorm(100)
x=w
for (t in 2:100) x[t] <- x[t - 1] + w[t]
par(mfrow=c(2,2))
ts.plot(x, main="Passejada aleatoria Xt")
acf(x, main="ACF Xt",ylim=c(-.5,.5),ci.col="black",ylab="")
d=diff(x)
ts.plot(d,main="Primera diferència de Xt")
acf(d, main="ACF primera diferència",ylim=c(-.5,.5),ci.col="black",ylab="")
```

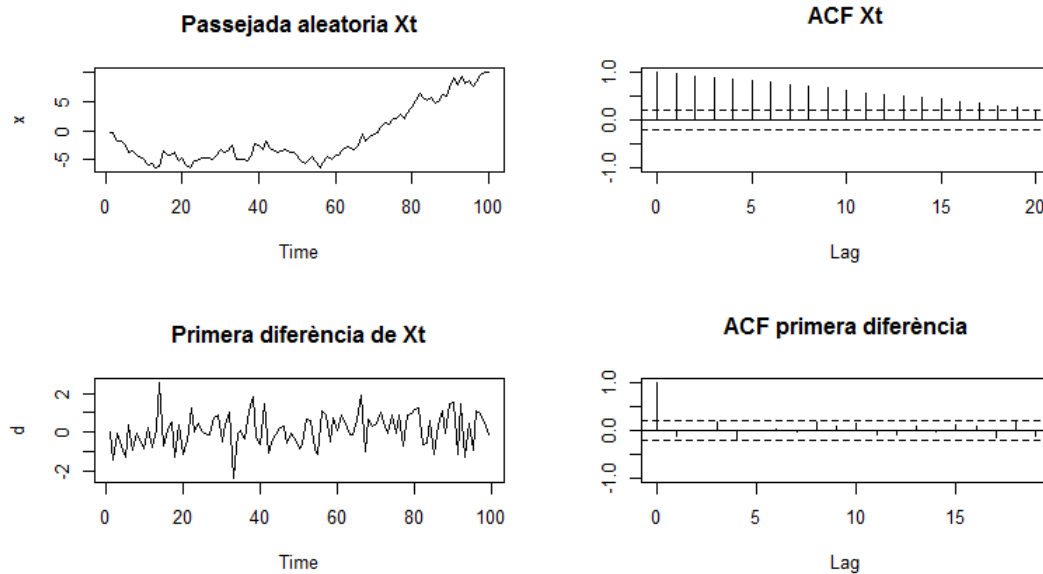


Figura 3.6: Procés estocàstic  $X_t$  i la primera diferència del procés, amb les ACF's corresponents.

En l'ACF del procés aleatori s'observa clarament la no estacionarietat de la sèrie, ja que decreix molt lentament cap a zero. En canvi, l'ACF de les primeres diferències decreix ràpidament cap a zero, fet que indica l'estacionarietat de la nova sèrie. A més, podem observar que el comportament del nou procés és com el d'un soroll blanc.

### 3.3.3 Mitjanes Mòbils (MA)

En aquest model el valor actual es pot predir a partir del comportament aleatori d'aquest instant, i en menor mesura, dels impulsos aleatoris anteriors.

Suposam que  $\{Z_t\}$  és un soroll blanc amb mitjana zero i variància  $\sigma_Z^2$ . Llavors un procés  $\{X_t\}$  s'anomena un procés de mitjanes mòbils d'ordre  $q$ , MA( $q$ ), si

$$X_t = \theta_0 Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}$$

on els  $\{\theta_i\}$  són constants. Els  $Z$ s normalment tenen l'escala modificada de manera que  $\theta_0 = 1$ .

Com que els  $Z$ s són independents, tenim que

$$E(X_t) = 0 \quad \text{i} \quad \text{Var}(X_t) = \sigma_Z^2 \sum_{i=0}^q \theta_i^2.$$

A més, també

$$\text{Cov}(Z_s, Z_t) = \begin{cases} \sigma_Z^2 & s = t \\ 0 & s \neq t \end{cases}$$

tenim que

$$\gamma(k) = \text{Cov}(X_t, X_{t+k}) = \text{Cov}(\theta_0 Z_t + \dots + \theta_q Z_{t-q}, \theta_0 Z_{t+k} + \dots + \theta_q Z_{t+k-q}) =$$

$$= \begin{cases} 0 & k > q \\ \sigma_Z^2 \sum_{i=0}^{q-k} \theta_i \theta_{i+k} & k = 0, 1, \dots, q \\ \gamma(-k) & k < 0 \end{cases}$$

Com que  $\gamma(k)$  no depèn de  $t$ , i la mitjana és constant, el procés és estacionari de segon ordre per a qualssevol valors de  $\{\theta_i\}$ . A més, si les  $Z_s$  estan distribuïdes normalment, llavors també ho estan les  $X_s$ , i tenim un procés estrictament estacionari.

Per tant, un procés de mitjanes mòbils,  $MA(q)$ , és sempre estacionari [12].

A partir de l'expressió anterior de la funció autocovariància d'un procés  $MA(q)$ , obtenim immediatament que la seva ACF ve donada per

$$\rho(k) = \begin{cases} 1 & k = 0 \\ \frac{\sum_{i=0}^{q-k} \theta_i \theta_{i+k}}{\sum_{i=0}^q \theta_i^2} & k = 1, \dots, q \\ 0 & k > q \\ \rho(-k) & k < 0 \end{cases}$$

A continuació s'explicarà la condició d'invertibilitat de la sèrie temporal. Considerem els processos  $MA(1)$  següents:

$$A: X_t = Z_t + \theta Z_{t-1}$$

$$B: X_t = Z_t + \frac{1}{\theta} Z_{t-1}$$

Es comprova immediatament que aquests dos processos diferents tenen exactament la mateixa funció d'autocorrelació. Per tant, no podem identificar un procés  $MA$  donant únicament la seva funció d'autocorrelació. Ara, si expressam els models  $A$  i  $B$  posant  $Z_t$  en termes de  $X_t, X_{t-1}, \dots$  (invertim el procés), trobam per substitució successiva que

$$A \quad Z_t = X_t - \theta X_{t-1} + \theta^2 X_{t-2} - \dots$$

$$B \quad Z_t = X_t - \frac{1}{\theta} X_{t-1} + \frac{1}{\theta^2} X_{t-2} - \dots$$

Si  $|\theta| < 1$ , la sèrie  $A$  convergeix mentre que la  $B$  no ho fa. Llavors, si  $|\theta| < 1$ , el model  $A$  es diu que és invertible mentre que  $B$  no ho és. La imposició de la condició d'invertibilitat assegura que hi ha un únic procés  $MA$  per a una determinada funció d'autocorrelació [12].

A continuació introduïrem un operador que és molt útil per estudiar els diferents models per a sèries temporals.

**Definició 3.3.1** *Definim l'operador de retard  $B$  per  $BX_t = X_{t-1}$ .*

Es pot estendre a  $B^2 X_t = B(BX_t) = BX_{t-1} = X_{t-2}$ , i així successivament. Llavors,

$$B^k X_t = X_{t-k}.$$

Així, utilitzant l'operador de retard  $B$ , un procés  $MA(q)$  es pot escriure en la forma

$$X_t = (\theta_0 + \theta_1 B + \dots + \theta_q B^q) Z_t$$

o més breument com

$$X_t = \theta(B)Z_t$$

on  $\theta(B) = \theta_0 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$ .

Es pot demostrar que un procés MA(q) és invertible si les arrels del polinomi  $\theta(B) = \theta_0 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$  cauen fora del cercle unitat. Aquesta condició és equivalent a que les arrels de l'equació:  $\theta_0 X^q + \theta_1 X^{q-1} + \dots + \theta_{q-1} X + \theta_q = 0$  siguin totes inferiors a 1 en mòdul [10].

Vegem un exemple d'un procés MA(1) amb  $\theta = 0.5$  (figura ), simulat en R [1] amb el següent codi:

```
#Simulació d'un procés MA(1) amb theta=0.5
layout(matrix(c(1,1,2,3), 2, 2, byrow = TRUE))
MA=arima.sim(list(order=c(0,0,1), ma=.5), n=100)
plot(MA, ylab=" ", main=(expression(MA(1)~~~theta==+.5)))
acf(MA, main="ACF",ylim=c(-.5,.5),ci.col="black",ylab="")
```

L'expressió del procés és:  $X_t = Z_t + 0.5Z_{t-1}$ .

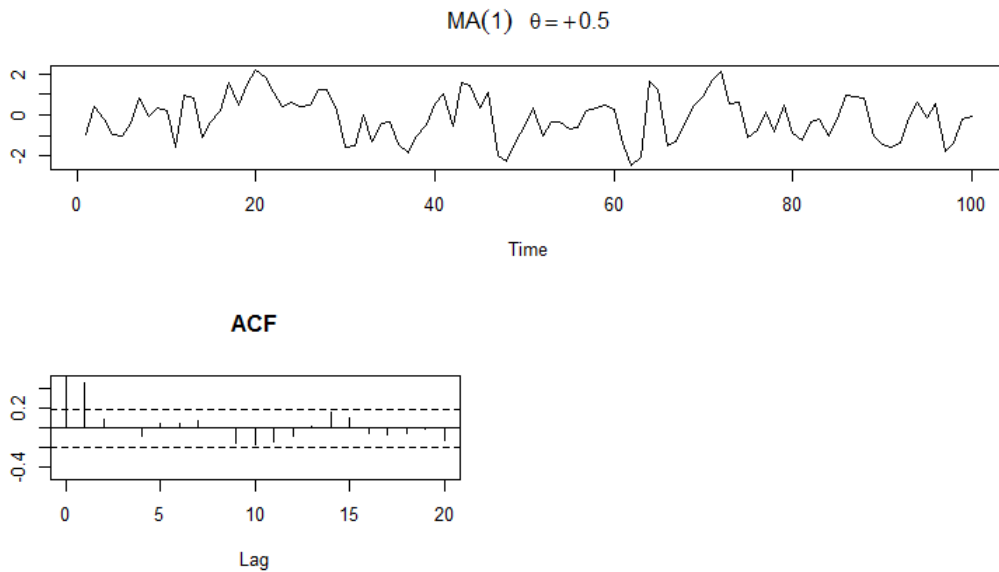


Figura 3.7: Procés MA(1) amb l'ACF corresponent.

Observem que l'ACF no és significativa a partir del segon retard ( $k = 2$ ).

Vegem ara un exemple d'un procés MA(2) amb  $\theta_1 = 0.8$  i  $\theta_2 = 0.6$  (figura 3.8):

```
#Simulació d'un procés MA(2) camb theta=(0.8,0.6)
layout(matrix(c(1,1,2,3), 2, 2, byrow = TRUE))
MA=arima.sim(list(order=c(0,0,2), ma=c(.8,.6)), n=100)
plot(MA, ylab=" ", main=(expression(MA(1)~~~theta[1]==+.8~~~
```

```
theta[2]==.6))
acf(MA, main="ACF",ylim=c(-.5,.5),ci.col="black",ylab="")
```

L'expressió del procés és:  $X_t = Z_t + 0.8Z_{t-1} + 0.6Z_{t-2}$ .

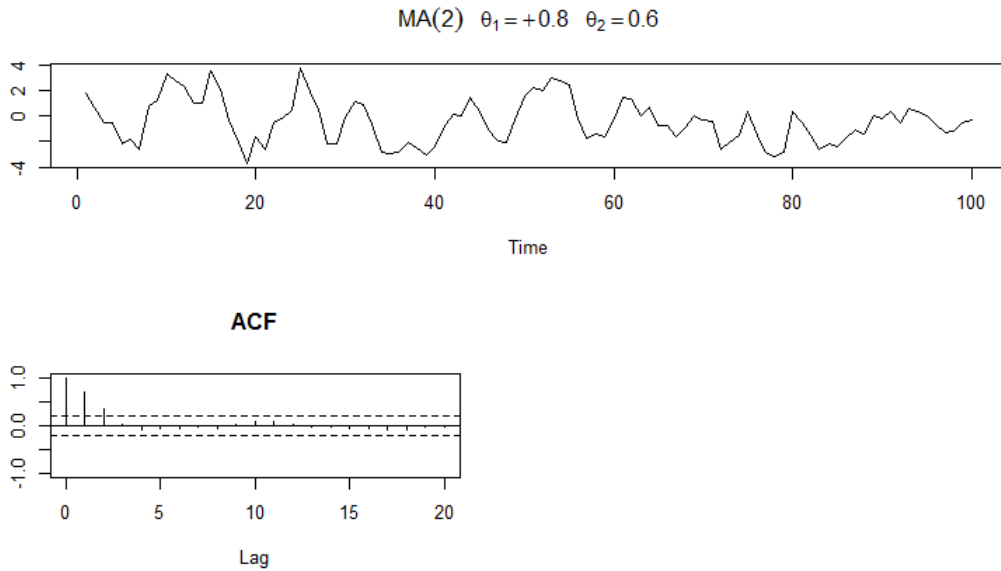


Figura 3.8: Procés MA(2) amb l'ACF corresponent.

Observem que ara l'ACF no és significativa a partir del tercer retard ( $k = 3$ ). Aquest fet és general: per a un procés MA( $q$ ), hi ha  $q$  valors significatius i per  $k > q$  els valors de l'ACF són tan petits (entre les franges de significació) que es consideren nuls. Per tant, a través del gràfic podem saber que l'ordre del procés MA és  $q$ , ja que tenim  $q$  valors de l'ACF no nuls.

### 3.3.4 Processos Autoregressius (AR)

Un model autoregressiu descriu un tipus particular de procés en que les observacions en un moment donat són previsible a partir de les observacions passades del procés més un terme d'error.

Suposam que  $\{Z_t\}$  és un soroll blanc amb mitjana zero i variància  $\sigma_Z^2$ . Llavors un procés  $\{X_t\}$  s'anomena un procés autoregressiu d'ordre  $p$ , AR( $p$ ), si

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + Z_t. \quad (3.1)$$

on  $\phi_1, \dots, \phi_p \in \mathbb{R}$ .

Observem que  $X_t$  es pot considerar com una regressió lineal múltiple; en aquest cas, les variables explicatives són les  $p$  observacions passades del propi procés:  $X_{t-1}, \dots, X_{t-p}$ . Per això, el procés s'anomena autoregressiu.

Es pot expressar un procés AR d'ordre finit com un procés MA d'ordre infinit. Això es pot fer mitjançant la substitució successiva, o utilitzant l'operador de retard B. Llavors,



un procés AR(p) pot posar-se mitjançant l'operador de retard B en la forma:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) X_t = Z_t$$

o

$$X_t = Z_t / (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) = \phi(B) Z_t,$$

on  $\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)^{-1} = 1 + \theta_1 B + \theta_2 B^2 + \dots$ .

Llavors, es pot trobar la relació entre les  $\phi$ s i les  $\theta$ s. Al tenir  $X_t$  expressat com un procés MA, es dedueix que  $E(X_t) = 0$ . La variància finita requereix que  $\sum \theta_i^2$  convergeixi, i aquesta és una condició necessària per la estacionarietat de la sèrie temporal.

Per trobar l'ACF, la forma més senzilla és assumir que el procés és estacionari, llavors es multiplica l'equació (3.1) per  $X_{t-k}$ , s'agafen esperances i es divideix per  $\sigma_X^2$ , suposant que la variància de  $X_t$  és finita. Després, utilitzant el fet que  $\rho(k) = \rho(-k)$  per a tot  $k$ , obtenim

$$\rho(k) = \phi_1 \rho(k-1) + \dots + \phi_p \rho(k-p) \text{ per a tot } k > 0.$$

Aquest conjunt d'equacions s'anomenen les equacions de Yule-Walker degut a G.U. Yule i Gilbert Walker. Es tracta d'un conjunt d'equacions en diferències que té la solució general

$$\rho(k) = A_1 \pi_1^{|k|} + \dots + A_p \pi_p^{|k|}$$

on  $\{\pi_i\}$  són les arrels de l'anomenada equació auxiliar

$$y^p - \phi_1 y^{p-1} - \dots - \phi_p = 0$$

Les constants  $\{A_i\}$  es trien per satisfer les condicions inicials  $\rho(0) = 1$ , el que significa que  $\sum A_i = 1$ . Les primeres  $(p-1)$  equacions de Yule-Walker proporcionen  $(p-1)$  noves restriccions a les  $\{A_i\}$  utilitzant  $\rho(0) = 1$  i  $\rho(k) = \rho(-k)$ .

Per la forma general de  $\rho(k)$ , és evident que  $\rho(k)$  tendeix a zero a mesura que  $k$  augmenta sempre que  $|\pi_i| < 1$  per a tot  $i$ , i aquesta és una condició necessària i suficient perquè el procés sigui estacionari [12].

Una forma equivalent d'expressar l'estacionarietat és si les arrels del polinomi  $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$  cauen fora del cercle unitat. Aquesta condició és equivalent a que les arrels de l'equació:  $X^p - \phi_1 X^{p-1} - \phi_2 X^{p-2} - \dots - \phi_{p-1} X - \phi_p = 0$  siguin totes inferiors a 1 en mòdul.

A més, un procés autoregressiu sempre és invertible [10].

Per exemple, considerem el cas  $p = 1$ , és a dir, el procés AR(1), també anomenat procés de Markov.

Tenim  $X_t = \phi X_{t-1} + Z_t$  amb  $|\phi| < 1$ . Aleshores

$$(1 - \phi B) X_t = Z_t.$$

I per tant,

$$X_t = \frac{1}{1 - \phi B} Z_t = (1 + \phi B + \phi^2 B^2 + \dots) Z_t$$

Així,

$$\text{Var}(X_t) = \sigma_Z^2 (1 + \phi^2 + \phi^4 + \dots) = \frac{\sigma_Z^2}{1 - \phi^2}.$$

D'altra banda,

$$\gamma(k) = E[X_t X_{t+k}] = E\left\{\left[\sum \phi^i Z_{t-i}\right] \left[\sum \phi^j Z_{t+k-j}\right]\right\} = \sigma_Z^2 \sum_{i=0}^{\infty} \phi^i \phi^{k+i} = \sigma_Z^2 \frac{\phi^k}{1-\phi^2}.$$

Aleshores,

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \phi^k, \quad k = 0, 1, 2, \dots$$

Vegem un exemple d'un procés AR(1) amb  $\phi = 0.4$  (figura 3.9), simulat en R amb el codi següent:

```
# Simulació d'un procés AR(1) amb phi=0.4
layout(matrix(c(1,1,2,3), 2, 2, byrow = TRUE))
AR=arima.sim(list(order=c(1,0,0), ar=.4), n=100)
plot(AR, ylab=" ", main=(expression(AR(1) ~~~ phi ==+.4)))
acf(AR, main="ACF", ylim=c(-.5, .5), ci.col="black", ylab="")
```

L'expressió del procés és:  $X_t = 0.4X_{t-1} + Z_t$ .

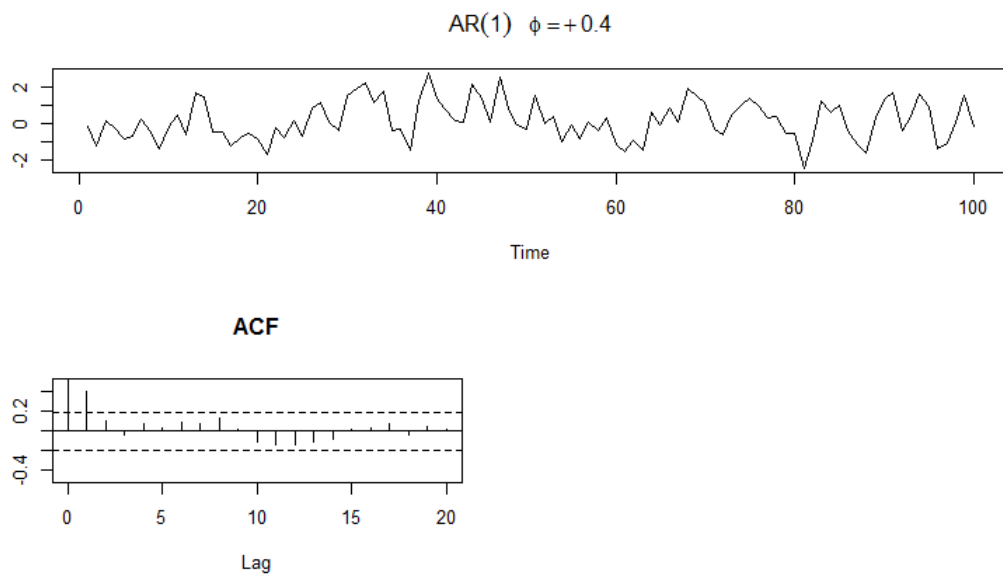


Figura 3.9: Procés AR(1) amb l'ACF corresponent.

Vegem un altre exemple d'un procés AR(1), en aquest cas amb  $\phi = -0.4$  (figura 3.10). L'expressió del procés és:  $X_t = -0.4X_{t-1} + Z_t$ .

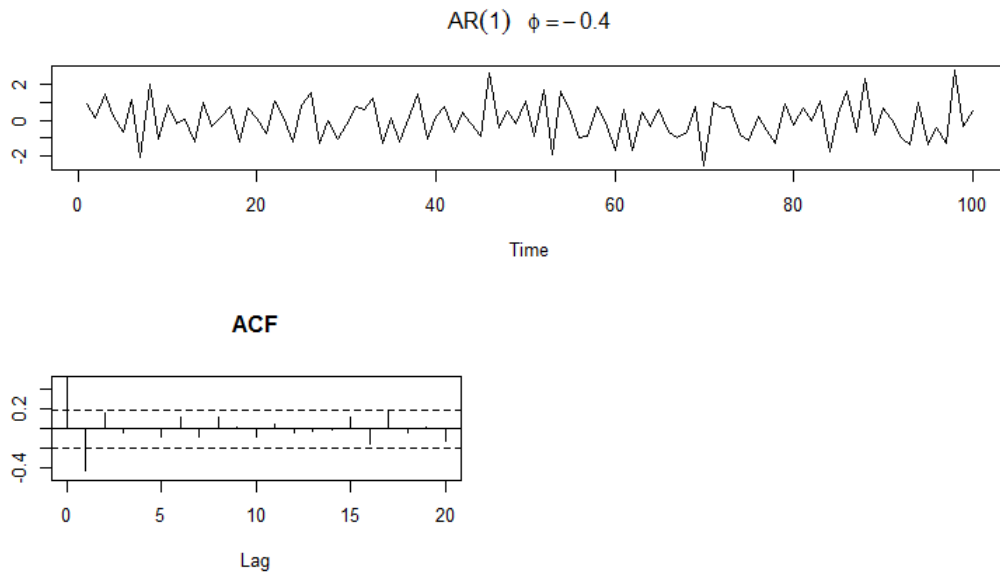


Figura 3.10: Procés AR(1) amb l'ACF corresponent.

Aquest comportament que hem vist en els dos exemples anteriors és general: l'ACF d'un procés AR(1) mostra un comportament amortitzant cap a zero amb tots els valors positius, en cas que  $\phi > 0$ , o bé alternant el signe, començant amb negatiu, si  $\phi < 0$ .

Vegem ara un exemple d'un procés AR(2) amb  $\phi_1 = 0.6$  i  $\phi_2 = 0.2$  (figura 3.11):

```
# Simulació d'un procés AR(2) amb phi=(0.6,0.2)
layout(matrix(c(1,1,2,3), 2, 2, byrow = TRUE))
AR=arima.sim(list(order=c(2,0,0), ar=c(.6,.2)), n=100)
plot(AR, ylab=" ", main=(expression(AR()~~~phi[1]==+.6~~~phi[2]==+.2)))
acf(AR, main="ACF",ylim=c(-.5,.5),ci.col="black",ylab="")
```

L'expressió del procés és:  $X_t = 0.6X_{t-1} + 0.2X_{t-2} + Z_t$ .

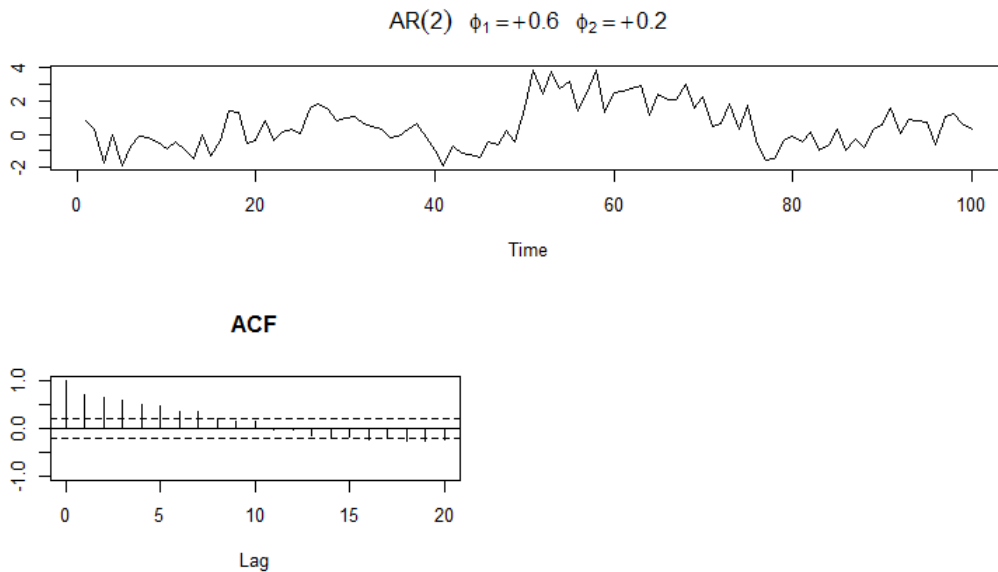


Figura 3.11: Procés AR(2) amb l'ACF corresponent.

En general, l'ACF d'un procés AR(p) mostra un decreixement ràpid exponencial atenuat o ones sinusoidals.

### 3.3.5 La funció d'autocorrelació parcial (PACF)

Hem vist en els models MA(q) que l'ACF serà zero per als desfasaments més grans que q. D'altra banda, a causa de les condicions d'estacionarietat, les funcions d'autocorrelació de processos autoregressius estacionaris d'ordre finit són sempre seqüències que convergeixen a zero però no es trenquen. Això fa que sigui difícil distingir entre processos de diferents ordres quan s'utilitza la funció d'autocorrelació. Per fer front a aquest problema, s'introdueix un nou concepte, la funció d'autocorrelació parcial, PACF. La correlació parcial entre dues variables aleatòries és la correlació que queda si el possible impacte de totes les altres variables aleatòries ha estat eliminat.

Un altre problema que presenta l'ACF és que si per exemple  $\phi_1$  és diferent de zero, llavors existeix una cadena d'influències separades per un retard.

$$X_1 \rightarrow X_2 \rightarrow \dots \rightarrow X_{t-1} \rightarrow X_t \rightarrow X_{t+1} \rightarrow \dots$$

Però si  $X_1 \rightarrow X_2$  i  $X_2 \rightarrow X_3$ , llavors  $X_1 \rightarrow X_3$ . Per tant, en general l'ACF, si  $\rho_1$  és diferent de zero, llavors trobarà que  $\rho_2, \rho_3, \dots$ , són diferents de zero.

Però és necessari distingir entre diferents cadenes d'influència possibles:

- La cadena d'influència general, a través de  $\rho_1$ .
- Les cadenes d'influència directa. Per exemple, com influeix  $X_1$  sobre  $X_3$  directament, és a dir, sense passar a través de  $X_2$ .

Per resoldre aquest problema es construeix la funció d'autocorrelació parcial, anomenada PACF.

La funció d'autocorrelació parcial proporciona la relació directa que existeix entre observacions separades  $k$  retards. Aquesta informació sobre l'estructura de la sèrie és molt valuosa, ja que elimina el problema que presentava l'ACF de que si  $X_1 \rightarrow X_2$  i  $X_2 \rightarrow X_3$ , llavors  $X_1 \rightarrow X_3$ . En l'ACF, la relació entre  $X_1$  i  $X_2$  o  $X_2$  i  $X_3$  serà significativa. I la relació entre  $X_1$  i  $X_3$  també.

En la PACF això no passa, la relació entre  $X_1$  i  $X_2$  o  $X_2$  i  $X_3$  serà significativa. Però la relació entre  $X_1$  i  $X_3$  no ho serà [13].

Vegem a continuació com definir la PACF.

Si  $X$ ,  $Y$  i  $Z$  són variables aleatòries, llavors la correlació parcial entre  $X$  i  $Y$  donat  $Z$  s'obté mitjançant la regressió de  $X$  sobre  $Z$  per obtenir  $\bar{X}$ , la regressió de  $Y$  sobre  $Z$  per obtenir  $\bar{Y}$ , i després calculant

$$\rho_{XY|Z} = \text{Corr}(X - \bar{X}, Y - \bar{Y}).$$

La idea és que  $\rho_{XY|Z}$  mesura la correlació entre  $X$  i  $Y$  amb l'efecte lineal de  $Z$  eliminat (o parcialment fora).

Per definir formalment la PACF per a la sèrie temporal estacionària de mitjana zero, sigui  $X_t^{h-1}$  la regressió de  $X_h$  sobre  $\{X_{h-1}, X_{h-2}, \dots, X_1\}$ , que escrivim com

$$X_h^{h-1} = \beta_1 X_{h-1} + \beta_2 X_{h-2} + \dots + \beta_{h-1} X_1. \quad (3.2)$$

A (3.2) no és necessari terme d'intersecció, ja que la mitjana de  $X_t$  és zero. A més, sigui  $X_0^{h-1}$  la regressió de  $X_0$  sobre  $\{X_1, X_2, \dots, X_{h-1}\}$ , llavors

$$X_0^{h-1} = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{h-1} X_{h-1}. \quad (3.3)$$

Es pot provar que els coeficients  $\beta_1, \beta_2, \dots, \beta_{h-1}$  són els mateixos en (3.2) i (3.3) [8].

**Definició 3.3.2** La funció d'autocorrelació parcial (PACF) d'un procés,  $X_t$ , que es denota per  $\phi_{hh}$ , per  $h = 1, 2, \dots$  es defineix com

$$\phi_{11} = \text{corr}(X_1, X_0) = \rho(1)$$

i

$$\phi_{hh} = \text{corr}(X_h - X_h^{h-1}, X_0 - X_0^{h-1}), \quad h \geq 2.$$

$X_h - X_h^{h-1}$  i  $X_0 - X_0^{h-1}$  no estan correlacionats amb  $\{X_1, X_2, \dots, X_{h-1}\}$ . Per estacionarietat, la PACF,  $\phi_{hh}$ , és la correlació entre  $X_t$  i  $X_{t-h}$  amb la dependència lineal de  $\{X_{t-1}, \dots, X_{t-(h-1)}\}$ , amb cada un, eliminat [8].

El coeficient d'autocorrelació parcial en el retard  $k$ ,  $\phi_{kk}$ , és la correlació entre  $X_t$  i  $X_{t-k}$  després d'extreure l'influència dels retards intermedis.

El càlcul de les autocorrelacions parcials pot basar-se en el model de regressió múltiple en desviacions respecte a les mitjanes poblacionals

$$\bar{X}_t = \phi_{1k} \bar{X}_{t-1} + \dots + \phi_{kk} \bar{X}_{t-k} + u_t.$$

Multiplicant el model per  $\bar{X}_{t-k}$  i fent l'esperança obtenim

$$E(\bar{X}_t \bar{X}_{t-k}) = \phi_{1k} E(\bar{X}_{t-1} \bar{X}_{t-k}) + \dots + \phi_{kk} E(\bar{X}_{t-k} \bar{X}_{t-k}).$$

d'on tenim

$$\gamma_k = \phi_{1k}\gamma_{k-1} + \dots + \phi_{kk}\gamma_{t-k}.$$

Dividint per  $\gamma_0$  podem especificar el sistema d'equacions de Yule-Walker com

$$\begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{pmatrix} = \begin{pmatrix} \rho_0 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & \rho_0 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_0 \end{pmatrix} \begin{pmatrix} \phi_{1k} \\ \phi_{2k} \\ \vdots \\ \phi_{kk} \end{pmatrix}$$

que ens permet obtenir l'autocorrelació parcial  $\phi_{kk}$  en termes de les autocorrelacions  $\rho_1, \dots, \rho_k$ . Aplicant la regla de Cramer, tenim que

$$\phi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{vmatrix}}$$

El primer terme de la funció d'autocorrelació parcial,  $\phi_{11}$ , es pot estimar transformant la sèrie  $X_t$  en desviacions respecte a la seva mitjana mostral  $Y_t = X_t - \bar{X}$  i a continuació estimant una regressió de  $Y_t$  sobre  $Y_{t-1}$ . El model de regressió és  $Y_t = \phi_{11}Y_{t-1} + u_t$ , on la pendent estimada d'aquesta regressió és  $\phi_{11}$  i  $u_t$  és l'ordenada a l'origen de la regressió lineal, que ha de ser un procés de mitjana 0, ja que  $EY_t = 0$ .

El segon valor de la PACE,  $\phi_{22}$ , s'estima mitjançant una regressió de  $Y_t$  sobre  $Y_{t-1}$  i  $Y_{t-2}$ . El model de regressió és  $Y_t = \phi_{21}Y_{t-1} + \phi_{22}Y_{t-2} + u_t$ .

El tercer valor de la PACE,  $\phi_{33}$ , s'estima mitjançant una regressió de  $Y_t$  sobre  $Y_{t-1}$ ,  $Y_{t-2}$  i  $Y_{t-3}$ . El model de regressió és  $Y_t = \phi_{31}Y_{t-1} + \phi_{32}Y_{t-2} + \phi_{33}Y_{t-3} + u_t$ .

Així, la funció d'autocorrelació parcial pot estimar-se a través d'una sèrie de regressions, on cada una conté com a variable explicativa un retard més que l'anterior, i en cada cas s'elegeixen els coeficients estimats en els retards més alts  $\{\phi_{11}, \phi_{22}, \phi_{33}, \dots\}$ , que són els valors estimats de la PACF [15].

En el cas d'un procés AR(1), hem vist que  $\rho_k = \phi^k$ , de manera que

$$\phi_{11} = \rho_1 = \phi$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\begin{vmatrix} 1 & \phi \\ \phi & \phi^2 \end{vmatrix}}{\begin{vmatrix} 1 & \phi \\ \phi & 1 \end{vmatrix}} = 0$$

i, en general,  $\phi_{kk} = 0 \forall k > 1$ .

Vegem l'exemple del procés AR(1), amb expressió :  $X_t = 0.4X_{t-1} + Z_t$ , amb la PACF corresponent (figura 3.12), s'ha simulat en R [1] amb el codi següent:

```
# Simulació d'un procés AR(1) amb phi=0.4
layout(matrix(c(1,1,2,3), 2, 2, byrow = TRUE))
AR=arima.sim(list(order=c(1,0,0), ar=.4), n=100)
plot(AR, ylab=" ", main=(expression(AR(1)~~~phi==+.4)))
acf(AR, main="ACF",ylim=c(-.5,.5),ci.col="black",ylab="")
pacf(AR,main="PACF",ylim=c(-.5,.5),ci.col="black")
```

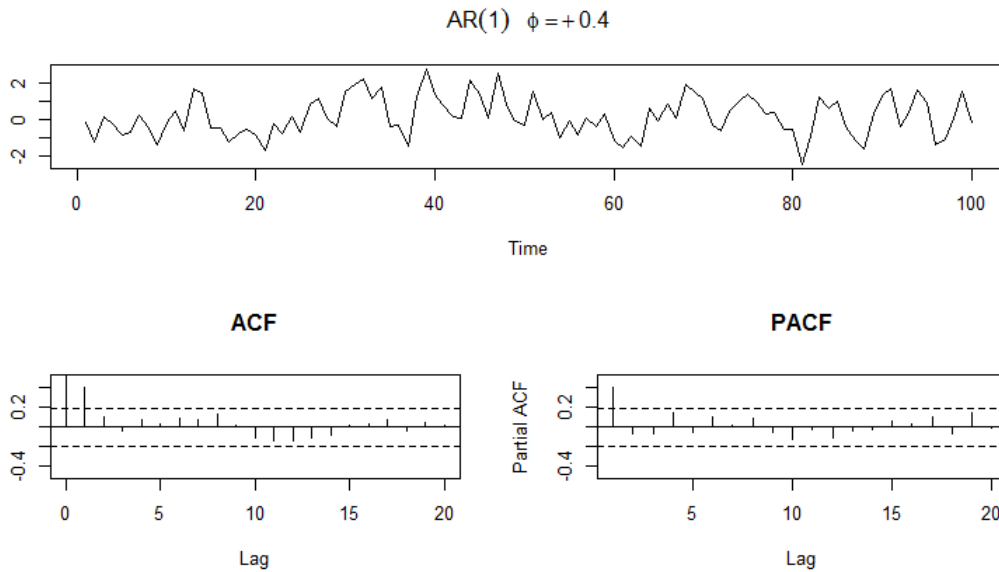


Figura 3.12: Procés AR(1) amb l'ACF i la PACF corresponents.

Es pot observar que la PACF s'anul·la per retards superiors a 1 (l'ordre del model), per tant es comporta com l'ACF per als models MA.

Vegem ara l'exemple del procés AR(2) també amb la PACF corresponent (figura 3.13):

```
# Simulació d'un procés AR(2) amb phi=(0.6,0.2)
layout(matrix(c(1,1,2,3), 2, 2, byrow = TRUE))
AR=arima.sim(list(order=c(2,0,0), ar=c(.6,.2)), n=100)
plot(AR, ylab=" ", main=(expression(AR()~~~phi[1]==+.6~~~phi[2]==+.2)))
acf(AR, main="ACF",ylim=c(-.5,.5),ci.col="black",ylab="")
pacf(AR,main="PACF",ylim=c(-.5,.5),ci.col="black")
```

Recordem que l'expressió del procés era:  $X_t = 0.6X_{t-1} + 0.2X_{t-2} + Z_t$ .

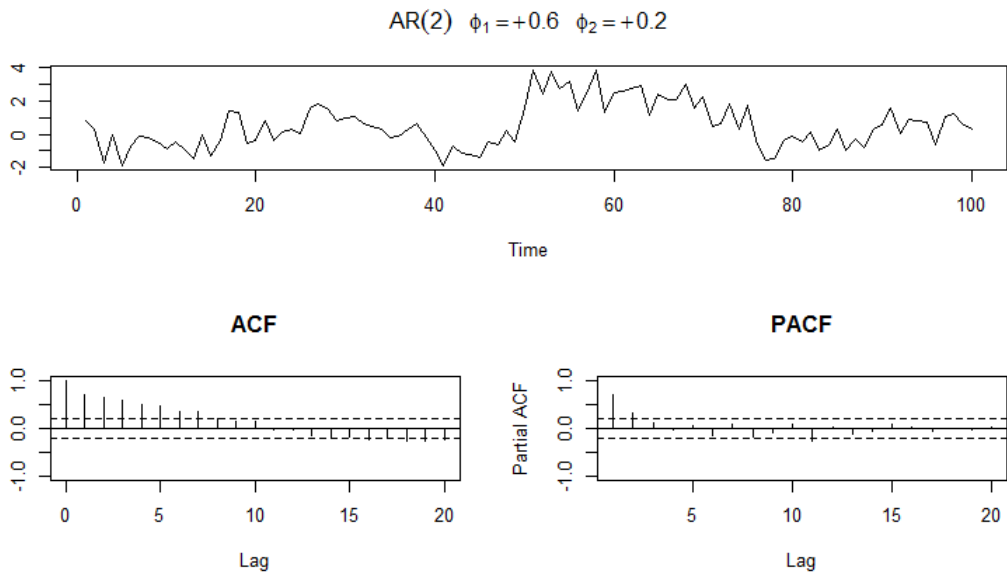


Figura 3.13: Procés AR(2) amb les ACF i PACF corresponents.

Observem que ara la PACF només té dos valors significatius. Això és un fet general: la PACF d'un procés AR( $p$ ) només té els  $p$  primers coeficients significatius, els altres s'anul·len, és a dir, té els coeficients igual a zero per als retards majors que  $p$  [8].

Ara vegem com es comporta la PACF dels models MA. Vegem els exemples dels processos MA(1) i MA(2) (figures 3.14 i 3.15).

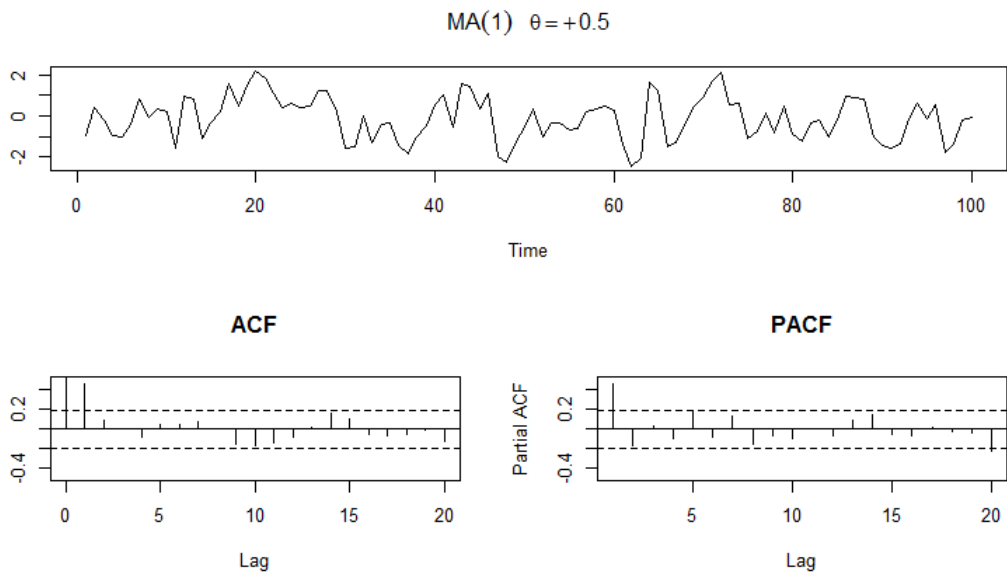


Figura 3.14: Procés MA(1), que és  $X_t = Z_t + 0.5Z_{t-1}$ , amb les ACF i PACF corresponents.



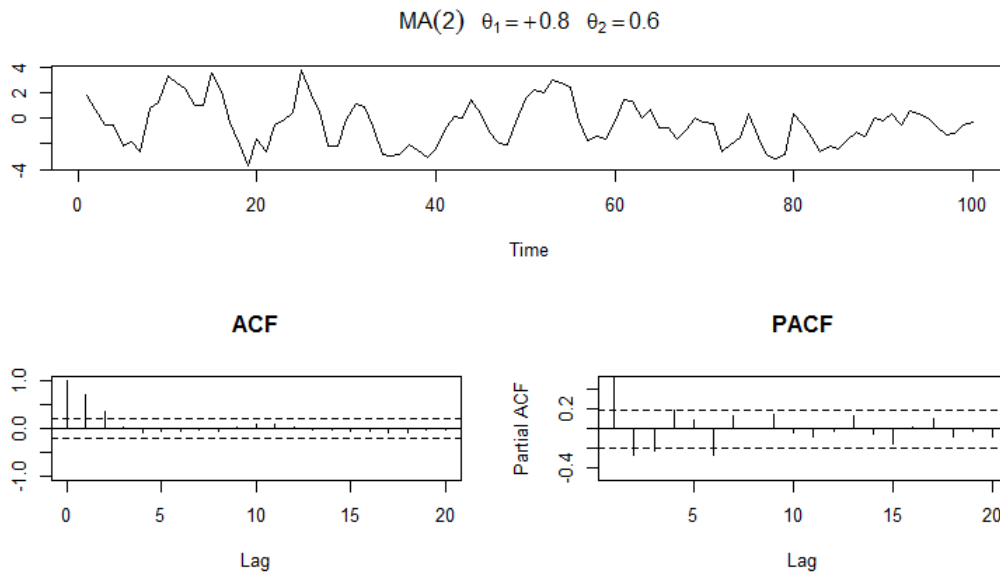


Figura 3.15: Procés MA(2), que és  $X_t = Z_t + 0.8Z_{t-1} + 0.6Z_{t-2}$ , amb la ACF i la PACF corresponents.

La PACF dels models MA està dominada per una mescla d'exponencials i oscil·lacions amortitzants, per tant, té el mateix comportament que l'ACF dels AR.

La taula següent mostra el comportament dels processos AR(p) i MA(q) respecte de les seves ACF i PACF.

	Estacionarietat	Invertibilitat	ACF	PACF
MA(q)	Sempre	Si les arrels de $\theta(B) = 0$ fora del cercle unitat	Igual a 0 després de q retards	Estructura exponencial o sinusoidal
AR(p)	Si les arrels de $\phi(B) = 0$ fora del cercle unitat	Sempre	Estructura exponencial o sinusoidal	Igual a 0 després de p retards

Taula 3.1: Comparació-resum dels AR(p) i MA(q).

Considerem ara un exemple d'ACF i PACF d'un procés MA(1) i un AR(1), respectivament.

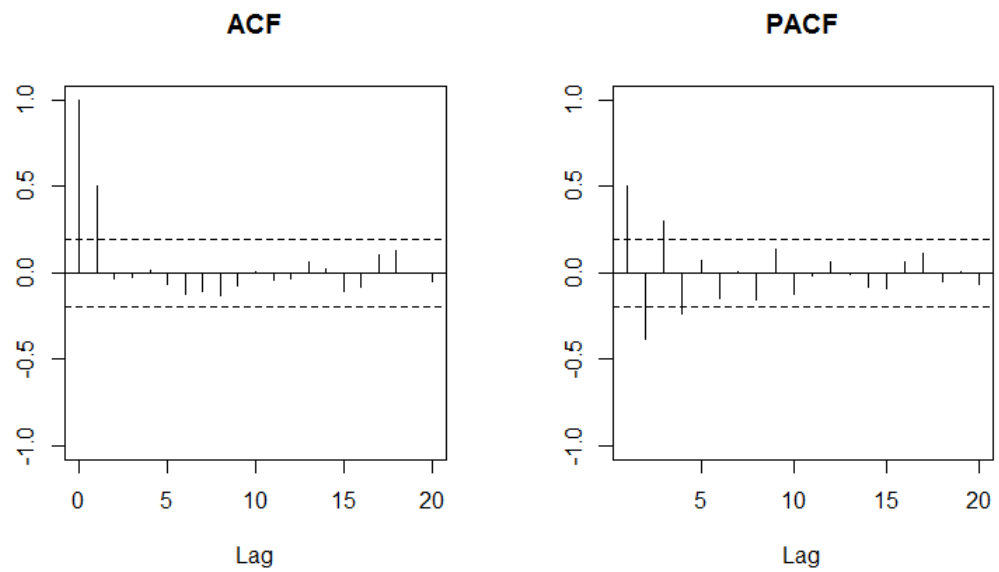


Figura 3.16: Les ACF i PACF corresponents d'un procés MA(1).

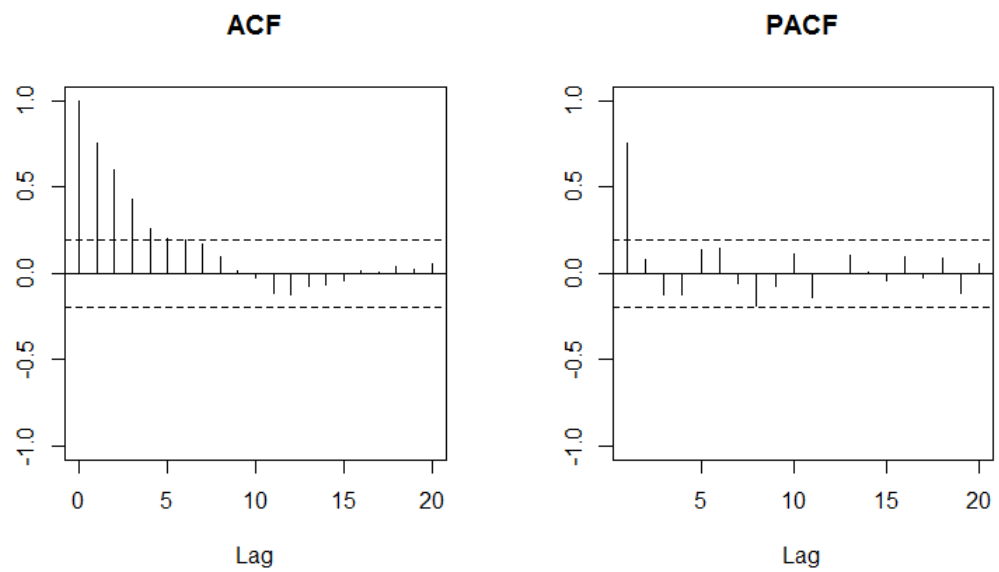


Figura 3.17: Les ACF i PACF corresponents d'un procés AR(1).

En aquests exemples podem comprovar el que es resumeix en la taula anterior. L'ACF s'anul·la després de  $q = 1$  en el cas del MA i la PACF segueix una estructura sinusoidal. I el procés AR presenta la PACF nul·la a partir de  $p = 1$  i l'ACF segueix una estructura sinusoidal.

### 3.3.6 Processos ARMA

Una extensió natural dels models AR(p) i MA(q) és un tipus de model que inclou tant termes autoregressius com de mitjanes mòbils i es defineixen com ARMA(p,q). Llavors, un procés ARMA(p,q) tindrà  $p$  termes autoregressius i  $q$  termes de mitjanes mòbils. Aquest ve donat per

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}$$

on els paràmetres  $p$  i  $q$  s'anomenen els ordres autoregressius i de mitjanes mòbils, respectivament.

Així, com s'ha indicat anteriorment, quan  $q = 0$ , el model s'anomena model autoregressiu d'ordre  $p$ , AR(p), i quan  $p = 0$ , el model s'anomena model de mitjanes mòbils d'ordre  $q$ , MA(q).

El model ARMA(p,q) es pot escriure en termes de l'operador de retards com

$$(1 - \phi_1 B - \dots - \phi_p B^p) X_t = (1 + \theta_1 B + \dots + \theta_q B^q) Z_t$$

o més breument com

$$\phi_p(B) X_t = \theta_q(B) Z_t \quad (3.4)$$

on  $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$  és el polinomi autoregressiu i  $\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$  és el polinomi de mitjanes mòbils.

El procés ARMA(p,q) és estacionari si ho és la seva component autoregressiva, i és invertible si ho és la seva component de mitjanes mòbils.

Llavors, un model ARMA(p,q) és estacionari si les arrels del polinomi  $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$  cauen fora del cercle unitat.

Anàlogament és invertible si les arrels del polinomi  $\theta(B) = \theta_0 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$  cauen fora del cercle unitat.

El model ARMA(p,q) té mitjana zero, i variància constant i finita. La funció d'autocorrelació és infinita, decreixent ràpidament cap a zero. És una mescla d'exponencials i oscil·lacions amortitzant després de  $k = q - p$ . I la PACF també és infinita, és una mescla d'exponencials i oscil·lacions amortitzant després de  $k = p - q$  [1].

La importància d'un procés ARMA està en el fet que una sèrie temporal estacionària pot ser descrita sovint mitjançant un model ARMA que implica menor nombre de paràmetres que un MA pur o un procés AR per si mateix.

A vegades pot ser útil expressar un model ARMA com un procés MA pur en la forma

$$X_t = \psi(B) Z_t \quad (3.5)$$

on  $\psi(B) = \sum \psi_i B^i$  és l'operador de MA que pot ser d'ordre finit. Els pesos  $\psi$ ,  $\{\psi_i\}$ , poden ser útils en el càlcul de les previsions i en l'avaluació de les propietats d'un model. Comparant amb l'equació (3.4), veim que  $\psi(B) = \theta_q(B) / \phi_p(B)$ . Alternativament, pot ser útil expressar un model ARMA com un procés AR pur en la forma

$$\pi(B) X_t = Z_t \quad (3.6)$$

on  $\pi(B) = \phi_p(B) / \theta_q(B)$ . Per convenció s'escriu  $\pi(B) = 1 - \sum_{i \geq 1} \pi_i B^i$ , ja que la forma natural per escriure un model AR és de la forma

$$X_t = \sum_{i=1}^{\infty} \pi_i X_{t-i} + Z_t.$$

Comparant (3.5) i (3.6), podem veure que

$$\pi(B)\psi(B) = 1.$$

Els pesos  $\psi$  o els pesos  $\pi$  poden ser obtinguts directament dividint o igualant potències de  $B$  en l'equació donada per

$$\psi(B)\phi(B) = \theta(B).$$

Considerem un procés ARMA(1,1) simulat en R amb el següent codi:

```
#Simulació d'un procés ARMA(1,1) amb phi=0.8 i theta=-0.3
layout(matrix(c(1,1,2,3), 2, 2, byrow = TRUE))
ARMA=arima.sim(list(order=c(1,0,1), ar=.8, ma=-.3), n=200)
plot(ARMA, ylab=" ", main=(expression(MA(1)~~~phi==+.8~~~theta==-.3)))
acf(ARMA, main="ACF",ylim=c(-.5,.5),ci.col="black",ylab="")
pacf(ARMA,main="PACF",ylim=c(-.5,.5),ci.col="black")
```

L'expressió del procés és:  $X_t = 0.8X_t + Z_t - 0.3Z_{t-1}$ .

La figura següent mostra la gràfica del procés i les seves funcions ACF i PACF.

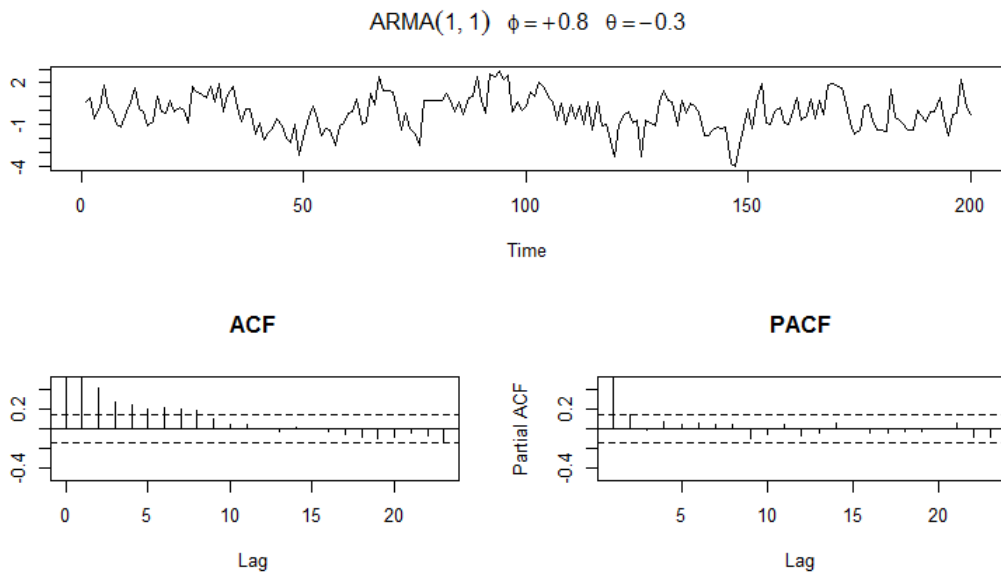


Figura 3.18: Procés ARMA(1,1) amb les seves ACF i PACF

Considerem ara un altre procés ARMA(1,1), amb els coeficients:  $\phi = 0.3$  i  $\theta = -0.8$ . Per tant, la seva expressió és:  $X_t = 0.3X_t + Z_t - 0.8Z_{t-1}$ .

El codi que el genera és:

```
#Simulació d'un procés ARMA(1,1) amb phi=0.3 i theta=-0.8
layout(matrix(c(1,1,2,3), 2, 2, byrow = TRUE))
ARMA=arima.sim(list(order=c(1,0,1), ar=.3, ma=-.8), n=200)
plot(ARMA, ylab=" ", main=(expression(MA(1)~~~phi==+.3~~~theta==-.8)))
```

```
acf(ARMA, main="ACF",ylim=c(-.5,.5),ci.col="black",ylab="")
pacf(ARMA,main="PACF",ylim=c(-.5,.5),ci.col="black")
```

La figura següent mostra la gràfica del procés i les seves funcions ACF i PACF

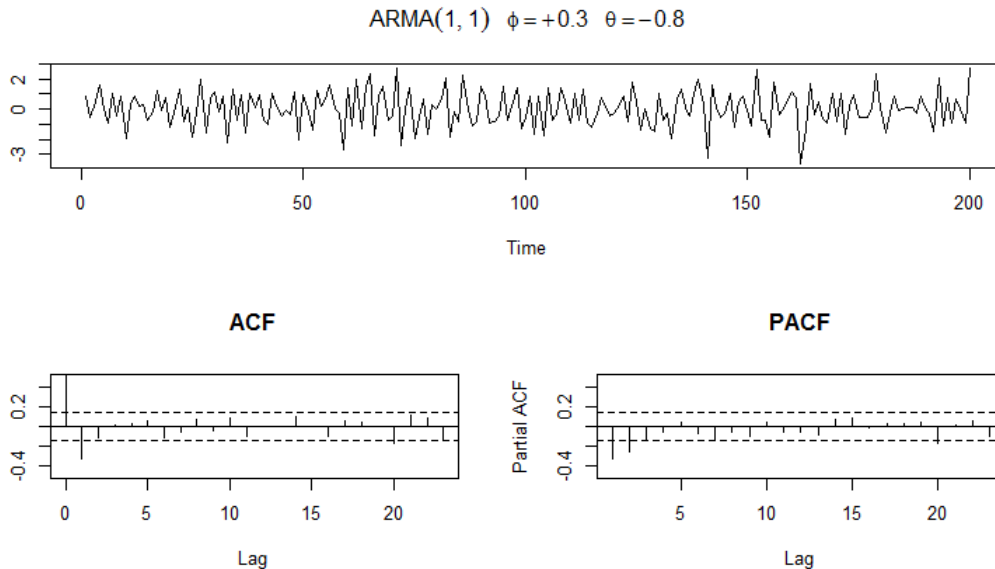


Figura 3.19: Procés ARMA(1,1) amb les seves ACF i PACF

### 3.3.7 Processos ARIMA

Els models de sèries temporals analitzats fins ara es basen en el supòsit d'estacionarietat, és a dir, la mitjana i la variància són constants en el temps i la covariància és invariant en el temps. Però moltes sèries temporals no són estacionàries, per exemple perquè la variància no és constant en el temps. Si un procés no és estacionari però les seves diferències d'ordre  $d$  ho són, el procés s'anomena procés integrat d'ordre  $d$ . Aleshores, a la sèrie temporal se li han d'aplicar diferències  $d$  vegades per fer-la estacionària, i llavors aplicar-li a aquesta sèrie un model ARMA( $p,q$ ). En aquest cas, direm que la sèrie original és ARIMA( $p,d,q$ ), és a dir, una sèrie temporal autoregressiva integrada de mitjanes mòbils.  $p$  indica el nombre de termes autoregressius,  $d$  el nombre de vegades que a la sèrie se li han d'aplicar diferències per fer-la estacionària, i  $q$  el nombre de termes de mitjana mòbil.

Un model ARIMA(0, $d$ ,0) és una sèrie temporal que es converteix en un soroll blanc després de ser diferenciat  $d$  vegades. Aquest model s'expressa mitjançant:

$$(1 - B)^d X_t = Z_t.$$

El model general ARIMA( $p,d,q$ ) té l'expressió:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - B)^d X_t = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q) Z_t$$

o més breument

$$\phi(B)(1 - B)^d X_t = \theta(B) Z_t.$$

Aquesta fórmula és pot aplicar a qualsevol model. Si hi ha alguna component  $p, d, q$  igual a zero, s'elimina el terme corresponent de la fórmula general.

Aquest model s'anomena un model "integrat" perquè el model estacionari diferenciat  $d$  vegades, que s'ajusta a les dades, ha de ser sumat (o "integrat") per proporcionar un model per les dades no estacionàries.

El model per  $X_t$  és clarament no estacionari, ja que l'operador AR,  $\phi(B)(1-B)^d$  té  $d$  arrels dins el cercle unitat [12]. En efecte,  $(1-B)^d$  té  $d$  arrels igual a 1, i que per tant, estan dins el cercle unitat.

### 3.4 Validació del model

Si la sèrie temporal és estacionària, o si s'ha transformat en una sèrie temporal estacionària diferenciant  $d$  vegades, el següent pas és seleccionar el model ARIMA apropiat, que vol dir trobar els valors de  $p$  i  $q$  més adequats per a un ARIMA(p,d,q). Per això, normalment és necessari examinar l'ACF i la PACF de la sèrie temporal estacionària.

Un test per comprovar l'estacionarietat d'una sèrie temporal, que s'utilitza amb sèries prèviament desestacionalitzades, és l'anomenat test de Dickey-Fuller augmentat. La hipòtesi nul·la en aquest test es que s'ha de diferenciar la sèrie perquè sigui estacionària i la hipòtesi alternativa és que no s'ha de diferenciar. Per tant, si el p-valor que obtenim en aquest test supera el nivell crític de significació (0.05), llavors s'haurà de diferenciar la sèrie, és a dir, s'assumeix que no és estacionària. En cas contrari, es rebutja la hipòtesi nul·la, llavors, no s'haurà de diferenciar la sèrie [3].

De vegades, aquest procediment ens donarà més d'un model ARIMA(p,d,q), d'entre els quals haurem de triar el millor.

Per això, existeixen una sèrie de criteris que ens permeten seleccionar el "millor" model. Després de donar les definicions, matisarem una mica què entenem per "millor".

Parlarem aquí només de dos d'aquests criteris, basats en la divergència de Kullback-Leibler, el criteri d'Akaike, AIC, i el criteri d'informació bayesiana, BIC.

Mitjançant el F test es poden provar diversos models d'un contra l'altre. Aquestes tests s'han utilitzat en el passat d'una manera esglaonada, on les variables eren afegides o eliminades quan els valors del F test o bé excedeix o no aconseguix superar alguns nivells predeterminats. El procediment anomenat de regressió múltiple per passos és útil per arribar a un conjunt de variables útils. Una altra alternativa és centrar-se en un procediment per a la selecció del model que no procedeixi de forma seqüencial, sinó que simplement avalua cada model pels seus propis mèrits. Suposem que considerem un model de regressió amb  $k$  coeficients i denotem l'estimador de màxima versemblança de la variància com

$$\hat{\sigma}_k^2 = \frac{SSE_k}{n} \quad (3.7)$$

on  $SSE_k$  indica la suma d'errors al quadrat sota el model amb  $k$  coeficients de regressió. Llavors, Akaike suggereix el mesurament de la bondat d'ajustament per a aquest model particular, equilibrant l'error de l'ajustament contra el nombre de paràmetres en el model; la definició del criteri d'informació d'Akaike és

**Definició 3.4.1** *El Criteri d'Informació d'Akaike, AIC, ve donat per*

$$AIC = \ln \hat{\sigma}_k^2 + \frac{n+2k}{n},$$

on  $\hat{\sigma}_k^2$  ve donat per (3.7) i  $k$  és el nombre de paràmetres en el model.

El valor de  $k$  que produeix el mínim AIC especifica el millor model. La idea és més o menys que la minimització de  $\hat{\sigma}_k^2$  seria un objectiu raonable, excepte que disminueixi monòtonament a mesura que  $k$  augmenta. Per tant, hem de penalitzar la variància de l'error per un terme proporcional al nombre de paràmetres. L'elecció del terme de penalització donada en l'AIC no és l'única, hi ha diferents termes de penalització. Una forma corregida va ser suggerida per Sugiura (1978), i ampliada per Hurvich i Tsai (1989), pot basar-se en petites mostres distribuïdes com a resultat del model de regressió lineal, en la forma següent:

**Definició 3.4.2** L'AIC de biaix corregit,  $AICc$ , ve donat per

$$AICc = \ln \hat{\sigma}_k^2 + \frac{n+k}{n-k-2},$$

on  $\hat{\sigma}_k^2$  ve donat per (3.7),  $k$  és el nombre de paràmetres en el model, i  $n$  és la mida de la mostra.

També podem derivar un terme de correcció basat en arguments bayesians, com en Schwarz(1978), que porta a la definició següent:

**Definició 3.4.3** El Criteri d'Informació Bayesian,  $BIC$ , ve donat per

$$BIC = \ln \hat{\sigma}_k^2 + \frac{k \ln n}{n},$$

on  $\hat{\sigma}_k^2$  ve donat per (3.7),  $k$  és el nombre de paràmetres en el model, i  $n$  és la mida de la mostra.

El BIC també es diu Criteri d'Informació de Schwarz (SIC). Diversos estudis de simulació han tendit a verificar que BIC aconsegueix l'ordre correcte en mostres grans, mentre que AICc tendeix a ser superior en mostres més petites on el nombre relatiu de paràmetres és gran [8].

Tots aquests criteris consten de dos sumands, el primer decreix a mesura que augmentam el nombre de termes, i el segon augmenta amb el nombre de sumands.

El model amb un valor d'AIC més petit és millor. De totes formes, s'ha de tenir present que l'AIC no dona cap informació sobre si un model donat és bo, només indica que és millor que els altres models considerats. I si tots els models candidats encaixen malament, l'AIC no donarà cap avís d'això.

Per la construcció d'un model adequat de sèries temporals, hem de considerar també el *principi de parsimònia*. D'acord amb aquest principi, sempre s'ha de seleccionar el model amb el menor nombre possible de paràmetres per tal de proporcionar una representació adequada de les dades subjacents a la sèrie temporal.

D'altra banda, com més complicat és el model, més possibilitats té de presentar desviació dels supòsits del model real. Amb l'augment dels paràmetres del model, el risc de sobreajust també augmenta. Un model de sèrie temporal més ajustat pot descriure les dades mostrals molt bé, però pot no ser adequat per al pronòstic futur. Com que potenciar el sobreajust afecta la capacitat d'un model per a pronosticar bé, la

parsimònia s'utilitza sovint com un principi rector per superar aquest problema. Així, en resum, es pot dir que, en fer pronòstics de sèries temporals, s'hauria de donar gran atenció per seleccionar el model més parsimoniós entre totes les altres possibilitats [11].

Una vegada s'ha triat el millor model, després s'ha de validar si és adequat per modelitzar la sèrie temporal donada. Concretament, una sèrie estarà ben identificada si quan s'ajusta el model els residus es comporten com un soroll blanc. Recordem que un soroll blanc és una sèrie estacionària en la que cap observació depèn de les altres, i per tant els valors de l'ACF i la PACF són nuls. A més, el gràfic dels residus ha de mostrar que varien al voltant del zero, sense tendència, la variància és constant i no hi ha valors atípics. Aproximadament el 95% dels residus estandaritzats han d'estar entre  $-2$  i  $2$  desviacions típiques [3].

Un test per comprovar si els residus són un soroll blanc és l'anomenat test de Ljung-Box. La hipòtesi nul·la d'aquest test és que les primeres autocorrelacions són nul·les. La hipòtesi alternativa d'aquest test implica que algunes de les correlacions són diferents de zero i, per tant, no es pot assumir que els residus siguin un soroll blanc. Llavors si el p-valor que obtenim en aquest test supera 0.05, s'assumeix que els residus són un soroll blanc [3].

### 3.5 Exemple

La sèrie temporal següent recull el nombre de naixements mensuals a Espanya des de gener del 2000 fins desembre de 2004 [16].

En total es tracta de 60 observacions i estan recopilades a la taula 3.2.

t	$X_t$	t	$X_t$	t	$X_t$	t	$X_t$	t	$X_t$	t	$X_t$
1	32995	11	33317	21	34176	31	35962	41	38593	51	37091
2	30696	12	33780	22	36266	32	35470	42	36100	52	36389
3	33608	13	34198	23	34322	33	36943	43	37849	53	37487
4	32196	14	29881	24	33692	34	37123	44	36814	54	37001
5	34026	15	33424	25	34023	35	35003	45	38655	55	39275
6	31832	16	32653	26	31307	36	36008	46	37978	56	37975
7	33754	17	35466	27	33680	37	36189	47	36117	57	39741
8	33666	18	32706	28	34656	38	33051	48	36738	58	40051
9	33758	19	34788	29	35733	39	36853	49	37603	59	38208
10	34004	20	34808	30	32938	40	36944	50	34515	60	39255

Taula 3.2: Naixements mensuals

El següent codi converteix la taula de dades en una sèrie temporal, i la representam en una gràfica respecte el temps. Es pot observar la representació gràfica en la figura 3.20.

```
Neixer <- read.table("nacimientos_totales.txt")
Neixer.ts <- ts(data=Neixer, start=c(2000,1), end=c(2004,12),
frequency = 12)
```



```
plot(Neixer.ts, type="o", pch=20 ,ylab="Naixaments Espanya",xlab="t")
```

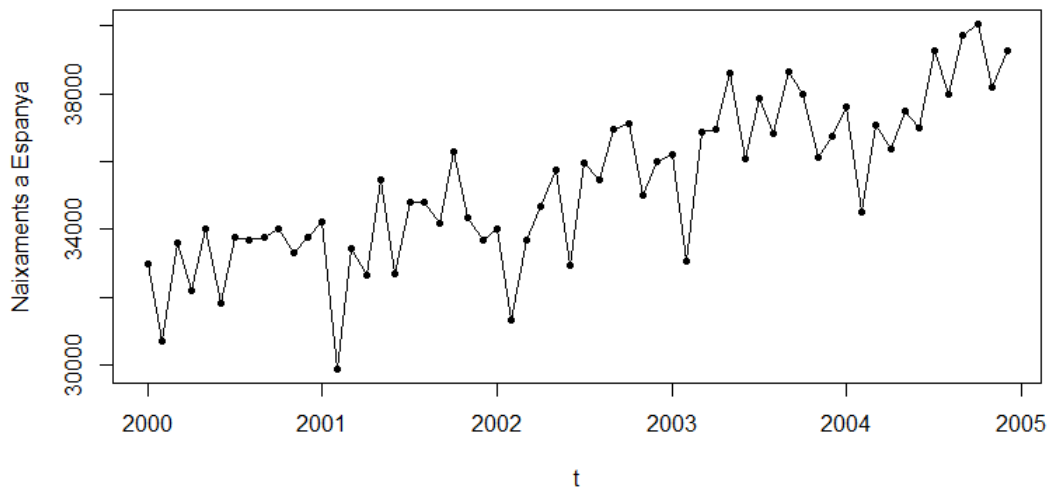


Figura 3.20: Evolució cronològica de les dades.

El primer que s’ha de fer és la descomposició en components de la sèrie. La descomposició de les sèries temporals implica la separació de la sèrie en tres components; la tendència, la component estacional i la component irregular, és a dir, l’estimació d’aquestes tres components. Per estimar aquestes components es pot utilitzar la funció en R: “`decompose(serie)`”. A continuació trobam el codi per fer aquesta descomposició i la posterior representació gràfica.

```
decomp.Neixer <-decompose(Neixer.ts)
str(decomp.Neixer)
plot(decomp.Neixer)
```

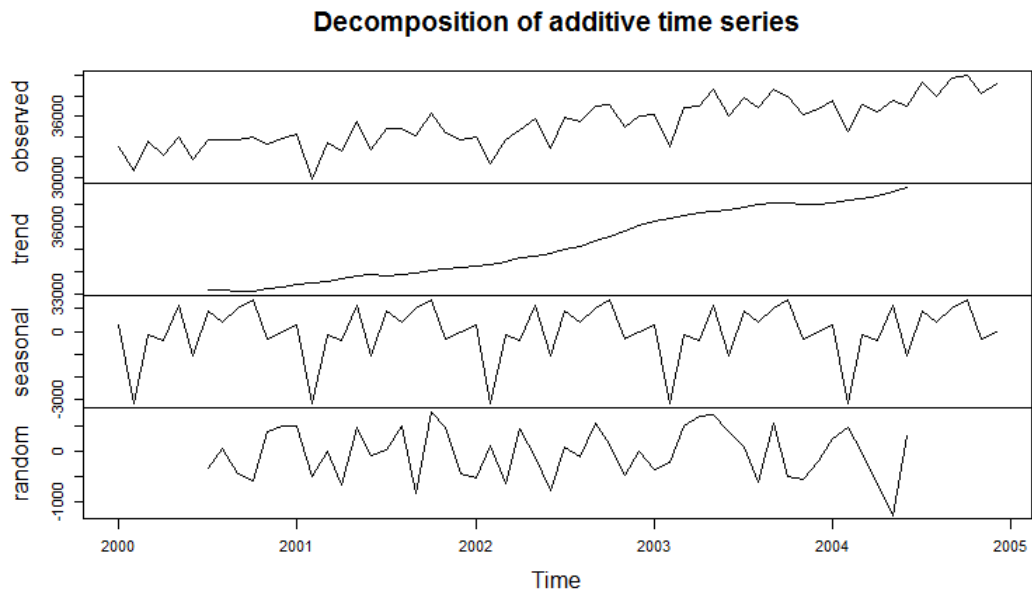


Figura 3.21: Descomposició en components de la sèrie temporal.

Si es té una sèrie temporal estacional que pot ser descrita utilitzant un model additiu, es pot ajustar per l'estacionalitat de la sèrie temporal mitjançant l'estimació de la component estacional, i restant aquesta component a la sèrie temporal original. Es pot fer mitjançant l'estimació de la component estacional calculada per la funció “decompose(serie)”, que correspon a “\$seasonal”. Vegem-ho en el següent codi, on també es fa la representació gràfica de la nova sèrie temporal.

```
Neixer1 <- Neixer.ts-decomp.Neixer$seasonal
plot(Neixer1,ylab="Naixements",xlab="t")
```

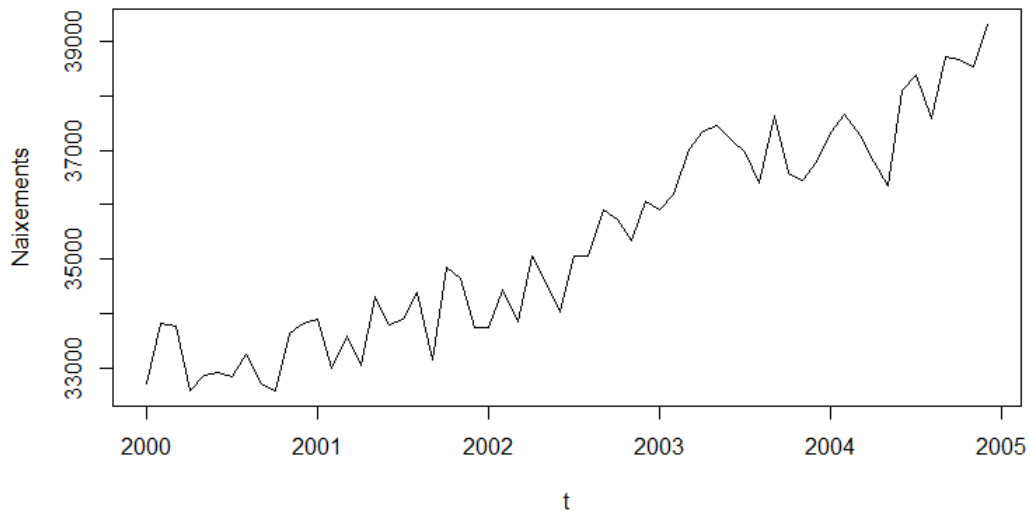


Figura 3.22: Sèrie sense la part estacional.

Els models ARMA es defineixen per a sèries temporal estacionàries. Per tant, si es comença amb una sèrie temporal no estacionària, primer s'haurà de diferenciar la sèrie fins a obtenir-ne una estacionària. La següent instrucció en R del paquet “forecast” indica quantes diferències s’han de fer per fer la sèrie estacionària.

```
require(forecast)
ndiffs(Neixer1)
```

Llavors, amb el codi següent aplicam les diferències a la sèrie: “diff()”, d’acord al que ens ha donat la instrucció anterior. I amb “adf.test()” de la llibreria “tseries” comprovam l’estacionarietat de la sèrie temporal, aquesta instrucció aplica el test de Dickey-Fuller augmentat. Que com ja s’ha comentat si el p-valor és menor que 0.05, indica que la sèrie és estacionària.

```
Neixer2 <- diff(Neixer1, differences = 1)
# test Dickey-Fuller
library(tseries)
adf.test(Neixer2)
```

El test ens dóna:

Augmented Dickey-Fuller Test

```
data: Neixer2
Dickey-Fuller = -6.4699, Lag order = 3, p-value = 0.01
alternative hypothesis: stationary
```

### 3. MODELS PROBABILÍSTICS

---

Per tant, el test de Dickey-Fuller augmentat després d'aplicar una diferència ha donat 0.01, i aleshores, la sèrie és estacionària. A més, hem representat la nova sèrie temporal i és pot veure clarament que és estacionària.

```
plot.ts(Neixer2)
```

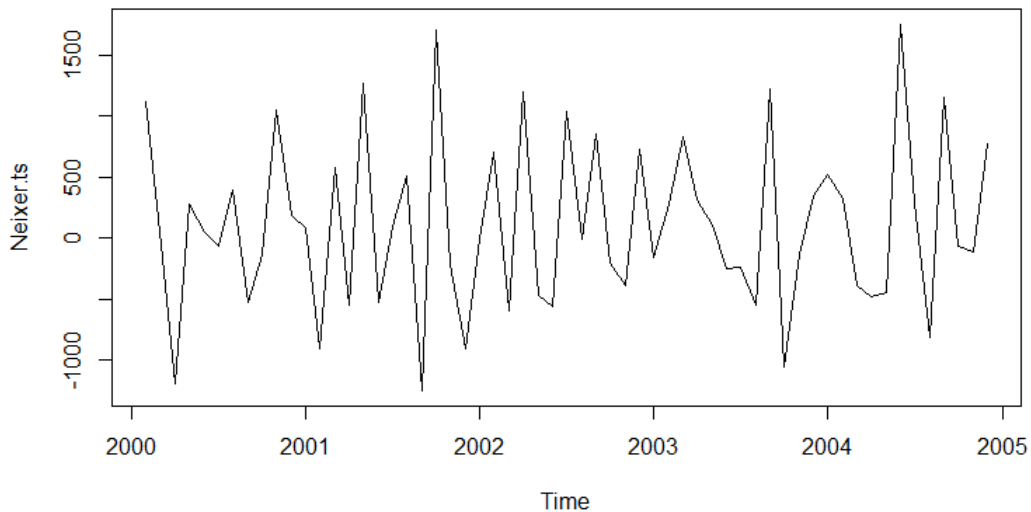


Figura 3.23: Sèrie estacionaritzada.

Tant si la sèrie temporal era estacionària, com si s'ha transformat en una sèrie estacionària diferenciant  $d$  vegades, el següent pas és seleccionar el model ARIMA apropiat, que vol dir trobar els valors  $p$  i  $q$  de ARIMA( $p,d,q$ ) més adequats. Per això és necessari examinar l'ACF i la PACF de la sèrie temporal estacionària. Per representar aquestes funcions podem utilitzar les funcions en R "ACF()" i "PACF()", respectivament.

```
require(astsa)
dev.new()
acfp <- acf(Neixer2,plot = FALSE)
acfp$lag <- acfp$lag *12
plot(acfp,ylab="",main="ACF")
dev.new()
pacfp <- pacf(Neixer2,plot = FALSE)
pacfp$lag <- pacfp$lag*12
plot(pacfp,ylab="",main="PACF")
```

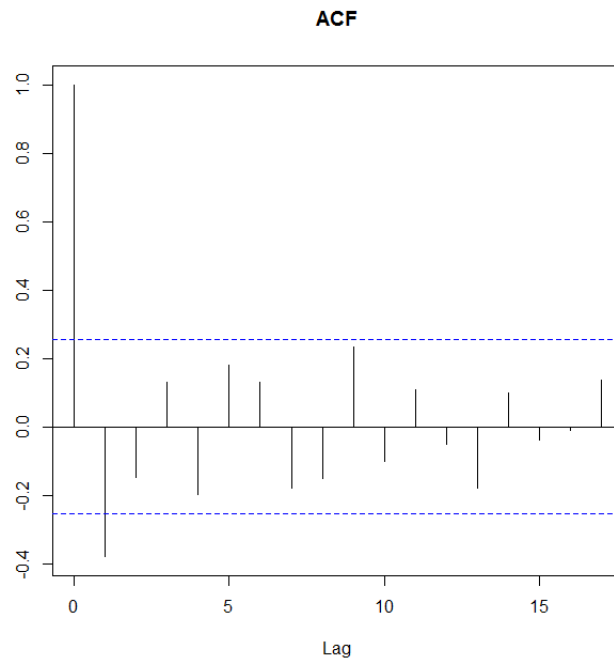


Figura 3.24: ACF de la sèrie Naixements.

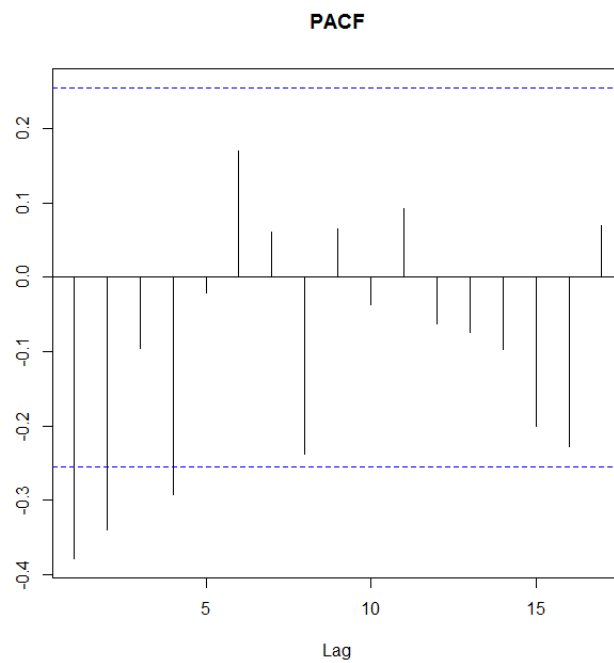


Figura 3.25: PACF de la sèrie Naixements.

Les gràfiques de l'ACF i la PACF mostren un comportament que es pot interpretar

### 3. MODELS PROBABILÍSTICS

---

com un procés MA(1): l'ACF és nul·la després d'un retard i la PACF sembla tenir un comportament sinusoidal. Per tant, el model podria ser un ARMA(0,1); de fet, com que hem fet una diferència per convertir la sèrie original en estacionària, el model elegit serà un ARIMA(0,1,1).

Amb el següent codi estimam els paràmetres del model ARIMA(0,1,1). A més, ens donarà el valor del criteri d'Akaike, AIC.

```
arima.Neixer <- arima(Neixer1,order = c(0,1,1))

> arima.Neixer

Call:
arima(x = Neixer1, order = c(0, 1, 1))

Coefficients:
      ma1
    -0.4844
s.e.    0.1022

sigma^2 estimated as 403187:  log likelihood = -464.61,  aic = 933.22
```

En el codi següent veim que el paràmetre del model ARIMA(0,1,1) és  $\theta = -0.4844$  i el valor de l'AIC és 933.22. Per tant, el model seria:

$$(1 - B)X_t = (1 - 0.4844B)Z_t$$

Però també es pot pensar que es tracta d'un model AR(2), interpretant l'ACF i la PACF com: la primera té un comportament sinusoidal i la segona és nul·la després de dos retards. I, com hem dit abans, com que hem aplicat una diferència, seria ARIMA(2,1,0).

Amb el següent codi estimam els paràmetres del model ARIMA(2,1,0). A més, ens donarà el valor del criteri d'Akaike, AIC.

```
arima.Neixer1 <- arima(Neixer1,order = c(2,1,0))

> arima.Neixer1

Call:
arima(x = Neixer1, order = c(2, 1, 0))

Coefficients:
      ar1      ar2
    -0.4811  -0.3026
s.e.    0.1277   0.1259

sigma^2 estimated as 393644:  log likelihood = -463.94,  aic = 933.88
```

En el codi següent veim que els paràmetres del model ARIMA(0,1,1) són  $\phi_1 = -0.4811$  i  $\phi_2 = -0.3026$  el valor de l'AIC és 933.88. Per tant, el model seria:

$$(1 - 0.4811B - 0.3026B^2)(1 - B)X_t = Z_t.$$

Però s'ha d'elegir el millor model, i com que el valor de l'AIC per el cas de ARIMA(0,1,1) és menor que per ARIMA(2,1,0), ens quedariem amb l'ARIMA(0,1,1).

A més, tenim la instrucció "auto.arima()" que ens diu quin és el millor model per la nostre sèrie temporal. Vegem que ens dona l'opció que hem elegit:

```
library(forecast)
auto.arima(Neixer1,allowdrift=F)

Series: Neixer1
ARIMA(0,1,1)

Coefficients:
      ma1
    -0.4844
s.e.    0.1022

sigma^2 estimated as 410157:  log likelihood=-464.61
AIC=933.22  AICc=933.44  BIC=937.38
```

Ens dona que el millor model és l'ARIMA(0,1,1), com havíem vist amb el valor de l'AIC.

Finalment, utilitzarem el test de Ljung-Box per veure si els residus són un soroll blanc, i per tant, el model s'ajusta correctament a les dades de la sèrie temporal.

Amb el codi següent calculam els residus i aplicam el test de Ljung-Box:

```
# Càlcul dels residus
res<-residuals(arima.Neixer)
# Tests de Ljung-Box:
Box.test(res,lag=20,type = "Ljung")
```

I hem obtingut:

```
Box-Ljung test

data:  res
X-squared = 24.74, df = 20, p-value = 0.2116
```

Llavors, com que surt un p-valor de 0.2116, acceptam que sigui un soroll blanc.

Per acabar de comprovar-ho, vegem que l'ACF i la PACF valen pràcticament zero.

```
# Representació dels residus i les funcions d'autocorrelació
plot(res)
dev.new()
acfp <- acf(res,plot = FALSE)
```

### 3. MODELS PROBABILÍSTICS

---

```
acfp$lag <- acfp$lag *12
plot(acfp,ylab="",main="ACF")
dev.new()
pacfp <- pacf(res,plot = FALSE)
pacfp$lag <- pacfp$lag*12
plot(pacfp,ylab="",main="PACF")
```

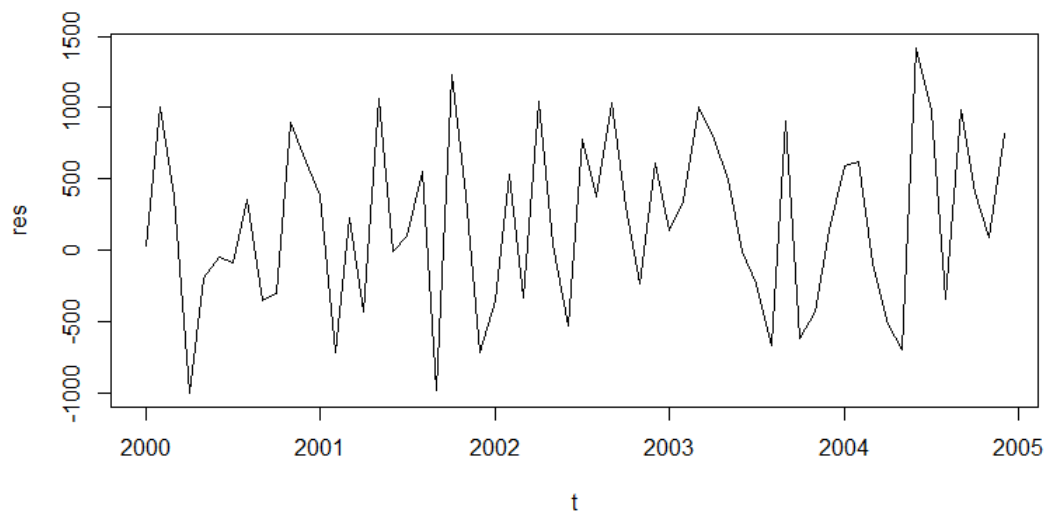


Figura 3.26: Residus.



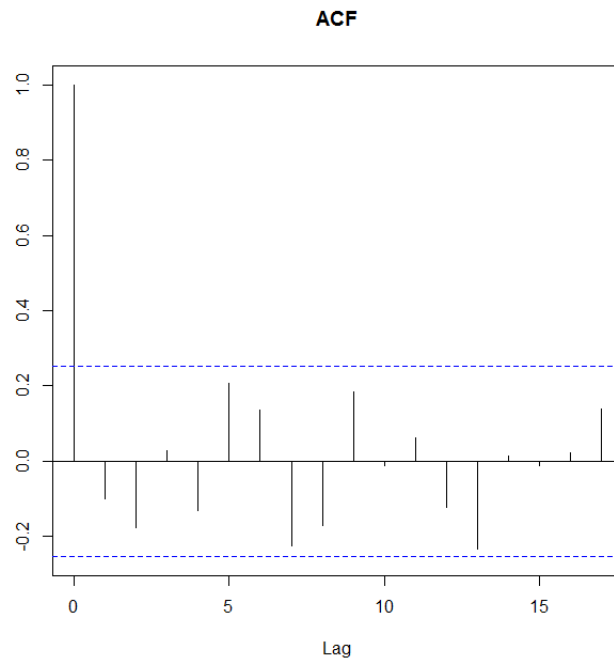


Figura 3.27: ACF dels residus.

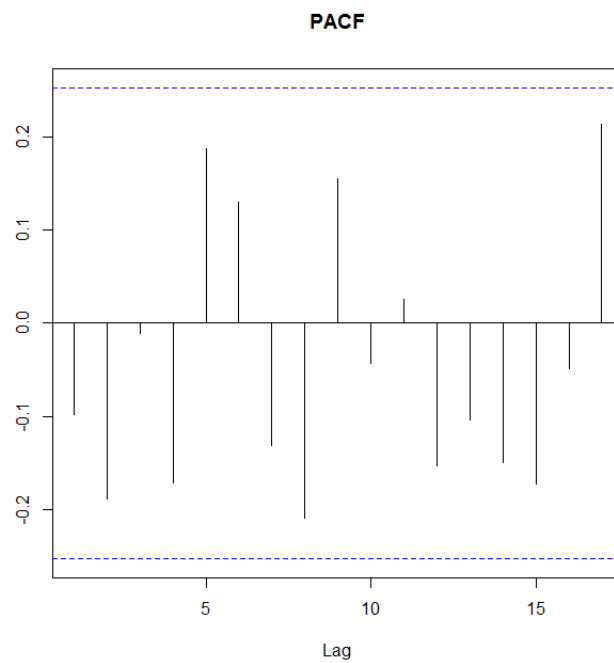


Figura 3.28: PACF dels residus.

### 3.6 Prediccions

Una vegada que s'ha seleccionat el millor model ARIMA(p,d,q) per a les dades de la sèrie temporal, i s'han estimat els paràmetres per aquest model ARIMA, es pot utilitzar com a model per fer prediccions dels valors futurs de la sèrie temporal.

Per fer-ho es pot utilitzar la instrucció en R "forecast.Arima()" del paquet "forecast". Es pot especificar el nivell de confiança per als intervals de predicció en "forecast.Arima()" amb l'argument "level". Per exemple, per obtenir un interval de confiança del 99.5%, s'hauria d'escriure: "forecast.Arima(serie,level=c(95.5))" [14].

Vegem algunes prediccions per l'exemple anterior (naixements mensuals a Espanya). La instrucció següent ens mostra els valors de l'estimació, així com els extrems dels intervals de confiança del 80% i del 95%.

```
#Predicció
library("forecast")
predict <- forecast.Arima(arima.Neixer, h=12)

> predict
      Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
Jan 2005      38914.03 38100.28 39727.78 37669.51 40158.55
Feb 2005      38914.03 37998.50 39829.56 37513.85 40314.21
Mar 2005      38914.03 37906.95 39921.11 37373.84 40454.22
Apr 2005      38914.03 37823.06 40005.00 37245.54 40582.53
May 2005      38914.03 37745.17 40082.89 37126.42 40701.64
Jun 2005      38914.03 37672.16 40155.90 37014.76 40813.30
Jul 2005      38914.03 37603.21 40224.85 36909.31 40918.76
Aug 2005      38914.03 37537.71 40290.35 36809.13 41018.93
Sep 2005      38914.03 37475.19 40352.87 36713.51 41114.55
Oct 2005      38914.03 37415.27 40412.79 36621.88 41206.18
Nov 2005      38914.03 37357.66 40470.40 36533.77 41294.29
Dec 2005      38914.03 37302.11 40525.95 36448.81 41379.26
```

Finalment, representem les prediccions.

```
plot.forecast(predict)
```

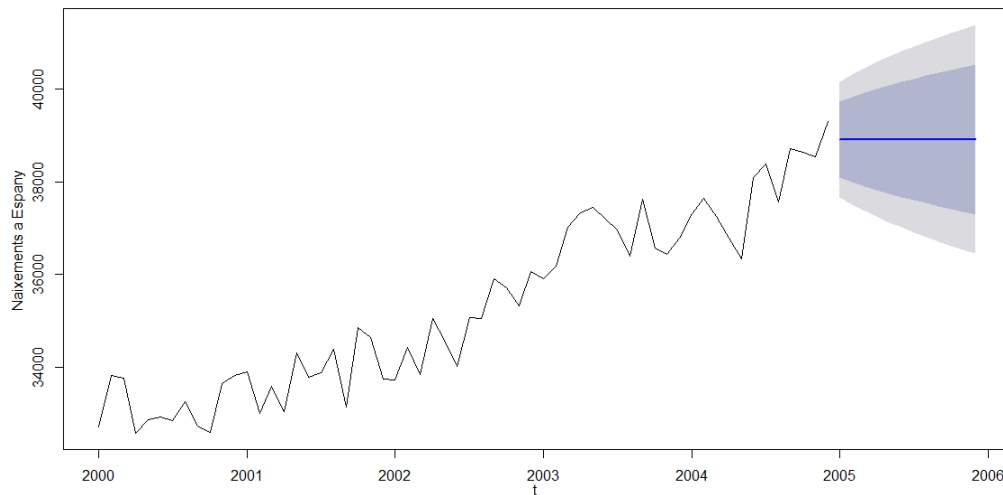


Figura 3.29: Prediccions dels naixements mensuals a Espanya al 2015.

### 3.7 Altres models

En aquest capítol s'han estudiat els models probabilístics per a sèries temporals estacionàries o que la no estacionarietat és deguda a la presència d'una tendència. En aquests casos, com hem vist, podem convertir la sèrie temporal en estacionària aplicant una o varies diferències a la sèrie, i després l'ajustam per un model ARIMA convenient. Quan la no estacionarietat és deguda a altres causes, hi ha altres models probabilístics adequats.

Així, si la sèrie presenta estacionalitat, es pot modelitzar mitjançant un procés ARIMA mixt, indicat  $ARIMA(p, d, q)(P, D, Q)_m$  on  $(p, d, q)$  són com abans i  $(P, D, Q)$  corresponen a la part estacional de la que  $m$  indica el període.

Per exemple, un model  $ARIMA(1, 1, 1)(1, 1, 1)_4$  té l'expressió

$$(1 - \phi_1 B)(1 - \Phi_1 B^4)(1 - B)(1 - B^4)X_t = (1 + \theta_1 B)(1 + \Theta_1 B^4)Z_t$$

on  $(1 - \phi_1 B)$ ,  $(1 - \Phi_1 B^4)$ ,  $(1 - B)$ ,  $(1 - B^4)$ ,  $(1 + \theta_1 B)$  i  $(1 + \Theta_1 B^4)$  corresponen a AR(1) no estacional, AR(1) estacional, diferència no estacional, diferència estacional, MA(1) no estacional i MA(1) estacional; respectivament.

En el cas d'una sèrie temporal de tipus autoregressiu que presenti heterocedasticitat, un model adequat seria ARCH, o GARCH (ARCH generalitzat) introduït per Engle (1982) i posteriorment generalitzat per Bollerslev (1986).



## CONCLUSIONS

Amb aquest treball s'ha volgut fer una ampliació de l'assignatura d'Inferència estadística, aprofundint en l'anàlisi de sèries temporals (tema que no s'ha donat en la carrera). Tot i això, el treball té gran part de probabilitat.

Les sèries temporals s'usen per estudiar la relació causal entre diverses variables que canvien amb el temps i que s'influeixen entre sí. Des del punt de vista probabilístic, una sèrie temporal és una seqüència de valors observats d'una variable ordenats en el temps.

L'estudi de les sèries temporals té dos objectius principals; el primer és el coneixement del seu patró de comportament, és a dir, determinar una funció de temps que modelitzi la sèrie a partir de les dades observades. El segon objectiu és, a partir del model obtingut, predir el comportament futur de la sèrie.

En la primera part hem estudiat com fer la descomposició de la sèrie temporal en components. En el cas de la tendència, hem estudiat el procés de les mitjanes mòbils, tot i que hem notat que són freqüents les situacions en què és difícil modelitzar la tendència a través d'un simple model polinòmic de menor o major grau. Per això s'exposen altres mètodes de modelitzar la tendència, com és l'allisat exponencial. Tot i que no s'ha explicat en el treball, s'ha comentat que hi ha tres tipus d'allisat exponencial: el simple, el de Holt i el de Holt-Winters. S'aplica un allisat exponencial o l'altre depenent si la sèrie té tendència i/o estacionalitat.

A més, s'ha explicat com estudiar l'estacionalitat a través de les mitjanes mòbils. S'ha donat un procés a seguir per calcular els índexs estacionals, aquest procés varia si el model és additiu o multiplicatiu. Finalment, tant per al càlcul de les mitjanes mòbils com dels índexs estacionals es necessita tenir el període de la sèrie. Per això s'ha d'utilitzar el correlograma, explicat en el darrer apartat d'aquest capítol.

A la segona part del treball hem estudiat els models ARIMA només per al cas de sèries sense estructura estacional. Hem vist com identificar el model ARIMA més adequat, com estimar els seus paràmetres i com validar el model. Al final del treball, hem inclòs un exemple complet. En cas que la sèrie temporal tengui estructura estacional, hi ha altres models per tal d'ajustar la sèrie, els models ARIMA(p,d,q)(P,D,Q). Un model

#### 4. CONCLUSIONS

---

ARIMA estacional es forma mitjançant la inclusió de termes estacionals en els models ARIMA vists en el treball. Els paràmetres  $(p,d,q)$  són de la part no estacional del model i els paràmetres  $(P,D,Q)$  són de la part estacional.

La part estacional del model consisteix en termes que són molt similars als components no estacionals del model, però implica un retard del període estacional. Els termes estacionals addicionals simplement es multipliquen amb els termes no estacionals.

## APÈNDIX

En aquest treball hem utilitzat R, un llenguatge i entorn de programació per a l'anàlisi estadístic i gràfic, per implementar les rutines necessàries per calcular el model que cerquem. A continuació s'expliquen les rutines que s'utilitzen en el segon capítol, que correspon a la modelització per components. A més en aquest apèndix inclourem els codis corresponents.

- **MM(dades,p)**. Aquesta rutina genera les mitjanes mòbils de la sèrie temporal *dades*, on *p* és el període de la sèrie temporal, que es determina amb el correlograma de la sèrie.

Es necessari definir la funció **parell(p)**, ja que com en vist en la teoria, es calculen de manera diferent les mitjanes mòbils si el període és parell o senar. Aquesta funció retorna un booleà, que és vertader si *p* és parell i fals si *p* és senar.

```
MM <- function(dades,p){
  l <- length(dades)
  llista <- c()
  for (i in 1:(l-p+1)) {
    aux <- 0
    for (j in 1:p) {
      aux <- aux+dades[i+j-1]
    }
    llista <- c(llibra,aux/p)
  }
  if (parell(p)){
    llista1 <- c()
    for (i in 1:length(llibra)-1) {
      llista1 <- c(llibra1,(llibra[i]+llibra[i+1])/2)
    }
    return(llibra1)
  }
}
```

```
    }
    return(llista)
}

parell <- function(p){
  if (p%%2==0){
    return (TRUE)
  }else{
    return (FALSE)
  }
}
```

- **IE(dades,MitMov,p,Additiu)**. Aquesta rutina calcula els índexs estacionals de la sèrie *dades*. Per fer-ho es necessita passar com a paràmetres: les mitjanes mòbils, el període *p* i un valor booleà *Additiu* per indicar si el model que farem servir és additiu (TRUE) o multiplicatiu (FALSE). Com abans, necessitam la funció **parell**.

```
IE <- function(dades,MitMov,p,Additiu){
  llista <- c()
  if(parell(p)){
    i0 <- p/2
  }
  if(Additiu){
    for (i in 1:length(MitMov)) {
      llista <- c(llista,dades[i0+i]-MitMov[i])
    }
  }else{
    for (i in 1:length(MitMov)) {
      llista <- c(llista,dades[i0+i]/MitMov[i])
    }
  }
  l <- length(llista)
  indexEst <- c()
  for (i in 0:(p-1)) {
    aux <- 0
    noTrobat <- 1
    cont <- 0
    while (noTrobat) {
      if ((i0+cont)%p == i){
        noTrobat <- 0
      }else{
        cont <- cont+1
      }
    }
  }
  ns <- 0
  if (cont == 0){
```



---

```

    cont <- p
  }
  while (cont+p*ns<=l) {
    aux <- aux+llista[cont+p*ns]
    ns <- ns+1
  }
  aux <- aux/ns
  indexEst <- c(indexEst,aux)
}
mitjana <- sum(indexEst)/p
if (Additiu){
  for (i in 1:length(indexEst)) {
    indexEst[i] <- indexEst[i]-mitjana
  }
}else{
  for (i in 1:length(indexEst)) {
    indexEst[i] <- (indexEst[i]/mitjana)*100
  }
}
aux <- indexEst[1]
for (i in 1:(length(indexEst)-1)) {
  indexEst[i] <- indexEst[i+1]
}
indexEst[length(indexEst)] <- aux
return(indexEst)
}

```

- Atres funcions:

- La rutina **f(m,x)** crea la funció tendència, és a dir, amb els mateixos valors de temps que les mitjanes mòbils  $m$  i donant  $x$ , els coeficients de la funció tendència, retornarà el valor de la tendència per unitat de temps.

```

f <- function(m,x){
  k <- c()
  for (i in 1:length(m)) {
    sum <- 0
    for (j in 1:length(x)) {
      sum <- sum+x[j]*i^(j-1)
    }
    k<-c(k,sum)
  }
  return(k)
}

```

- La rutina **x(m,t)** recalcula els valors de temps, ja que no tenim la mateixa quantitat d'observacions de la sèrie i de les mitjanes mòbils, llavors alhora de fer les representacions gràfiques necessitam desplaçar les mitjanes

mòbils  $p/2$ . Els paràmetres d'entrada corresponen a  $m$ , mitjanes mòbils, i  $t$  observacions de la sèrie. Aquesta funció només és necessària per fer algunes de les representacions gràfiques.

```
x <- function(m,t){
  z <- length(t)-length(m)
  t <-z/2
  x <- c()
  for (i in 1:length(m)) {
    x <- c(x,t+i)
  }
  return(x)
}
```

- La rutina **mult(x1,x2,y)** retorna tots els nombres múltiples de  $y$  que es troben entre els nombres  $x1$  i  $x2$ .

```
mult <- function(x1,x2,y){
  t <- c()
  for (i in x1:(x2-1)) {
    if(i%%y==0){
      t<-c(t,i)
    }
  }
  return(t)
}
```

- La rutina **Y(mult,s,E,Additiu)** calcula els valors de  $\tilde{X}_t = X + E_s$  o  $\tilde{X}_t = X \frac{E_s}{100}$  amb  $s = 1, \dots, p$  i  $t = s + \dot{p}$ , on  $X$  és la modelització de la tendència i  $E_s$  els índexs estacionals. Per això, s'han de passar com a paràmetres: els múltiples de  $p$   $mult$ , el període  $p$ , els índexs estacionals  $E$  i un booleà que indica si és additiu o multiplicatiu.

```
Y <- function(mult,p,E,Additiu){
  x <- 0
  f <- 0
  Y <- c()
  for (i in 1:length(mult)) {
    for (j in 1:p) {
      x <- mult[i]+j
      f <- (1.841e+01-1.095e-01*x+2.100e-03*x^2-1.085e-05*x^3)
      if (Additiu){
        Y <- c(Y,f+E[j])
      }else{
        Y <- c(Y,f*E[j]/100)
      }
    }
  }
}
```

---

```

    }
    return(Y)
  }

```

Finalment, el codi utilitzat en l'exemple de les temperatures mensuals a Palma de Mallorca, utilitzant les rutines descrites anteriorment, és:

```

# Dades
Temp <- read.table("TablaTEMP.txt")
Temp.ts <- ts(data=Temp$V2, start=min(Temp$V1), end=max(Temp$V1))
plot(Temp.ts, type="o", pch=20 ,ylab="Temperatures",xlab="t")

# Correlograma
acf(Temp.ts, main="ACF",ylim=c(-1,1),ci.col="black",ylab="")

# Mitjanes mobils
m <- MM(Temp.ts,12)

# Tendència (cubica)
t2 <- time(m)^2
t3 <- time(m)^3
lm.temp <- lm(m~time(m)+t2+t3)
summary(lm.temp)

X2<-x(m,Temp.ts)
fun <- f(m,c(1.845e+01,-1.151e-01,2.265e-03,-1.210e-05))
plot(Temp.ts, type="o", pch=20 ,ylab="Temperatures",xlab="t",
ylim=c(5,30),lwd=1.5)
lines(X2,m, type="o", pch=20 ,ylab="",xlab="", col="blue",
ylim=c(6,27),lwd=1.5)
lines(X2,fun, col="red",lwd=2)
legend(84,29,legend = c("Dades reals", "Mitjanes Mòbils", "Tendència"),
col = c("black","blue","red"),lwd = c(2.5,2.5,2.5),cex = 0.6,
xjust = 0, yjust = 0.6, x.intersp = 0.5, y.intersp = 0.5)

# Estacionalitat
E <-IE(Temp.ts,m,12,TRUE)
plot(E, type = "o", pch=19, ylab = "E", xlab = "s")

# Comprovacio
mult1 <- mult(0,108,12)
Ytilde <- Y(mult1,12,E,TRUE)
plot(Ytilde, type = "l", col="red", ylab = "Temperatures", xlab = "t",
ylim = c(5,30),lwd=1.5)
points(Temp.ts, pch=20,lwd=1.5)
legend(85,29,legend = c("Dades reals","Model"), col = c("black","red"),
lwd = c(2.5,2.5),cex = 0.6,xjust = 0, yjust = 0.6, x.intersp = 0.5,

```

```
y.intersp = 0.5)

# Residus
plot((Temp.ts-Ytilde), type="o",pch=19,ylab="R",xlab="t")

# Predicció
Temp2 <- read.table("TablaTEMP220216.txt")
Temp2.ts <- ts(data=Temp2$V2, start=min(Temp2$V1), end=max(Temp2$V1))
mult2 <- mult(108,120,12)
Ytilde2 <- Y(mult2,12,E,TRUE)
x <- c(109:120)
plot(Ytilde, type = "l", col="red", ylab = "Temperatures", xlab = "t",
ylim = c(5,30),xlim = c(0,120),lwd=1.5)
lines(x,Ytilde2,col="blue",lwd=1.5)
points(Temp2.ts, pch=20,lwd=1.5)
legend(85,29,legend = c("Dades reals","Model","Previsions"),
col = c("black","red","blue"),lwd = c(2.5,2.5),cex = 0.6,xjust = 0,
yjust = 0.6, x.intersp = 0.5, y.intersp = 0.5)
```

## BIBLIOGRAFIA

- [1] J. Villavicencio. (04 Juliol 2016) Introducción a Series de Tiempo. [Online]. Available: [http://www.estadisticas.pr/iepr/LinkClick.aspx?fileticket=4\\_BxecUaZmg%3D&tabid=100](http://www.estadisticas.pr/iepr/LinkClick.aspx?fileticket=4_BxecUaZmg%3D&tabid=100) 1, 2.0.1, 3, 3.2, 3.3.1, 3.3.3, 3.3.5, 3.3.6
- [2] G. Kirchgässner, J. Wolters, and U. Hassler, *Introduction to Modern Time Series Analysis*, 2nd ed. Springer, 2013. 1
- [3] L. Cayuela and A. Justel, *Series temporales en R*, 2012. 2.1, 3.4, 3.4
- [4] (22 Febrer 2016) Clima en Palma de Mallorca/Son San Juan. [Online]. Available: [http://www.tutiempo.net/clima/Palma\\_De\\_Mallorca\\_Son\\_San\\_Juan/83060.htm](http://www.tutiempo.net/clima/Palma_De_Mallorca_Son_San_Juan/83060.htm) 2.1
- [5] M. Pepió Viñals, *Series Temporales*. Edicions UPC, 2001. 2.3.1, 3
- [6] M. S. Bartlett, *On the Theoretical Specification and Sampling Properties of Autocorrelated Time-Series*. Royal Statistical Society. 2.3.1
- [7] A. O. D., *Time Series Analysis and Forecasting*. Batlerworth & Lo Publishers Limited, 1975. 2.3.1
- [8] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications*, 3rd ed. Springer, 2006. 3, 3.1.1, 3.2.2, 3.2, 3.2, 3.3.5, 3.3.5, 3.3.5, 3.4
- [9] A. Aguirre Jaime, *Introducción al tratamiento de series temporales*. Ediciones Diaz de Santos, S.A., 1994. 3
- [10] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, 1970. 3.1.1, 3.3.3, 3.3.4
- [11] R. Adhikari and R. K. Agrawal, *An Introductory Study on Time Series Modeling and Forecasting*. 3.1.1, 3.4
- [12] C. Chatfield, *The Analysis of Time Series. An introduction*, 5th ed. Chapman & Hall, 1996. 3.2, 3.3.1, 3.3.3, 3.3.4, 3.3.7
- [13] T. Villagarcía, *Series Temporales*. 3.2, 3.3.5
- [14] A. Coghlan. (04 Juliol 2016) A Little Book of R For Time Series. [Online]. Available: <https://media.readthedocs.org/pdf/a-little-book-of-r-for-time-series/latest/a-little-book-of-r-for-time-series.pdf> 3.2, 3.6

## BIBLIOGRAFIA

---

- [15] S. de la Fuente Fernández. (06 Juliol 2016) Series Temporales: Modelo ARIMA. [Online]. Available: <http://www.estadistica.net/ECONOMETRIA/SERIES-TEMPORALES/modelo-arima.pdf> 3.3.5
- [16] M. González Velasco and I. M. del Puerto García, *Series temporales*. Universidad de Extremadura. 3.5