

---

# Nous models per a sèries temporals

---

David MORIÑA SOLER

Tesi doctoral presentada al Departament de  
Matemàtiques de la Universitat Autònoma de Barcelona

*Director:* Pere PUIG, PhD.

**UAB**  
Universitat Autònoma de Barcelona

CERTIFICO que aquesta memòria ha estat realitzada per David Moríña Soler sota la direcció del Dr. Pere Puig Casado

Bellaterra, Juliol de 2013.

Dr. Pere Puig Casado.

---

A Queralt, la teva aparició m'ha il·luminat i m'ha donat forces quan més falta feia

A Laia, per la teva paciència durant aquest llarg procés que ara (per fi) acaba

Als meus pares, que han cregut i m'han fet creure



---

# ÍNDEX

<b>Índex de taules</b>	<b>v</b>
<b>Índex de figures</b>	<b>vii</b>
<b>Agraïments</b>	<b>ix</b>
<b>Introducció</b>	<b>xiii</b>
<b>0 Eines matemàtiques</b>	<b>1</b>
0.1 Mètode delta . . . . .	1
0.2 Equacions en diferències . . . . .	6
0.3 Equacions funcionals . . . . .	10
0.4 Contrastos de bondat de l'ajust basats en la funció de distribució empírica . . . . .	13
0.4.1 Distribució asimptòtica de $W_n^2$ . . . . .	14
0.5 Distribucions autodescomposables . . . . .	15

<b>1</b>	<b>Anàlisi de sèries temporals</b>	<b>17</b>
1.1	Introducció . . . . .	17
1.2	Sèries temporals discretes . . . . .	20
1.2.1	Models INAR . . . . .	20
1.2.2	Models INMA(q) . . . . .	23
1.2.3	Models INARMA(1,q) . . . . .	26
1.2.4	Altres models per a sèries temporals discretes . . . . .	27
1.3	Anàlisi espectral de sèries temporals . . . . .	27
<b>2</b>	<b>Aplicació als ingressos hospitalaris per malalties estacionals</b>	<b>31</b>
2.1	Introducció . . . . .	32
2.2	Definició del model . . . . .	35
2.3	Estimació dels paràmetres i validació del model . . . . .	38
2.3.1	Validació del model . . . . .	41
2.4	Prediccions . . . . .	43
2.4.1	Mitjanes a llarg termini . . . . .	44
2.4.2	Estimació de la variància . . . . .	46
2.4.3	Regions de predicció a curt termini . . . . .	48
2.4.4	Regions de predicció a llarg termini . . . . .	52
2.5	Conclusions . . . . .	55
<b>3</b>	<b>Models autoregressius de primer ordre amb innovacions no gaussianes</b>	<b>59</b>
3.1	Introducció . . . . .	59
3.2	La funció d'autocorrelació de l'exponencial d'un AR(1) . . . . .	61
3.3	Caracterització de la distribució de les innovacions . . . . .	68
3.4	Un contrast de bondat de l'ajust per a les innovacions . . . . .	73
3.5	Dos exemples . . . . .	78
3.5.1	Captures de peix a l'Atlàntic i al golf del riu St. Lawrence	78

3.5.2	Deflactor del producte interior brut (PIB) espanyol . . . .	80
3.6	Estudi de la potència del contrast . . . . .	83
3.7	Conclusions . . . . .	84
<b>Bibliografia</b>		<b>87</b>





---

# ÍNDEX DE TAULES

2.1	Estimadors de màxima versemblança . . . . .	40
2.2	Mitjanes a llarg termini estimades per un any (48 setmanes) . . .	46
3.1	Distribucions marginals de $X_t$ amb la funció generatriu de moments de la distribució marginal de $Y_t$ , $\psi_Y(z)$ ( $Y = \log(X)$ ), i la ACF estandarditzada $\rho_X(s)$ ( $s = a^k$ ) . . . . .	65
3.2	p-valors corresponents a l'exemple de les captures de peix . . . . .	79
3.3	p-valors corresponents a l'exemple del PIB espanyol . . . . .	82
3.4	Anàlisi de la potència per innovacions uniformes, Laplace, logística i gamma . . . . .	84



---

# ÍNDEX DE FIGURES

1.1	Dades de Fürth . . . . .	22
1.2	Funcions d'autocorrelació i autocorrelació parcial per a les dades de Fürth . . . . .	23
1.3	Sèrie original de Fürth i simulació basada en un model INAR(2) . . . . .	24
2.1	Arribades al Servei d'Urgències amb símptomes de grip. Dades observades, model trigonomètric ajustat i mitjanes estimades en base al model. . . . .	37
2.2	Funcions d'autocorrelació dels residus corresponents als models d'ordre 1 i 2 . . . . .	42
2.3	Funció d'autocorrelació parcial dels residus corresponents als models d'ordre 1 i 2 . . . . .	43
2.4	Arribades setmanals mitjanes a llarg termini estimades per un any . . . . .	48
2.5	Regió de predicció per 2009 . . . . .	52
2.6	Regió de predicció per 2008 i valors observats . . . . .	53
2.7	Regions de predicció a curt i llarg termini . . . . .	54

2.8	Regió de predicció per 2009 i valors observats . . . . .	57
3.1	Funcions de densitat de $Y$ en el cas d'innovacions uniformes . . . .	67
3.2	Funcions d'autocorrelació de $X$ en el cas d'innovacions uniformes	68
3.3	$\rho_X(s)$ per algunes marginals estandarditzades . . . . .	69
3.4	Funcions de densitat de les innovacions en el cas $\rho_X(s) = as^2 + bs + c$	72
3.5	Funció d'autocorrelació parcial (captures de peix) . . . . .	79
3.6	Comparació de les distribucions espectrals mostral i teòrica (cap- tures de peix) . . . . .	80
3.7	Funció d'autocorrelació parcial (PIB espanyol) . . . . .	81
3.8	Comparació de les distribucions espectrals mostral i teòrica (PIB espanyol) . . . . .	82

---

# AGRAÏMENTS

El treball que acaba amb aquesta memòria va començar, com és habitual, per una successió de casualitats. En finalitzar el Màster en Matemàtica Avançada a la Universitat Autònoma de Barcelona, enfocat cap a l'Àlgebra no commutativa, em vaig trobar treballant com a estadístic amb gent molt propera a l'Hospital Clínic de Barcelona. A ells he d'agrair-los haver decidit emprendre aquesta aventura, especialment a José Ríos i Ferran Torres, que des del primer dia em van acollir com un més, i em van permetre assistir a les interessants trobades periòdiques que organitzen amb el seu grup, on va sorgir la qüestió primigènia que va motivar aquest treball. Allà van arribar unes dades d'ingressos hospitalaris per grip, que em van fer pensar en les tècniques de sèries temporals de recomptes que havia introduït a les classes del Màster el professor Pere Puig, a qui he d'agrair la infinita paciència que ha tingut en assumir la tasca de dirigir-me en tot aquest procés, i amb qui tindrè un deute perpetu per haver accedit a compartir amb mi una

petita part dels seus coneixements. Ni tan sols un greu problema de salut patit en la fase final de redacció d'aquesta memòria ha pogut reduir la seva dedicació ni la seva increïble capacitat de treball.

El Grup de Recerca *Advanced Statistical Modelling* m'ha permès participar de les seves interessants activitats i seminaris, que han estat sense dubte enormement enriquidors, així com formar part des de la Unitat de Bioestadística del Centre Tecnològic de Nutrició i Salut de Reus d'un dels nodes de la xarxa *BioStatNet*, que durant els seus pocs anys d'existència ha facilitat, entre altres coses, la participació de joves investigadors en congressos internacionals, experiència que ha resultat també molt enriquidora.

Al professor Albert Navarro, de la Unitat de Bioestadística de la Universitat Autònoma de Barcelona, li he d'agrair l'oportunitat de descobrir la vocació de la docència que em va donar ja fa quatre anys, tot just acabada la Diplomatura d'Estadística, i les impagables hores de teràpia que hem compartit a la cafeteria de la Facultat de Medicina. Precisament, des de la Unitat de Bioestadística, i més concretament des del *Grup de Recerca d'Amèrica i Àfrica Llatines (GRAAL)* també se m'ha facilitat la participació en activitats ben engrescadores com cursos i congressos internacionals.

Al llarg de tot aquest temps, hem tingut la sort de comptar amb el suport del professor Jordi Valero, de la Universitat Politècnica de Catalunya, recolzament

que ha estat crucial en la part final del projecte, que forma el gruix del tercer capítol d'aquesta memòria, part en la qual el professor Valero ens ha donat un gran cop de mà.

També he d'agrair al professor Manuel Scotto, de la Universidade de Aveiro, el fet d'haver-me convidat a passar uns dies a la seva institució, on vaig ser tractat de forma immillorable, vaig tenir ocasió de presentar un seminari i assistir-ne a un altre ofert pel professor Brendan McCabe de la University of Liverpool, i em va proporcionar l'oportunitat d'endegar una col·laboració amb tots dos que de ben segur donarà els seus fruits.

Aquest treball mai no hagués arribat a bon port sense el suport dels amics que m'han acompanyat tot aquest temps, en especial Carlos Giraldo, amb qui vaig compartir bona part del Màster, i vam acabar compartint molt més que això, Sara Fontdecaba, Isabel Serra i tota la gent fantàstica que he anat coneixent durant aquesta aventura.

I evidentment, sense la paciència i el suport de la meva família, aquest treball hagués quedat probablement com a projecte pendent indefinidament, però la confiança (fregant els límits de la inconsciència) dels meus pares i la meva dona em van donar la força necessària per emprendre aquest projecte, i en darrer terme, per acabar-lo amb aquesta memòria.





---

# INTRODUCCIÓ

A l'octubre de 2009, just acabada la Diplomatura d'Estadística i el Màster en Matemàtica Avançada, em vaig trobar col·laborant com a estadístic amb un grup molt proper a l'Hospital Clínic de Barcelona. Allà s'organitzaven setmanalment unes trobades entre alguns estadístics dels diversos grups de l'Hospital per parlar dels projectes en els quals cadascú participava, i en definitiva, col·laborar i compartir els coneixements. En aquest marc, es van presentar unes dades relatives a les arribades al Servei d'Urgències de l'Hospital diagnosticades com a grip, que els responsables havien proposat d'analitzar a través d'una regressió de Poisson com a primera aproximació.

L'estructura de les dades (recomptes setmanals del número d'ingressos) em va fer pensar en les tècniques de sèries temporals discretes que el professor Pere Puig havia introduït a les seves classes del Màster. Per aquest motiu, vaig anar a parlar amb ell sobre la qüestió, per veure si podia assessorar-me sobre la millor

manera d'abordar l'anàlisi d'aquelles dades. El seu entusiasme des del primer moment va fer-me decidir a aprofundir una mica més en aquest projecte, de manera que de seguida vam posar-nos a treballar i vam arribar a la conclusió que el fenomen podia ser estudiat a través d'una aproximació basada en els models de sèries temporals discretes de tipus INAR, definits al capítol següent, un cop introduïda d'alguna manera l'estacionalitat clara que presentaven les dades en aquests models. Aquest treball va donar lloc als resultats que s'introdueixen al capítol 2, publicats com a article de recerca a la revista *Statistics in Medicine*, i presentats en forma de comunicació oral als congressos *ERCIM* de 2012 a Oviedo, a l'*IWSM* de 2010 a Glasgow i a la primera *Jornada de joves investigadors en Matemàtiques* de 2010 a Barcelona.

La segona qüestió que s'aborda en aquest treball, presentada en el tercer capítol, va sorgir com un problema de caracterització de distribucions, en un context de models autoregressius de primer ordre, motivada per un resultat sorprenent de McKenzie presentat a [31].

En aquest treball es prova que donat un procés  $Y_t$  amb estructura  $AR(1)$ , i considerant la sèrie exponenciada  $X_t = e^{Y_t}$ , la funció d'autocorrelació de  $X_t$  és la mateixa que la de la sèrie original  $Y_t$  si i només si la distribució estacionària de  $X_t$  és una gamma.

Amb aquest punt de partida, el nostre objectiu principal va ser generalitzar aquest resultat de McKenzie, en el sentit de caracteritzar la distribució de les

innovacions en aquest context, i desenvolupar un contrast de bondat de l'ajust basat en la funció distribució empírica -qüestió que introduïrem en el proper capítol i que estudiarem amb tot detall en el tercer capítol- per tal de decidir si és raonable pensar, amb un cert nivell de confiança, que la distribució de les innovacions és una distribució concreta. En particular, aquest contrast es pot utilitzar, en la situació clàssica, per tal de comprovar si les innovacions en un model autoregressiu de primer ordre són gaussianes.

Aquest contrast s'ha aplicat en primer lloc sobre les captures de peix a l'oceà Atlàntic i golf del riu St. Lawrence, entre 1990 i 1996 per tal d'estudiar si l'assumpció de normalitat de les innovacions és raonable o no. En segon lloc s'ha realitzat el contrast sobre les dades del defactor del producte interior brut espanyol des de 1962 fins a 2011.

Finalment, es presenta un estudi de la potència del contrast, en diferents situacions, considerant diversos valors per al primer coeficient d'autocorrelació, diferents mides mostrals i diverses distribucions per les innovacions.

Aquest treball, introduït al capítol 3, ha estat presentat com a comunicació oral als congressos *IWSM* de 2012 a Praga i a l'*ERCIM* de 2011 a Londres, i s'ha enviat com a article de recerca a la revista *Scandinavian Journal of Statistics*.



---

---

# CAPÍTOL 0

---

## EINES MATEMÀTIQUES

En aquest capítol presentarem algunes de les eines matemàtiques bàsiques i dels resultats més importants que necessitarem fer servir per assolir els objectius plantejats en els propers capítols.

### 0.1 Mètode delta

El mètode delta permet aproximar, sota algunes condicions, la distribució d'una funció d'un estimador asimptòticament normal si se'n coneix el límit de la variància. En aquest sentit, és una generalització molt utilitzada del Teorema Central del Límit. Donarem a continuació la definició d'un concepte clau que ens ajudarà a provar l'objectiu d'aquesta secció, enunciat al teorema 0.1.2 i al teorema 0.1.5

## 0.1. Mètode delta

---

en la seva versió multivariada.

**Definició 0.1.1** *Sigui  $X_n$  una successió de variables aleatòries. Direm que  $X_n$  és asimptòticament normal amb mitjana  $\mu_n$  i variància  $\sigma_n^2$  si  $\sigma_n > 0$  per a tot  $n$  prou gran i es satisfà*

$$\frac{X_n - \mu_n}{\sigma_n} \xrightarrow{d} N(0, 1). \quad (1)$$

El mètode delta es pot enunciar formalment com

**Teorema 0.1.2** *Sigui  $X_n$  una successió de variables aleatòries tal que  $\sqrt{n}(X_n - \mu)$  és asimptòticament normal amb mitjana 0 i variància  $\sigma^2$ . Sigui  $g : \mathbb{R} \rightarrow \mathbb{R}$  una funció arbitrària diferenciable a  $x = \mu$  i complint  $g'(\mu) \neq 0$ . Aleshores,  $\sqrt{n}(g(X_n) - g(\mu))$  és asimptòticament normal amb mitjana 0 i variància  $[g'(\mu)]^2 \sigma^2$ .*

Fent servir el ben conegut desenvolupament en sèrie de potències de Taylor (Definició 0.1.3), podem tenir una prova intuïtiva d'aquest resultat.

**Definició 0.1.3** *Si una funció  $g : \mathbb{R} \rightarrow \mathbb{R}$  té derivades d'ordre  $r$ , és a dir,  $g^{(r)}(x) = \frac{\partial^r}{\partial x^r} g(x)$  existeix, aleshores, per qualsevol constant  $a \in \mathbb{R}$ , el polinomi de Taylor d'ordre  $r$  al voltant d' $a$  és*

$$T_r(x) = \sum_{i=0}^r \frac{g^{(i)}(a)}{i!} (x - a)^i \quad (2)$$

**Teorema 0.1.4 (Taylor)** *Si  $g^{(r)}(a) = \frac{\partial^r}{\partial x^r} g(x) |_{x=a}$  existeix, aleshores*

$$\lim_{x \rightarrow a} \frac{g(x) - T_r(x)}{(x - a)^r} = 0, \quad (3)$$

on  $T_r(x)$  és el polinomi de Taylor d'ordre  $r$  al voltant d' $a$  definit a 0.1.3.

Considerem una variable aleatòria  $X$  amb  $\mathbb{E}(X) = \mu \neq 0$ . Si volem estimar  $g(\mu)$  per una certa funció  $g$ , podem utilitzar el Teorema de Taylor (0.1.4), considerant només el primer terme del polinomi de Taylor per obtenir

$$g(X) \approx g(\mu) + g'(\mu)(X - \mu) \quad (4)$$

Si prenem  $g(X)$  com a estimador de  $g(\mu)$ , podem concloure que aproximadament,

$$\begin{aligned} \mathbb{E}(g(X)) &\approx g(\mu), \\ \text{Var}(g(X)) &\approx [g'(\mu)]^2 \text{Var}(X) \end{aligned} \quad (5)$$

Si  $g'(\mu) = 0$ , caldria fer servir termes d'ordre superior del polinomi de Taylor (0.1.3), i la distribució asimptòtica seria en aquest cas del tipus  $\chi^2$  ([40]).

La versió del mètode delta presentada a 0.1.2 es pot estendre a un context multivariat, que serà, de fet, la situació on ho aplicarem més endavant. El resultat que presenta el mètode delta en un context multivariat és el següent:

**Teorema 0.1.5** *Considerem una variable aleatòria  $\mathbf{X} = (X_1, \dots, X_p)$  a valors a  $\mathbb{R}^p$  i sigui  $\mathbf{X}_1, \dots, \mathbf{X}_n$  una mostra aleatòria amb  $\mathbb{E}(X_{ij}) = \mu_i, i = 1, \dots, n, j = 1, \dots, p$  i  $\text{Cov}(X_{ik}, X_{jk}) = \sigma_{ij}, i = 1, \dots, n, j = 1, \dots, n, k = 1, \dots, p$ . Sigi una funció  $g : \mathbb{R}^p \rightarrow \mathbb{R}$  amb primeres derivades parcials contínues i un valor específic de  $\mu = (\mu_1, \dots, \mu_p)$  complint*

$$\tau^2 = \sum \sum \sigma_{ij} \frac{\partial g(\mu)}{\partial \mu_i} \cdot \frac{\partial g(\mu)}{\partial \mu_j} > 0. \quad (6)$$

## 0.1. Mètode delta

---

Aleshores, si  $\nabla g(\mu) \neq 0$ ,

$$\sqrt{n}[g(\bar{X}_1, \dots, \bar{X}_p) - g(\mu_1, \dots, \mu_p)] \rightarrow N(0, \tau^2) \text{ en distribuci3.} \quad (7)$$

La demostraci3 de 0.1.2 i 0.1.5 es pot trobar, per exemple, a [12] i a [29].

Sovint pot resultar d'utilitat l'aplicaci3 del mètode delta en combinaci3 amb el teorema de Slutsky enunciat a 0.1.6:

**Teorema 0.1.6** (Slutsky) *Donades dues successions de variables aleat3ries  $X_n$  i  $Y_n$  tals que  $X_n$  convergeix en distribuci3 cap a una variable aleat3ria  $X$  i  $Y_n$  convergeix en probabilitat a una constant  $c$ , aleshores*

- $X_n + Y_n$  convergeix en distribuci3 cap a  $X + c$ .
- $Y_n \cdot X_n$  convergeix en distribuci3 cap a  $c \cdot X$ .
- Si  $c \neq 0$ , aleshores  $Y_n^{-1} \cdot X_n$  convergeix en distribuci3 cap a  $c^{-1} \cdot X$ .

Anem a veure un exemple d'aplicaci3 del mètode delta, en un cas senzill i univariat.

**Exemple 0.1.7** [12] *Considerem una variable aleat3ria  $X$  i  $g(x) = \frac{1}{x}$ . En aquest cas, tenim que  $g'(\mu) = -\frac{1}{\mu^2}$ , i per tant, podem escriure*

$$\begin{aligned} \mathbb{E}\left(\frac{1}{\bar{X}}\right) &\approx \frac{1}{\mu}, \\ \text{Var}\left(\frac{1}{\bar{X}}\right) &\approx \left(\frac{1}{\mu}\right)^4 \text{Var}(X) \end{aligned} \quad (8)$$



Suposem que tenim la mitjana  $\bar{X}$  de la variable  $X$  calculada en una mostra aleatòria. Per  $\mu \neq 0$ , tindrem

$$\sqrt{n} \left( \frac{1}{\bar{X}} - \frac{1}{\mu} \right) \rightarrow N \left( 0, \left( \frac{1}{\mu} \right)^4 \text{Var}(X) \right) \text{ en distribució.} \quad (9)$$

Si no coneixem la variància de  $X$ , necessitarem una estimació  $s^2$  per obtenir l'aproximació a  $\text{Var} \left( \frac{1}{\bar{X}} \right)$ . Com que tampoc coneixem  $\mu$ , haurem d'estimar també  $\frac{1}{\mu}$ , obtenint finalment

$$\widehat{\text{Var}} \left( \frac{1}{\bar{X}} \right) \approx \left( \frac{1}{\bar{X}} \right)^4 s^2 \quad (10)$$

I com a conseqüència del teorema de Slutsky (0.1.6) es pot concloure que per  $\mu \neq 0$  tenim que

$$\frac{\sqrt{n} \left( \frac{1}{\bar{X}} - \frac{1}{\mu} \right)}{\left( \frac{1}{\bar{X}} \right)^2 \cdot s} \rightarrow N(0, 1) \text{ en distribució.} \quad (11)$$

El següent exemple mostra una possible aplicació del mètode delta en una situació multivariada:

**Exemple 0.1.8** [13] *Imaginem que es vol estudiar la distribució asimptòtica de l'índex de dispersió mostral, definit com*

$$id = \frac{s^2}{\bar{x}},$$

*és a dir, el quocient entre la variància i la mitjana mostrals. Aquest estadístic és interessant, entre altres situacions, quan convé estudiar diferències en la dispersió d'un paràmetre en diferents grups. Per exemple, Chambers et al. proposen a [13]*

## 0.2. Equacions en diferències

---

*l'estudi de les variacions morfològiques de les larves de certs peixos. Per a fer-ho, necessitem conèixer la distribució assintòtica de la mitjana i variància mostrals, que per a una variable aleatòria amb moments finits d'ordre 4 és ([39]):*

$$(\bar{x}, s^2) \approx N \left[ (\mu, \sigma^2), \frac{1}{n} \begin{pmatrix} \sigma^2 & \mu_3 \\ \mu_3 & \mu_4 - \sigma^4 \end{pmatrix} \right]$$

*on  $n$  representa la mida mostral,  $\mu$  i  $\sigma^2$  són respectivament la mitjana i variància poblacionals i  $\mu_3$  i  $\mu_4$  són els moments d'ordre 3 i 4 centrats en l'esperança respectivament. Aplicarem el teorema 0.1.5 amb  $g(x_1, x_2) = \frac{x_2}{x_1}$ , obtenint  $\nabla g(x_1, x_2) = (-\frac{x_2}{x_1^2}, \frac{1}{x_1})$ . Per tant, tindrem que*

$$id = \frac{s^2}{\bar{x}} \approx N \left[ \frac{\sigma^2}{\mu}, \frac{1}{n} \left( \frac{\sigma^6 - 2\sigma^2\mu_3\mu + \mu^2\mu_4 - \mu^2\sigma^4}{\mu^4} \right) \right]$$

*Per exemple, en el cas d'una variable aleatòria amb distribució de Poisson, aquest resultat és útil per contrastar la sobredispersió. En aquest cas, tenint en compte que  $\sigma^2 = \mu_3 = \mu$  i que  $\mu_4 = \mu + 3\mu^2$ , l'expressió anterior es pot escriure com*

$$id = \frac{s^2}{\bar{x}} \approx N \left[ 1, \frac{2}{n} \right]$$

## 0.2 Equacions en diferències

Una relació de recurrència o equació en diferències és una equació que defineix una seqüència de forma recursiva, partint d'un o diversos termes inicials. La forma de la solució dependrà de l'estructura de l'equació en diferències, nosaltres estarem

interessats en trobar la solució d'equacions en diferències lineals no homogènies amb coeficients constants, que són de la forma

$$b_n = a_n + c_1 a_{n-1} + c_2 a_{n-2} + \dots + c_d a_{n-d}, \quad (12)$$

on tots els coeficients  $c_i$  són constants,  $b_n$  és coneguda i es vol determinar  $a_n$  a partir d'unes certes condicions inicials.

A partir de l'equació homogènia, es construeix el *polinomi característic* com

$$p(t) = t^d - c_1 t^{d-1} - c_2 t^{d-2} - \dots - c_d \quad (13)$$

Com veurem al teorema 0.2.1, les arrels del polinomi característic tenen un paper crucial en la recerca de les solucions d'una equació en diferències d'aquest tipus. Si les arrels  $r_1, r_2, \dots$  són totes diferents, la solució de l'equació homogènia pren la forma

$$a_n = k_1 r_1^n + k_2 r_2^n + \dots + k_d r_d^n, \quad (14)$$

on els coeficients  $k_i$  es poden determinar d'acord amb les condicions inicials de l'equació. En cas que hi hagi arrels múltiples, els termes de (14) corresponents a les aparicions posteriors a la primera de la mateixa arrel es multipliquen per  $n^i$ , on  $i$  és la potència corresponent. Per exemple, si el polinomi característic es pot factoritzar com  $(x - r)^3$ , amb una arrel  $r$  que apareix tres cops, aleshores la

## 0.2. Equacions en diferències

---

solució seria de la forma

$$a_n = k_1 r^n + k_2 n r^n + k_3 n^2 r^n \quad (15)$$

La solució d'una equació en diferències lineal no homogènia és la suma d'una solució particular, que pot trobar-se pel mètode dels coeficients indeterminats, i la solució general de l'equació homogènia. Un punt important, que farem servir més endavant és el que afirma el resultat següent:

**Teorema 0.2.1** *La solució de l'equació en diferències lineal homogènia d'ordre  $d$*

$$a_n = c_1 a_{n-1} + c_2 a_{n-2} + \dots + c_d a_{n-d}, \quad (16)$$

*convergeix asimptòticament a zero si i només si les arrels del polinomi característic són menors que 1 en valor absolut (o en mòdul en el cas complex).*

La demostració del teorema 0.2.1 i una explicació en un context més general de les equacions en diferències i els diferents mètodes per resoldre-les es pot trobar en abundant bibliografia, com per exemple a [7] o [33].

**Exemple 0.2.2** *La successió de Fibonacci és un exemple clàssic d'equació en diferències amb coeficients constants homogènia. Es defineix mitjançant la relació de recurrència*

$$a_n = a_{n-1} + a_{n-2},$$

*amb les condicions inicials  $a_0 = 0$  i  $a_1 = 1$ . El polinomi característic d'aquesta equació en diferències és  $p(t) = t^2 - t - 1$ . La solució es pot trobar de la manera*

descrita, tenint en compte que les arrels del polinomi característic són  $r_1 = \frac{\sqrt{5}}{2} + \frac{1}{2}$  i  $r_2 = \frac{1}{2} - \frac{\sqrt{5}}{2}$  obtenint en aquest cas que la solució ve donada per l'expressió

$$a_n = k_1 \left( \frac{\sqrt{5}}{2} + \frac{1}{2} \right)^n + k_2 \left( \frac{1}{2} - \frac{\sqrt{5}}{2} \right)^n$$

Imposant les condicions inicials és clar que  $k_1 = -k_2 = \frac{1}{\sqrt{5}}$ , de manera que la solució serà

$$a_n = \frac{1}{\sqrt{5}} \left( \frac{\sqrt{5}}{2} + \frac{1}{2} \right)^n - \frac{1}{\sqrt{5}} \left( \frac{1}{2} - \frac{\sqrt{5}}{2} \right)^n$$

Considerem ara una petita variació d'aquesta mateixa equació en diferències, per convertir-la en no homogènia, per exemple

$$n = a_n - a_{n-1} - a_{n-2},$$

on  $b_n = n$ . La solució de l'equació en diferències serà  $P_t + H_t$ , on  $P_t$  és qualsevol solució particular de l'equació no homogènia i  $H_t$  és la solució general de l'homogènia, que serà la que hem trobat anteriorment,

$$a_n = k_1 \left( \frac{\sqrt{5}}{2} + \frac{1}{2} \right)^n + k_2 \left( \frac{1}{2} - \frac{\sqrt{5}}{2} \right)^n$$

Per trobar  $P_t$  considerem com a candidata una expressió de la forma de  $b_n = n$ , és a dir, un polinomi de grau 1,  $P_t = k_3 n + k_4$ . Utilitzant aquesta expressió a

### 0.3. Equacions funcionals

---

*l'equació original tenim que*

$$n = k_3n + k_4 - k_3(n-1) - k_4 - k_3(n-2) - k_4$$

*Igalant els termes de grau 1 obtenim que  $k_3 = -1$ , i igualant els termes de grau 0 tenim que  $k_4 = -3$ . Per tant, la solució de l'equació no homogènia vindrà donada per l'expressió*

$$a_n = k_1 \left( \frac{\sqrt{5}}{2} + \frac{1}{2} \right)^n + k_2 \left( \frac{1}{2} - \frac{\sqrt{5}}{2} \right)^n - n - 3$$

*I imposant les condicions inicials  $a_0 = 0$  i  $a_1 = 1$ , trobem finalment que la solució de l'equació no homogènia és*

$$a_n = \left( 3 - \frac{1}{10}(3\sqrt{5} - 7)\sqrt{5} \right) \left( \frac{\sqrt{5}}{2} + \frac{1}{2} \right)^n + \frac{1}{10}(3\sqrt{5} - 7)\sqrt{5} \left( \frac{1}{2} - \frac{\sqrt{5}}{2} \right)^n - n - 3$$

## 0.3 Equacions funcionals

Una equació funcional és una equació en la qual les incògnites són funcions. A [25] es presenta una definició formal d'equació funcional, basada en la següent definició de *terme*.

**Definició 0.3.1** *Un terme es defineix per les següents condicions:*

1. *Qualsevol variable independent és un terme.*
2. *Si  $t_1, \dots, t_p$  són termes i  $f(x_1, \dots, x_p)$  és una funció de  $p$  variables, aleshores*

$f(t_1, \dots, t_p)$  també és un terme.

3. No existeix cap altre terme.

**Definició 0.3.2** *Una equació funcional és una equació entre dos termes que conté com a mínim una funció desconeguda i un nombre finit de variables independents.*

Una equació funcional molt coneguda, i que permet la caracterització de la funció gamma és la que apareix al teorema de Bohr-Mollerup (0.3.3). Una demostració es pot trobar per exemple a [6].

**Teorema 0.3.3** *(Bohr - Mollerup) La funció  $\Gamma(x)$ , definida per  $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ ,  $x > 0$  és l'única funció  $f$  logàritmicament convexa, amb  $f(1) = 1$  i que satisfà l'equació funcional  $f(x+1) = xf(x)$  per  $x > 0$ .*

Prenent logaritmes en l'equació funcional del teorema de Bohr-Mollerup 0.3.3, i dient  $\phi(x) = \log f(x)$ , obtenim l'equació funcional

$$\phi(x+1) - \phi(x) = \log(x) \tag{17}$$

El nostre interès se centrarà en equacions de la forma

$$\phi(x+1) - \phi(x) = F(x), \tag{18}$$

Observem que l'equació funcional (17) definida pel teorema de Bohr-Mollerup n'és un cas particular.

Les solucions d'aquestes equacions que ens interessaran al tercer capítol són, de fet, funcions generatrius de cumulants, i per tant, podrem restringir-nos al cas en el qual les solucions són funcions convexes. En aquest context, ens serà útil el teorema següent:

**Teorema 0.3.4** (Krull [25]). *Sigui  $F$  una funció còncaua a  $I = (a, \infty)$ ,  $a \geq \infty$ . Sigui  $x_0 \in (a, \infty)$  arbitràriament fixat i suposem que*

$$\lim_{x \rightarrow \infty} [F(x+1) - F(x)] = 0. \quad (19)$$

*Aleshores, per tot  $\nu_0 \in (-\infty, \infty)$  existeix una única funció  $\phi$  convexa en  $I$  satisfent l'equació (18) i complint la condició  $\phi(x_0) = \nu_0$*

**Exemple 0.3.5** A [26] es presenta un exemple molt senzill d'equació funcional en una sola variable:

$$\phi(x+1) = \phi(x)$$

*És clar que qualsevol funció periòdica de període 1 és solució d'aquesta equació funcional. Per tenir una única solució cal imposar certes condicions de regularitat sobre la funció  $\phi$ . Observem que exigir que  $\phi$  sigui contínua o analítica no són condicions suficients per tal d'assegurar la unicitat de la solució de l'equació funcional. Considerem l'equació funcional*

$$\phi\left(\frac{1}{2}x\right) = \phi(x)$$



Fent el canvi  $x = 2^u$ ,  $\psi(u) = \phi(2^{u-1})$ , l'equació funcional anterior es converteix en

$$\psi(u + 1) = \psi(u),$$

equivalent a la que consideràvem anteriorment per  $x < 0$  i  $x > 0$ . Ara bé, si considerem també  $x = 0$ , la solució canvia i les úniques solucions d'aquesta equació que són contínues en  $x = 0$  són les funcions constants. En efecte, per inducció tenim que

$$\phi\left(\frac{1}{2^n}x\right) = \phi(x),$$

i prenent límits quan  $n$  va cap a infinit tenim que si  $\phi$  és contínua en  $x = 0$ , aleshores  $\phi(x) = \phi(0) = k$ , on  $k$  és una constant.

## 0.4 Contrastos de bondat de l'ajust basats en la funció de distribució empírica

La verificació de l'ajust dels models és un procediment bàsic en la inferència estadística ([28]). Partint d'una mostra  $x_1, x_2, \dots, x_n$  en la qual es vol contrastar una certa hipotèsi  $H_0$  que prové d'una població amb funció de distribució acumulada  $F(x, \theta)$ , els contrastos de bondat de l'ajust basats en la funció de distribució empírica (EDF) es fonamenten en la comparació entre les funcions de distribució

#### 0.4.1. Distribució asimptòtica de $W_n^2$

---

empírica i hipotètica, és a dir, entre  $F(x, \theta)$  i  $F_n(x)$ , on

$$\begin{aligned} F_n(x) &= 0, & x < x_1 \\ F_n(x) &= \frac{i}{n}, & x_i \leq x < x_{i+1} \quad i = 1, \dots, n-1 \\ F_n(x) &= 1, & x_n \leq x \end{aligned} \tag{20}$$

La diferència entre les funcions de distribució empírica i teòrica es pot definir de moltes maneres. Per exemple, utilitzant estadístics de la forma

$$KS = \max_x | F_n(x) - F(x, \theta) |, \tag{21}$$

com en el cas del contrast de Kolmogorov-Smirnov, o bé utilitzant les diferències al quadrat ponderades, de forma genèrica

$$W_n^2 = n \int (F_n(x) - F(x, \theta))^2 \psi(x) dF(x), \tag{22}$$

on  $\psi(x)$  és la funció que pondera les diferències quadràtiques. Si triem  $\psi(x)$  de forma que es doni més pes a les cues de la distribució que no pas a la part central, obtindrem un estadístic de la forma del d'Anderson-Darling. L'estadístic de Cramér-von Mises es correspon al cas particular de (22) en el qual  $\psi(x) = 1$ , que serà clau al tercer capítol, on veurem com estimar-lo en la pràctica.

#### 0.4.1 Distribució asimptòtica de $W_n^2$

La distribució asimptòtica de l'estadístic de Cramér-von Mises en el context dels models autoregressius de primer ordre (i també per processos MA(1)) s'ha estu-

diat en detall per exemple a [4] en el cas en el qual els paràmetres són coneguts.

A [2] es tracta en detall, també en un context de models autoregressius de primer ordre, la situació en la qual els paràmetres són desconeguts, que evidentment, és la situació que presenta un major interès pràctic.

En el mateix context que a [4], a [5] i [2] es para especial atenció a la distribució asimptòtica de l'estadístic de Cramér-von Mises en el cas de models autoregressius de primer ordre, a més d'altres estadístics de bondat de l'ajust basats en la funció de distribució empírica.

En un context més general, les distribucions asimptòtiques i exacta de l'estadístic de Cramér-von Mises es presenten a [14], a més d'una correcció de la funció de distribució asimptòtica que aproxima la funció exacta i que resulta particularment útil per a mostres petites.

## 0.5 Distribucions autodescomposables

Les distribucions autodescomposables es defineixen com les distribucions amb funcions característiques autodescomposables, definides de la manera següent:

**Definició 0.5.1** [41] *La funció característica  $\phi$  és autodescomposable si per qual-*

## 0.5. Distributions autodescomposables

---

sevol  $0 < \alpha < 1$  satisfà

$$\phi(t) = \phi(\alpha t) \cdot \phi_\alpha(t), -\infty < t < \infty, \quad (23)$$

on  $\phi_\alpha$  és també una funció característica.

**Exemple 0.5.2** Considerem per exemple una variable aleatòria amb distribució de Bernoulli, amb probabilitat d'èxit  $p$ . La seva funció característica és  $\Phi(t) = (1 - p) + pe^{it}$ . Si fos autodescomposable, tindriem que

$$\Phi_\alpha(t) = \frac{\Phi(t)}{\Phi(\alpha t)} = \frac{1 - p + pe^{it}}{1 - p + pe^{i\alpha t}}$$

és una funció característica. Però observem que  $1 - p + pe^{i\alpha t}$  s'anul·la per  $t = -\frac{\ln(p-1)}{p\alpha}$ , de manera que  $\Phi(t)$  no és fitada, i per tant, no pot ser una funció característica.

Per altra banda, la distribució normal és descomposable, ja que la seva funció característica és  $\phi(t) = e^{i\mu t - \frac{1}{2}\sigma^2 t^2}$ , de manera que tindrem

$$\phi_\alpha(t) = \frac{\phi(t)}{\phi(\alpha t)} = e^{(1-\alpha)i\mu t - \frac{1}{2}\sigma^2 t^2(1-\alpha^2)},$$

que clarament és també una funció característica.

---

---

# CAPÍTOL 1

---

## ANÀLISI DE SÈRIES TEMPORALS

### 1.1 Introducció

Una sèrie temporal és un conjunt d'observacions  $X_t$ , cadascuna recollida en un moment de temps concret  $t$  [8]. L'evident correlació que s'introdueix en considerar moments de temps adjacents questiona l'ús de moltes tècniques inferencials habituals, basades en la hipòtesi que les observacions són independents i idènticament distribuïdes. L'enfocament sistemàtic que intenta donar resposta a les qüestions matemàtiques i estadístiques plantejades per aquestes correlacions és el que es coneix habitualment com anàlisi de sèries temporals, segons la definició de [42]. Un altre possible enfocament a aquesta problemàtica ens l'ofereix el domini de freqüències, que assumeix que les principals característiques d'interès en l'anàlisi

## 1.1. Introducció

---

de sèries temporals està relacionat amb les variacions periòdiques sinusoidals que s'observen de forma natural en molts fenòmens. Aquestes variacions periòdiques són habitualment causades per processos biològics, físics o ambientals, tal com s'indica a [42].

Una de les estructures més habituals i clàssiques en el context de l'anàlisi de sèries temporals és l'estructura dels models autoregressius, definits per

$$X_t = \sum_{i=1}^p \alpha_i X_{t-i} + \epsilon_t, \quad (1.1)$$

on  $\alpha_1, \dots, \alpha_p$  són els paràmetres del model i  $\epsilon_t$  és un soroll blanc. Per tal d'assegurar l'estacionarietat del procés, cal que totes les arrels  $z_1, \dots, z_p$  del polinomi característic  $P(z) = z^p - \sum_{i=1}^p \alpha_i z^{p-i}$  satisfacin  $|z_i| < 1$ .

Els paràmetres d'un model autoregressiu es poden estimar mitjançant diverses tècniques, entre les més habituals hi ha el mètode de màxima versemblança o el mètode dels moments, basat en les equacions de Yule-Walker, definides per

$$\gamma_m = \sum_{k=1}^p \alpha_k \gamma_{m-k}, \quad (1.2)$$

on  $\gamma_m$  és la funció d'autocovariància del procés:

**Definició 1.1.1** *Sigui  $X_t$  un procés estacionari. La seva funció d'autocovariàn-*

cia és una funció  $\gamma : \mathbb{Z} \rightarrow \mathbb{R}$  definida per:

$$\gamma(h) = \mathbb{E}(X_{t+h}X_t) - \mathbb{E}(X_{t+h})\mathbb{E}(X_t), h \in \mathbb{Z} \quad (1.3)$$

En el cas que ens ocuparà al tercer capítol, el dels models autoregressius de primer ordre, tenim que la funció d'autocorrelació és  $\rho(k) = \alpha^k$ , i les equacions de Yule-Walker en aquest cas ens permeten estimar fàcilment el paràmetre del model a través de l'expressió:

$$\hat{\alpha} = \frac{\hat{\gamma}_1}{\hat{\gamma}_0}, \quad (1.4)$$

on  $\hat{\gamma}_1$  és el coeficient d'autocorrelació empíric i  $\hat{\gamma}_0$  és la variància empírica.

Altres models de sèries temporals s'han descrit abastament en la literatura partint dels models autoregressius, com per exemple els models autoregressius de mitjana mòbil (ARMA), definits per

$$X_t = c + \epsilon_t + \sum_{i=1}^p \phi_i X_{t-i} + \sum_{j=1}^q \theta_j \epsilon_{t-j} \quad (1.5)$$

Aquests models van ser introduïts el 1951 per Peter Whittle i el seu ús és molt comú, especialment després de la publicació de la metodologia de Box-Jenkins al 1971 per a l'anàlisi d'aquest tipus de processos.

## 1.2 Sèries temporals discretes

La modelització de sèries temporals discretes ha estat una qüestió tractada recentment per nombrosos autors, i s'han considerat diversos models. A [32] es presenten, entre altres, nombrosos models basats en l'operador *binomial thinning* que definirem a (1.7), com els descrits a continuació.

### 1.2.1 Models INAR

En el proper capítol ens centrarem en els models enters autoregressius (INAR), que són una extensió natural al cas discret dels ben coneguts models autoregressius (1.1) (AR), i que sovint tenen una interpretació senzilla des d'un punt de vista pràctic. Aquests models es defineixen de la manera següent:

$$X_t = p_1 \circ X_{t-1} + \dots + p_k \circ X_{t-k} + W_t, \quad (1.6)$$

on  $p_1, \dots, p_k$  són paràmetres fixats complint  $0 < p_1, \dots, p_k < 1$  i s'assumeix que  $W_t$  segueix una distribució de Poisson de mitjana fixada  $\lambda$ . A més, també es demana que  $X_{t-i}$ ,  $i = 1, \dots, k$  i  $W_t$  siguin independents per tot temps  $t$ . L'operador  $\circ$  que apareix a l'expressió (1.6), anomenat *binomial thinning*, es defineix de la forma següent:

$$p_j \circ X_{t-j} = \sum_{i=1}^{X_{t-j}} Y_i, \quad (1.7)$$



on  $Y_i$  són variables aleatòries de Bernoulli independents i idènticament distribuïdes amb probabilitat d'èxit igual a  $p_j$ . Per tant, si  $X_{t-j} = x_{t-j}$ , aleshores  $p_j \circ x_{t-j}$  es distribueix com una binomial amb nombre de proves igual a  $x_{t-j}$ . Una de les primeres propostes més interessants sobre l'aplicació d'aquests models a l'anàlisi de la incidència de certes malalties és [11]. Una revisió més recent sobre els models INAR es pot trobar a [23], i recentment s'han desenvolupat algunes generalitzacions d'aquests models com per exemple la que es proposa a [18]. Una propietat molt interessant de l'operador *binomial thinning*, que utilitzarem més endavant, és que el *binomial thinning* d'una variable aleatòria de Poisson de mitjana  $\lambda$  es distribueix també seguint una Poisson de mitjana  $p \cdot \lambda$ .

Un bon resum de les propietats del *binomial thinning* es pot trobar a [45] o [46]. Entre les més importants, i que farem servir en el proper capítol, podríem citar les següents:

- $\mathbb{E}[p \circ X] = p\mathbb{E}[X]$
- $\mathbb{V}[p \circ X] = p^2\mathbb{V}[X] + p(1 - p)\mathbb{E}[X]$
- La funció generatriu de probabilitats de  $Y = p \circ X$  és  $f_Y(t) = f_X(1 - p + pt)$
- $(p_1 \circ p_2) \circ X = (p_1 \cdot p_2) \circ X$

Una caracterització de distribucions de recomptes basada en aquests tipus de propietats del *binomial thinning* es pot trobar a [36]. Qualsevol d'aquestes distribucions, que inclouen la de Poisson, poden ser distribucions apropiades per  $W_t$ .

**Exemple 1.2.1** Considerem l'exemple clàssic de les dades de Fürth. Aquestes dades corresponen a una sèrie de 505 observacions del nombre de persones que es poden comptar des d'una ubicació determinada, en intervals de 5 segons. La gràfica d'aquestes dades es mostra a la figura 1.1. Les funcions d'autocorrelació i

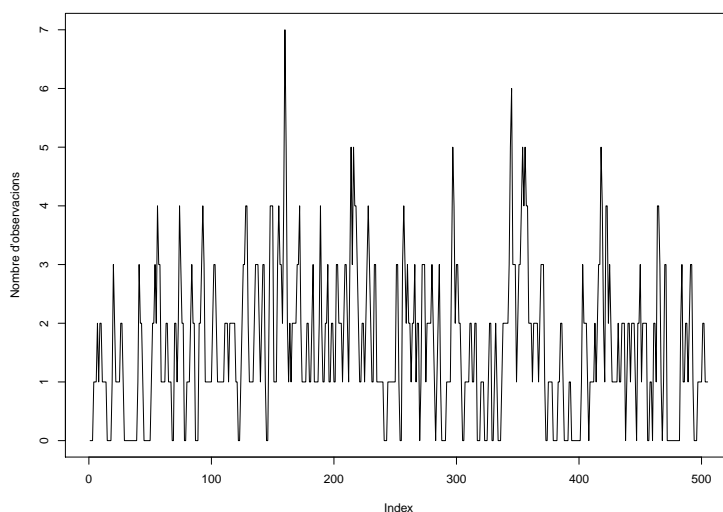


Figura 1.1: Dades de Fürth

d'autocorrelació parcial mostren que un model d'ordre 2 pot resultar adequat (figura 1.2). Estimant els paràmetres pel mètode de màxima versemblança, obtenim  $p_1 = 0,65 \pm 0,024$ ,  $p_2 = 0,22 \pm 0,027$  i  $\lambda = 0,56 \pm 0,054$ . És a dir, considerem el model

$$X_t = 0,65 \circ X_{t-1} + 0,22 \circ X_{t-2} + W_t(0,56) \quad (1.8)$$

De fet, a la literatura ja s'ha utilitzat aquest model per ajustar aquest conjunt de dades, per exemple a ([23]). Utilitzant les dues primeres observacions

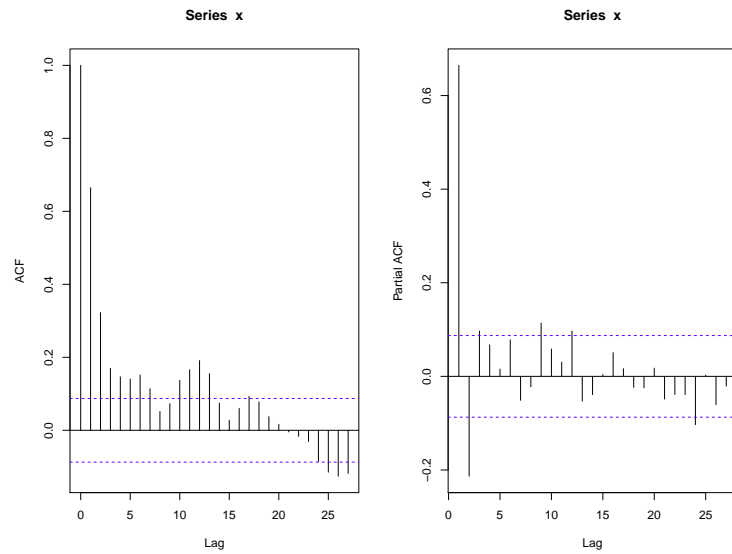


Figura 1.2: Funcions d'autocorrelació i autocorrelació parcial per a les dades de Fürth

*de les dades originals de Fürth, podem generar una sèrie simulada i comparar-la amb la sèrie original. A la figura 1.3 podem veure els darrers 100 valors de la sèrie original i la simulació basada en un model INAR(2), i observem que són extremadament properes al llarg del temps.*

### 1.2.2 Models INMA(q)

A [32] es presenten també els models INMA(1) i la seva extensió a ordre superior, els models INMA(q), definits per

$$X_t = c_1 \circ W_{t-1} + \dots + c_q \circ W_{t-q} + W_t,$$

### 1.2.2. Models INMA(q)

---

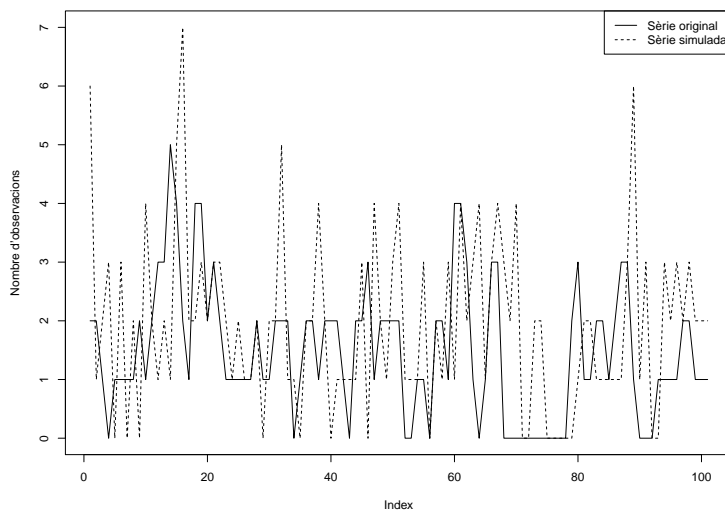


Figura 1.3: Sèrie original de Fürth i simulació basada en un model INAR(2)

on  $c_i \in [0, 1]$  per a tot  $i$ ,  $C = 1 + \sum_{i=1}^q c_i$  i  $W_{t-j}$ ,  $j = 0..q$  són independents i idènticament distribuïdes amb distribució de Poisson. Totes les operacions de *thinning* a temps  $t$  es realitzen de forma independent. En aquest cas,  $X_t$  és una seqüència de variables aleatòries de Poisson de mitjana  $\theta$ .

Aquests processos es poden interpretar com a generalitzacions de cues M/D/ $\infty$ .

Els processos INAR es poden pensar com a processos INMA( $\infty$ ). Per exemple, suposem que tenim unes dades que podem modelitzar per un procés INAR( $p$ ),

amb  $p = 1, 2$ . Aleshores, la distribució marginal de  $X_t$  es pot representar per

$$X_t = \sum_{i=1}^{\infty} c_j \circ W_{t-j} + W_t \quad (1.9)$$

on  $c_j = a^j$  si  $p = 1$  i  $c_1 = a_1$ ,  $c_j = a_1 c_{j-1} + a_2 c_{j-2}$ ,  $j = 2, 3, \dots$  si  $p = 2$ .

En el cas particular dels processos INMA(1), suposant que el procés segueix una llei de Poisson de paràmetre  $\lambda$ , la mitjana i la variància del procés vénen donats per  $\mathbb{E}[X_t] = Var(X_t) = (1+c_1)\lambda$ . Els estimadors pel mètode dels moments dels paràmetres es poden obtenir a partir de l'esperança del model:

$$\hat{c}_1 = \frac{\hat{\rho}(1)}{1 - \hat{\rho}(1)}; \hat{\lambda} = \frac{\bar{X}}{1 + \hat{c}_1}, \quad (1.10)$$

on  $\hat{\rho}$  és la funció d'autocorrelació empírica, definida per

$$\hat{\rho}(j) = \frac{\hat{Cov}(X_t, X_{t-j})}{\hat{Var}(X_t)} \quad (1.11)$$

El procés INMA(2) és una extensió dels processos INMA(1), amb estructura

$$X_t = c_1 \circ W_{t-1} + c_2 \circ W_{t-2} + W_t,$$

on els paràmetres  $c_i \in [0, 1]$ , per  $i = 1, 2$ . S'assumeix que les operacions  $c_i \circ W_{t-i}$  es realitzen de manera independent,  $i = 1, 2$  entre elles i també de manera independent en el temps. Els moments d'aquest procés són  $\mathbb{E}[X_t] = Var(X_t) =$

### 1.2.3. Models INARMA(1,q)

---

$\lambda(1 + c_1 + c_2)$ , i la seva ACF és

$$\rho = \begin{cases} \frac{\sum_{i=0}^{2-k} c_i c_{i+k}}{1+c_1+c_2} & \text{si } k = 1, 2 \\ 0 & \text{si } k > 2 \end{cases} \quad (1.12)$$

amb  $c_0 = 1$ . A partir de l'esperança del model que hem vist, es poden deduir els estimadors dels paràmetres pel mètode dels moments per al cas dels models INMA(2), suposant que el procés segueix una llei de Poisson de paràmetre  $\lambda$ :

$$\begin{aligned} \hat{c}_1 &= \frac{\hat{\rho}(1) - 1 + [(1 - \hat{\rho}(1))^2 + 4\hat{\rho}(1)\hat{\rho}(2)]^{1/2}}{2\hat{\rho}(2)} \\ \hat{c}_2 &= \frac{\hat{\rho}(1) + 2\hat{\rho}(2) - 1 + [(1 - \hat{\rho}(1))^2 + 4\hat{\rho}(1)\hat{\rho}(2)]^{1/2}}{2(1 - \hat{\rho}(2))} \\ \hat{\lambda} &= \frac{\bar{X}}{1 + \hat{c}_1 + \hat{c}_2}, \end{aligned}$$

on  $\rho$  és la funció d'autocorrelació empírica.

### 1.2.3 Models INARMA(1,q)

Aquest procés es construeix combinant els dos processos INAR(1) i MA(q) descrits anteriorment. Un procés INARMA(1,q) es defineix com

$$X_t = \alpha \circ X_{t-1} + \sum_{i=1}^q c_i \circ W_{t-i} + W_t, \quad (1.13)$$

on totes les operacions són independents, i  $W_t$  són independents i idènticament distribuïdes amb distribució de Poisson de mitjana  $\lambda$ .

Observem que si  $\alpha = 0$  obtenim el procés INMA(q) descrit anteriorment, i de forma similar, si  $c_k = 0$  per a tot  $k$ , obtenim el procés INAR(1).

### 1.2.4 Altres models per a sèries temporals discretes

En el mateix document [32] es descriuen altres models per a sèries temporals discretes, com models INAR multivariats o models amb altres distribucions marginals com la distribució geomètrica, binomial o binomial negativa.

## 1.3 Anàlisi espectral de sèries temporals

La funció d'autocovariància d'un procés estacionari  $X_t$ , definida a 1.1.1 satisfà les propietats següents:

1.  $\gamma \geq 0$
2.  $|\gamma(t)| \leq \gamma(0)$
3.  $\gamma(t) = \gamma(-t), t \in \mathbb{Z}$

**Definició 1.3.1** *L'espectre d'un procés estacionari a valors reals és la transformada de Fourier de la seva funció d'autocovariància teòrica. Dit d'una altra forma, si  $X_t$  és un procés estacionari a valors reals, amb funció d'autocovariància  $\gamma(t), t \in \mathbb{Z}$ . Aleshores, la densitat espectral o espectre de  $X_t$  és*

$$f(\lambda) = \sum_{t \in \mathbb{Z}} \gamma(t) e^{-i2\pi\lambda t} = \gamma(0) + 2 \sum_{t \in \mathbb{N}} \gamma(t) \cos(2\pi\lambda t) \quad (1.14)$$

El següent resultat caracteritza la densitat espectral d'un procés estacionari:

**Teorema 1.3.2** *Una funció  $f : [0, 1] \rightarrow \mathbb{R}$  és la densitat espectral d'un procés estacionari si i només si satisfà les tres condicions*

1.  $f(\lambda) \geq 0$
2.  $f(\lambda) = f(1 - \lambda)$
3.  $\int_0^1 f(\lambda) d\lambda < \infty$

Sovint parlarem de densitat espectral estandarditzada, indicant que la funció espectral està definida en termes d'autocorrelacions en comptes d'autocovariàncies ([3]).

**Proposició 1.3.3** *Si  $Y_t$  un procés estacionari amb densitat espectral  $f(\lambda)$ . Aleshores, és possible recuperar la funció d'autocovariància a partir de  $f(\lambda)$ .*

**Prova.** La funció d'autocovariància es pot recuperar considerant la transformada de Fourier inversa de la densitat espectral:

$$\gamma(t) = \int_0^1 f(\lambda) e^{i2\pi\lambda t} d\lambda = \int_0^1 f(\lambda) \cos(2\pi\lambda t) d\lambda$$

■

**Observació 1.3.4** *És important remarcar que, d'acord amb la proposició 1.3.3, és equivalent conèixer la funció d'autocovariància o bé la densitat espectral d'un procés.*



**Exemple 1.3.5** Considerem un procés  $AR(1)$  definit per  $X_t = \alpha X_{t-1} + W_t$ . És ben conegut que la funció d'autocovariància ve donada per l'expressió  $\gamma(t) = \frac{\sigma^2 \alpha^{|t|}}{1-\alpha^2}$ . A partir d'aquí podem calcular la funció de densitat espectral:

$$\begin{aligned}
 f(\lambda) &= \sum_{t=-\infty}^{\infty} \gamma(t) e^{-2\pi i \lambda t} = \frac{\sigma^2}{1-\alpha^2} \sum_{t=-\infty}^{\infty} \alpha^{|t|} e^{-2\pi i \lambda t} \\
 &= \frac{\sigma^2}{1-\alpha^2} \left( 1 + \sum_{t=1}^{\infty} \alpha^t (e^{-2\pi i \lambda t} + e^{2\pi i \lambda t}) \right) \\
 &= \frac{\sigma^2}{1-\alpha^2} \left( 1 + \frac{\alpha e^{-2\pi i \lambda}}{1-\alpha e^{-2\pi i \lambda}} + \frac{\alpha e^{2\pi i \lambda}}{1-\alpha e^{2\pi i \lambda}} \right) \\
 &= \frac{\sigma^2}{(1-\alpha^2)} \frac{1-\alpha e^{-2\pi i \lambda} \alpha e^{2\pi i \lambda}}{(1-\alpha e^{-2\pi i \lambda})(1-\alpha e^{2\pi i \lambda})} \\
 &= \frac{\sigma^2}{1-2\alpha \cos(2\pi \lambda) + \alpha^2}
 \end{aligned}$$

Observem que aquesta densitat espectral és la mateixa que en el cas d'un procés  $INAR(1)$ .



---

---

## CAPÍTOL 2

---

# APLICACIÓ ALS INGRESSOS HOSPITALARIS PER MALALTIES ESTACIONALS

En aquest capítol proposarem una modificació dels models INAR introduïts al capítol anterior amb l'objectiu d'incloure el comportament estacional que apareix en moltes situacions. En particular, veurem com un model basat en una sèrie discreta autoregressiva de segon ordre es pot aplicar per analitzar el nombre d'arribades al servei d'urgències d'un hospital motivades per malalties que presenten un comportament estacional.

## 2.1 Introducció

Sovint, la decisió d'hospitalitzar o no un pacient que arriba al Servei d'Urgències d'un hospital es pren no només en funció de la gravetat de la malaltia, sino que també està relacionada amb el nombre de places disponibles en aquell moment. És, doncs, d'una importància vital, tenir eines que permetin realitzar prediccions acurades del nombre futur d'arribades als serveis d'urgències en base al que ha succeït en el passat. Tenint en compte la proporció d'arribades als serveis d'urgències que acaben en ingrés hospitalari, les prediccions del nombre d'arribades poden ser molt útils a l'hora de preveure el nombre de places hospitalàries que caldrà tenir disponibles en el futur. El model teòric que definirem en la següent secció s'aplicarà a un conjunt de dades referents al nombre d'arribades per setmana al Servei d'Urgències de l'*Hospital Clínic i Provincial* de Barcelona per causes atribuïbles a la grip, tot i que cal fer notar que hi ha diverses malalties que poden presentar símptomes semblants (per exemple altres infeccions víriques que poden provocar pneumònia o altres patologies). Això complica notablement la tasca del diagnòstic inicial en el moment en què el pacient arriba al Servei d'Urgències de l'hospital.

La grip és una malaltia que acostuma a aparèixer en forma epidèmica un cop l'any, generalment a l'hivern. L'aparició de la pandèmia de grip H1N1 al 2009 es pot considerar una excepció, en tractar-se d'una soca diferent i tenir, com veurem més endavant, un comportament marcadament llunyà al de la grip estacional. Per fer front a una epidèmia amb garanties d'èxit cal disposar dels recursos necessaris per atendre les necessitats sanitàries immediates de la població, a més de tenir la

capacitat de gestionar els nous casos i donar servei a tots els pacients en el context de l'epidèmia, en el qual la majoria d'aspectes més rellevants, com el moment de la irrupció o la ràtio d'infecció, seran desconeguts.

Les dades que analitzarem es van recollir entre gener de 2004 i desembre de 2008, i el nostre objectiu ha estat estudiar el comportament del fenomen de les arribades al Servei d'Urgències i realitzar prediccions. No hem fet servir el dia com unitat de mesura del temps (malgrat que podria semblar una tria natural) perquè aquest enfocament hagués introduït un excés de zeros a les dades. A més, en un sentit més pràctic, la setmana és la unitat de temps utilitzada habitualment pels sistemes de vigilància epidemiològica de la grip per proporcionar la informació. En aquest treball s'ha analitzat el nombre d'arribades al Servei d'Urgències utilitzant tècniques de sèries temporals discretes. Molts models de sèries temporals discretes han estat considerats en la literatura (veure [32]), encara que nosaltres ens hem centrat en models enters autoregressius (INAR), que són una extensió natural dels ben coneguts models autoregressius (AR), i acostumen a ser fàcilment interpretables en els contextos pràctics.

Algunes malalties, com la grip, tenen una major incidència en determinats moments de l'any. Aquest comportament estacional no és cobert pels models INAR(k) clàssics, definits per

$$X_t = p_1 \circ X_{t-1} + \dots + p_k \circ X_{t-k} + W_t,$$

on  $p_1, \dots, p_k$  són paràmetres fixats complint  $0 < p_1, \dots, p_k < 1$  i s'assumeix que  $W_t$  segueix una distribució de Poisson de mitjana fixada  $\lambda$ . A més, també

## 2.1. Introducció

---

es demana que  $X_{t-i}$ ,  $i = 1, \dots, k$  i  $W_t$  siguin independents per tot temps  $t$ . Recordant la definició de *binomial thinning*, introduïda a (1.7),

$$p_j \circ X_{t-j} = \sum_{i=1}^{X_{t-j}} Y_i,$$

on  $Y_i$  són variables aleatòries de Bernoulli independents i idènticament distribuïdes amb probabilitat d'èxit igual a  $p_j$ , podem pensar que en el nostre context,  $X_t$  representa el nombre d'arribades al servei d'Urgències a temps  $t$ . La variable aleatòria de Bernoulli  $Y_i$ ,  $i = 1, \dots, x_{t-j}$ , és igual a 1 si l'individu  $i$ , que ja ha accedit al Servei a temps  $t-j$ , hi torna en el moment  $t$ , i igual a 0 en cas contrari. Com s'ha comentat, no hi ha una manera única d'entendre els paràmetres que apareixen en un model INAR(k). Una possible interpretació és pensar  $p_j$  com la proporció d'observacions aparegudes en temps  $t-j$  que continuen al temps  $t$ . Segons quin sigui el context, es pot interpretar en un sentit de ràtio de supervivència. D'acord amb la nostra situació específica, els paràmetres  $p_j$  de (1.6) es podrien interpretar com la proporció de pacients amb símptomes de grip que van anar al Servei d'Urgències en temps  $t-j$  i van tenir una recaiguda  $j$  setmanes després. Els recomptes  $W_t$  podrien interpretar-se com les innovacions o els nous casos de grip, pacients que arriben per primera vegada al Servei d'Urgències en temps  $t$ .

En la propera secció es planteja un model que té en consideració l'estacionalitat observada. La secció 3 està dedicada a l'estimació dels paràmetres del model pel mètode de màxima versemblança, i a la secció 4 es presentaran alguns mètodes

per realitzar prediccions basades en el model introduït.

## 2.2 Definició del model

Per tal d'incloure l'impacte estacional en el model, comencem explorant la variació següent dels models INAR(1) i INAR(2) definits a (1.6):

**Definició 2.2.1** *Sigui  $X_t$  una sèrie temporal discreta. Direm que  $X_t$  segueix un model INAR amb estacionalitat d'ordre 1 i 2 respectivament si satisfà*

$$X_t = p_1 \circ X_{t-1} + W_t(\lambda_t), \quad (2.1)$$

$$X_t = p_1 \circ X_{t-1} + p_2 \circ X_{t-2} + W_t(\lambda_t). \quad (2.2)$$

on els paràmetres  $p_1$  i  $p_2$  són com a (1.6) però ara assumim que les innovacions  $W_t$  segueixen distribucions de Poisson amb diferents mitjanes de manera que  $\lambda_t = \lambda_{t+\tau}$ , on  $\tau$  representa la periodicitat observada.

De la mateixa manera que a (1.6), a (2.2.1) també demanem que  $X_{t-1}$  i  $W_t$  per una banda i  $X_{t-2}$  i  $W_t$  per una altra, siguin independents. Aquest model INAR(1) és un cas particular del model presentat a [34]. El model INAR(2) fou introduït a [35]. De fet, la metodologia que s'introduirà en les properes seccions és directament generalitzable a models INAR(k) d'ordre superior.

Una altra manera d'introduir la periodicitat en aquest tipus de models es planteja a [47], on els autors extenen un model INAR(1) afegint una segona sèrie independent no estacionària amb mitjana variable. Un altre mètode és el descrit

## 2.2. Definició del model

---

a [22], on es considera que les observacions segueixen una distribució de Poisson amb mitjanes que poden seguir una sèrie temporal no estacionària que inclou l'estacionalitat.

Després d'una aproximació exploratòria basada en una anàlisi de regressió lineal s'ha estimat una estacionalitat o periodicitat de 12 mesos ( $\tau = 12$ ) en les nostres dades. Per a aquesta anàlisi en particular hem agrupat les dades per mesos (figura 2.1) per tal de simplificar els càlculs. La suma de les arribades al Servei d'Urgències per setmana  $X_t$  en un mes és denotada per  $Z_t$ . Així, s'ha considerat un model trigonomètric definit per

$$Z_t = \alpha + \beta \sin\left(\frac{2\pi t}{\tau}\right) + \gamma \cos\left(\frac{2\pi t}{\tau}\right), \quad (2.3)$$

amb  $\tau$  variant desde 1 fins a 24, trobant que el millor ajust ( $R^2$  més alt) correspon a  $\tau = 12$  ( $R^2 = 0,3175$ ). Com podem veure a la figura 2.1, el model de regressió trigonomètric és lluny d'ajustar bé les dades reals, encara que captura correctament la seva estacionalitat. L'estimació basada en el model de la figura 2.1 correspon als valors predits descrits a la propera secció, considerant la mitjana mensual.

En realitat, la nostra unitat de temps és la setmana, i per tant, si utilitzéssim un model complet, hauríem de considerar tantes mitjanes  $\lambda_t$  diferents per a les innovacions com setmanes hi hagués a l'any. Com que això implicaria estimar un nombre excessivament gran de paràmetres, s'ha fet servir una reducció mensual, fent que les mitjanes  $\lambda_t$  siguin iguals per qualsevol setmana dins del mateix mes. Per tant, tindrem només 12  $\lambda_t$  diferents, el que significa un total de 13 paràmetres



a estimar en el cas del model INAR(1) i 14 en el cas del model INAR(2), això és,  $\lambda_1, \dots, \lambda_{12}, p_1, p_2$ .

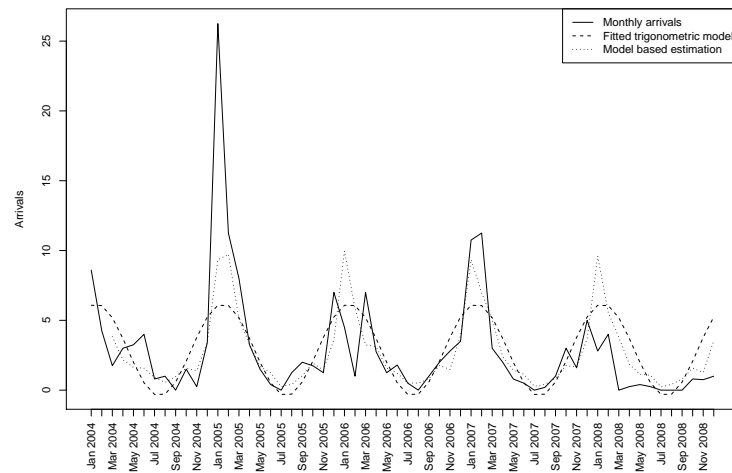


Figura 2.1: Arribades al Servei d'Urgències amb símptomes de grip. Dades observades, model trigonomètric ajustat i mitjanes estimades en base al model.

Les nostres dades s'han ajustat utilitzant els models definits a (2.1) i (2.2) però només el model INAR(2) (2.2) ha estat satisfactori, fet que es comentarà amb més profunditat a l'apartat 2.3.1. Per tant, en les properes seccions centrarem l'atenció només en aquest model. De tota manera, totes les tècniques desenvolupades aquí poden ser adaptades de forma senzilla al model (2.1) o fins i tot a models INAR d'ordres més alts.

## 2.3 Estimació dels paràmetres i validació del model

Hem estimat els paràmetres del model (2.2) ,  $\theta = (\lambda_1, \dots, \lambda_{12}, p_1, p_2)$  utilitzant el mètode de màxima versemblança condicionada. La funció de versemblança condicionada als primers dos valors  $x_1$  i  $x_2$  es pot escriure com

$$L(X; \theta) = P(x_1, x_2, \dots, x_n; \theta) = P(x_1; \theta)P(x_2 | x_1; \theta)P(x_3 | x_1, x_2; \theta) \cdots P(x_n | x_1, \dots, x_{n-1}; \theta) \quad (2.4)$$

Com que només considerem dependències d'ordre 1 i 2, podem escriure l'expressió anterior com

$$L(X; \theta) \sim P(x_3 | x_1, x_2; \theta) \cdot P(x_4 | x_2, x_3; \theta) \cdots P(x_n | x_{n-1}, x_{n-2}; \theta), \quad (2.5)$$

Aquí  $P(x_i | x_{i-1}, x_{i-2}; \theta)$  indica la funció de probabilitat de  $X_i$  condicionada a les dues observacions prèvies i avaluada a  $x_i$ . Observant l'expressió del model (2.2) veiem que aquesta distribució de probabilitat es correspon a la suma de dues variables aleatòries binomials independents amb paràmetres  $(x_{i-1}, p_1)$  i  $(x_{i-2}, p_2)$  i una variable aleatòria de Poisson amb mitjana  $\lambda_1, \dots, \lambda_{12}$  dependent del mes al qual correspongui l'observació. Per tal de calcular aquesta funció de probabilitat hem fet servir la següent expressió, basada en un resultat de [9]:

**Proposició 2.3.1** *Sigui  $X_t$  una sèrie temporal satisfent (2.2). Aleshores, la funció de probabilitat de  $x_i$  condicionada a les dues observacions prèvies  $x_{i-1}$  i  $x_{i-2}$*

és

$$P(x_i | x_{i-1}, x_{i-2}) = \sum_{r=0}^{x_i} \frac{e^{-\lambda_i} \lambda_i^{x_i-r}}{(x_i-r)!} \sum_{k=0}^l \binom{x_{i-2}}{l-k} p_2^{l-k} (1-p_2)^{x_{i-2}-(l-k)} \binom{x_{i-1}}{k} p_1^k (1-p_1)^{x_{i-1}-k} \quad (2.6)$$

**Prova.** La distribució de probabilitat que volem calcular es correspon a la suma de dues variables aleatòries binomials independents amb paràmetres  $(x_{i-1}, p_1)$  i  $(x_{i-2}, p_2)$  i una variable aleatòria de Poisson amb mitjana  $\lambda_1, \dots, \lambda_{12}$  dependent del mes al qual correspongui l'observació. Suposem que  $X_1 \sim Bin(x_{i-1}, p_1)$  i que  $X_2 \sim Bin(x_{i-2}, p_2)$ . Aleshores, tindrem que

$$\begin{aligned} P(X_1 + X_2 = l) &= \sum_{k=0}^l P(X_2 = l - k | X_1 = k) P(X_1 = k) = \\ &= \binom{x_{i-2}}{l-k} p_2^{l-k} (1-p_2)^{x_{i-2}-(l-k)} \binom{x_{i-1}}{k} p_1^k (1-p_1)^{x_{i-1}-k} \end{aligned}$$

Considerem ara una nova variable  $Y \sim Pois(\lambda_i)$ , i tindrem que

$$\begin{aligned} P(X_1 + X_2 + Y = x_i) &= \sum_{r=0}^{x_i} P(Y = x_i - r | X_1 + X_2 = r) P(X_1 + X_2 = r) = \\ &= \sum_{r=0}^{x_i} \frac{e^{-\lambda_i} \lambda_i^{x_i-r}}{(x_i-r)!} \sum_{k=0}^l \binom{x_{i-2}}{l-k} p_2^{l-k} (1-p_2)^{x_{i-2}-(l-k)} \\ &\quad \binom{x_{i-1}}{k} p_1^k (1-p_1)^{x_{i-1}-k} \end{aligned}$$

■

L'expressió (2.6) es pot generalitzar a models d'ordre més alt, com fan en el mateix treball [9]. La maximització de (2.5) s'ha realitzat amb un programa desenvolupat en **R** ([38]) fent servir la rutina de minimització no lineal *nlm*. Els

### 2.3. Estimació dels paràmetres i validació del model

errors estàndards dels estimadors dels paràmetres s'han calculat a partir de la matriu Hessiana corresponent. Els estimadors obtinguts es mostren a la taula 2.1.

Taula 2.1: Estimadors de màxima versemblança

Paràmetre	$\lambda_1$	$\lambda_2$	$\lambda_3$	$\lambda_4$	$\lambda_5$	$\lambda_6$	$\lambda_7$	$\lambda_8$	$\lambda_9$	$\lambda_{10}$	$\lambda_{11}$	$\lambda_{12}$	$p_1$	$p_2$
Estimador	6,90	3,44	2,50	1,56	0,83	0,61	0,14	0,32	0,61	1,37	0,78	2,68	0,28	0,18
Error Est.	0,66	0,53	0,40	0,31	0,23	0,21	0,09	0,13	0,19	0,27	0,23	0,38	0,03	0,02

Es pot observar que les innovacions de mitjana més alta corresponen als mesos de desembre, gener i febrer ( $\hat{\lambda}_{12}$ ,  $\hat{\lambda}_1$  and  $\hat{\lambda}_2$ ). De fet,  $\hat{p}_1 > \hat{p}_2$ , indicant que l'impacte del valor corresponent a la setmana anterior és més gran que l'impacte del valor corresponent a dues setmanes anteriors, amb una diferència al voltant del 10%. Podem veure també que tots els paràmetres són significativament diferents de zero exepete  $\lambda_7$ , que correspon al mes de juliol.

La matriu de variància-covariància estimada, calculada com la inversa de la matriu Hessiana, és la següent:

$$\hat{\Sigma} = \begin{pmatrix} 0.4344 & 0.0082 & 0.0049 & 0.0019 & 0.0014 & 0.0017 & 0.0002 & 0.0003 & 0.0008 & 0.0010 & 0.0020 & 0.0026 & -0.0016 & -0.0007 \\ 0.0082 & 0.2783 & 0.0075 & 0.0029 & 0.0021 & 0.0027 & 0.0003 & 0.0004 & 0.0014 & 0.0015 & 0.0031 & 0.0041 & -0.0024 & -0.0013 \\ 0.0049 & 0.0075 & 0.1617 & 0.0018 & 0.0012 & 0.0015 & 0.0002 & 0.0002 & 0.0007 & 0.0009 & 0.0017 & 0.0024 & -0.0017 & -0.0004 \\ 0.0019 & 0.0029 & 0.0018 & 0.0939 & 0.0005 & 0.0006 & 0.0001 & 0.0001 & 0.0003 & 0.0003 & 0.0007 & 0.0009 & -0.0006 & -0.0002 \\ 0.0014 & 0.0021 & 0.0012 & 0.0005 & 0.0507 & 0.0005 & 0.0001 & 0.0001 & 0.0002 & 0.0002 & 0.0005 & 0.0007 & -0.0003 & -0.0003 \\ 0.0017 & 0.0027 & 0.0015 & 0.0006 & 0.0005 & 0.0442 & 0.0001 & 0.0001 & 0.0003 & 0.0003 & 0.0007 & 0.0009 & -0.0005 & -0.0003 \\ 0.0002 & 0.0003 & 0.0002 & 0.0001 & 0.0001 & 0.0001 & 0.0081 & 0.0000 & 0.0000 & 0.0000 & 0.0001 & 0.0001 & 0.0000 & 0.0000 \\ 0.0003 & 0.0004 & 0.0002 & 0.0001 & 0.0001 & 0.0001 & 0.0000 & 0.0166 & 0.0000 & 0.0000 & 0.0001 & 0.0001 & -0.0001 & -0.0001 \\ 0.0008 & 0.0014 & 0.0007 & 0.0003 & 0.0002 & 0.0003 & 0.0000 & 0.0000 & 0.0357 & 0.0002 & 0.0004 & 0.0004 & -0.0002 & -0.0002 \\ 0.0010 & 0.0015 & 0.0009 & 0.0003 & 0.0002 & 0.0003 & 0.0000 & 0.0000 & 0.0002 & 0.0745 & 0.0004 & 0.0005 & -0.0003 & -0.0001 \\ 0.0020 & 0.0031 & 0.0017 & 0.0007 & 0.0005 & 0.0007 & 0.0001 & 0.0001 & 0.0004 & 0.0004 & 0.0544 & 0.0010 & -0.0005 & -0.0004 \\ 0.0026 & 0.0041 & 0.0024 & 0.0009 & 0.0007 & 0.0009 & 0.0001 & 0.0001 & 0.0004 & 0.0005 & 0.0010 & 0.1426 & -0.0008 & -0.0004 \\ -0.0016 & -0.0024 & -0.0017 & -0.0006 & -0.0003 & -0.0005 & 0.0000 & -0.0001 & -0.0002 & -0.0003 & -0.0005 & -0.0008 & 0.0007 & 0.0000 \\ -0.0007 & -0.0013 & -0.0004 & -0.0002 & -0.0003 & -0.0003 & 0.0000 & -0.0001 & -0.0002 & -0.0001 & -0.0004 & -0.0004 & 0.0000 & 0.0004 \end{pmatrix}$$

### 2.3.1 Validació del model

Les arribades al Servei d'Urgències a temps  $t$  es poden estimar en base a l'esperança condicionada a la darrera informació coneguda, això és,  $E(X_t | X_{t-1} = x_{t-1}, X_{t-2} = x_{t-2}) = p_1 x_{t-1} + p_2 x_{t-2} + \lambda_t$ . Això porta al següent estimador o *valor predit* de  $X_t$ :

**Definició 2.3.2** *Podem construir un estimador de  $X_t$  condicionant l'esperança a la darrera informació coneguda, obtenint*

$$\hat{X}_t = \hat{p}_1 x_{t-1} + \hat{p}_2 x_{t-2} + \hat{\lambda}_t. \quad (2.8)$$

Com és habitual, aquesta metodologia es pot fer servir per validar fins a quin punt el model s'ajusta a les dades reals. La variància de  $\hat{X}_t$  es pot calcular directament de l'expressió (2.8), obtenint

$$\begin{aligned} V(\hat{X}_t) = & V(\hat{p}_1)x_{t-1}^2 + V(\hat{p}_2)x_{t-2}^2 + V(\hat{\lambda}_t) + 2x_{t-1}x_{t-2}COV(\hat{p}_1, \hat{p}_2) + \\ & + 2x_{t-1}COV(\hat{p}_1, \hat{\lambda}_t) + 2x_{t-2}COV(\hat{p}_2, \hat{\lambda}_t). \end{aligned} \quad (2.9)$$

Per tal d'estimar  $V(\hat{X}_t)$ , les variàncies i covariàncies que apareixen a l'expressió (2.9) es poden substituir pels seus estimadors corresponents, mostrats a  $\hat{\Sigma}$ . A partir d'aquí, es poden construir intervals de confiança  $100(1 - \alpha)\%$  aproximats de la manera habitual,  $X_t = \hat{X}_t \pm z_{\alpha/2} \sqrt{\hat{V}(\hat{X}_t)}$ . La precisió d'aquests intervals, basats en propietats asimptòtiques de l'estimació per màxima versemblança, depen essencialment de la longitud de la sèrie. En general, la precisió creix amb la longitud de la sèrie.

### 2.3.1. Validació del model

---

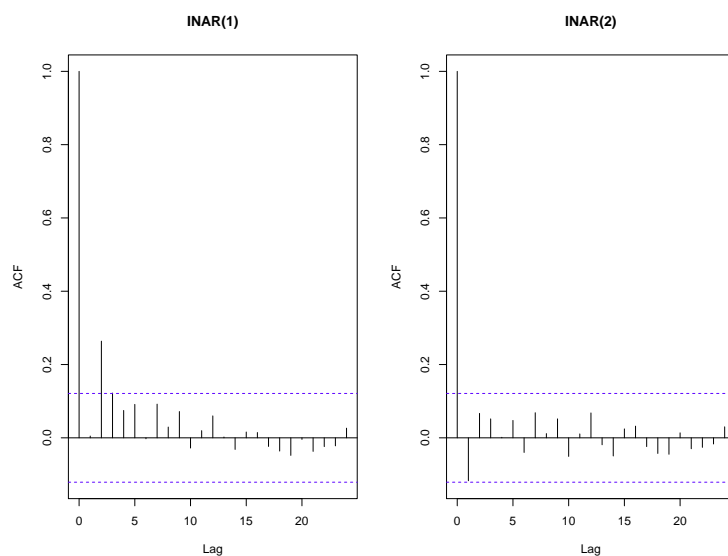


Figura 2.2: Funcions d'autocorrelació dels residus corresponents als models d'ordre 1 i 2

El residu a temps  $t$  es defineix com  $r_t = x_t - \hat{X}_t$ . Els perfils de les funcions d'autocorrelació (ACF) i d'autocorrelació parcial (PACF) que es mostra a les figures 2.2 i 2.3 suggereixen que hi ha correlacions d'ordre més alt que 1, però no correlacions d'ordre més alt que 2 en les nostres dades que calgui tenir en compte, i per tant, recolza la validació i justificació del model definit a (2.2).

Aquest procediment ha estat utilitzat també per comprovar la bondat d'ajust del model INAR(1) definit a (2.1), i les gràfiques de l'ACF i PACF dels residus no van ser tan bones, justificant l'ús del model INAR(2).

Donades  $n$  observacions  $x_1, \dots, x_n$  de les nostres arribades, l'expressió (2.8) es pot utilitzar també per realitzar prediccions puntuals a temps  $n + 1$ . Per un temps  $n + k$ ,  $k > 1$ , estimadors puntuals similars, encara que més complicats,

poden construir-se també de forma directa. En qualsevol cas, com s'estudiarà en la propera secció, hem preferit considerar mitjanes a llarg termini i regions de predicció amb l'objectiu de realitzar previsions.

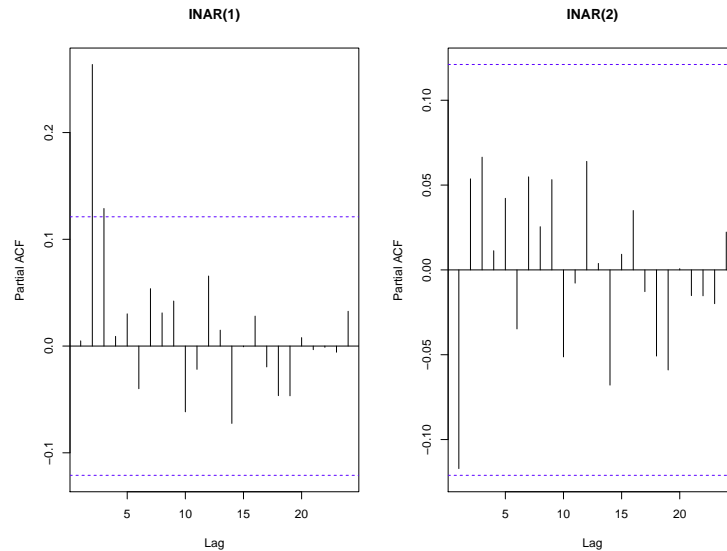


Figura 2.3: Funció d'autocorrelació parcial dels residus corresponents als models d'ordre 1 i 2

Altres mètodes per avaluar la bondat de l'ajust d'aquest model, basats en la previsió sobre un període de temps en el qual es disposa d'observacions reals, es consideraran a la propera secció.

## 2.4 Prediccions

Amb el propòsit de realitzar prediccions basades en el model definit a (2.2), considerarem aquí un any amb el mateix nombre de setmanes (4) a cada mes, és

### 2.4.1. Mitjanes a llarg termini

---

a dir, un any amb un total de 48 setmanes. El nombre de setmanes de cada mes varia segons l'any, per tant, hem triat aquest enfocament per raons pràctiques. En qualsevol cas, el mètode és també vàlid si hom volgués realitzar previsions sobre un any concret, tenint en compte la seva estructura temporal, afegint setmanes amb els paràmetres  $\lambda_j$  apropiats en els mesos corresponents.

### 2.4.1 Mitjanes a llarg termini

**Proposició 2.4.1** *Sigui  $E(X_t) = \mu_t$ . Aleshores,  $\mu_t$  convergeix asimptòticament cap a  $P_t$ , on  $P_t$  és qualsevol solució particular de l'equació en diferències*

$$\mu_t = p_1\mu_{t-1} + p_2\mu_{t-2} + \lambda_t, \quad (2.10)$$

*si els valors absoluts d'ambdues arrels del polinomi característic (o els seus mòduls si són complexos) són inferiors a 1.*

**Prova.** Prenent esperances a (2.2) obtenim la relació de recurrència definida a (2.10), que és una equació en diferències lineal no homogènia de segon ordre. És ben conegut que la solució és de la forma  $\mu_t = H_t + P_t$ , on  $H_t$  és la solució general de l'equació homogènia corresponent,

$$\mu_t = p_1\mu_{t-1} + p_2\mu_{t-2}, \quad (2.11)$$

i  $P_t$  és qualsevol solució particular de l'equació no homogènia.

La solució general de l'equació homogènia depen de les arrels de l'equació carac-



terística,  $y^2 - p_1y - p_2 = 0$ , és a dir,  $y_1$  i  $y_2$ . Aquesta solució és de la forma,

$$H_t = A(y_1)^t + B(y_2)^t, \quad (2.12)$$

on  $A$  i  $B$  són constants que es poden obtenir tenint en compte els valors inicials. Si els valor absoluts d'ambdues arrels (o els seus mòduls si són complexos) són inferiors a 1, aleshores  $H_t$  tendeix a zero quan  $t$  va cap a infinit. Per tant, en aquesta situació,  $\mu_t \sim P_t$  per  $t$  prou gran ■

Per tal d'estimar els valors asimptòtics de  $\mu_t$ , és a dir,  $P_t$ , podem substituir a (2.10) els paràmetres pels seus valors estimats. Aleshores obtenim l'equació en diferències

$$\hat{\mu}_t = \hat{p}_1\hat{\mu}_{t-1} + \hat{p}_2\hat{\mu}_{t-2} + \hat{\lambda}_t. \quad (2.13)$$

És immediat comprovar que els valors absoluts de les arrels de l'equació característica són inferiors a 1 ( $y_1 = 0,59$  i  $y_2 = -0,31$ ), de manera que  $\hat{\mu}_t \sim \hat{P}_t$  per  $t$  prou gran. A més, com que  $\hat{\lambda}_t$  és periòdic amb període de 48 setmanes, el mateix serà cert per a la solució particular  $\hat{P}_t$ .

**Observació 2.4.2** *La proposició anterior ens permet parlar d'una certa estabilitat asimptòtica, que anomenarem pseudo-estacionarietat.*

Aquesta solució particular  $\hat{P}_t$  es podria calcular explícitament obtenint una expressió complicada, encara que des d'un punt de vista pràctic és més senzill obtenir-la simplement iterant (2.13) fins que els valors de  $\hat{\mu}_t$  s'estabilitzen en un cicle de període 48. La taula 2.2 mostra les estimacions obtingudes per aques-

## 2.4.2. Estimació de la variància

---

tes mitjanes a llarg termini. Cal observar que si els valors absoluts de les arrels de l'equació característica no fossin menors que 1, aleshores la solució de l'equació homogènia definida a  $H_t$  a (2.12) no és necessàriament fitada quan el temps tendeix a infinit, de manera que no podríem garantir aquest tipus de pseudo-estacionarietat. En tot cas, és clar que sèries temporals amb mitjanes no fitades no resultarien adequades per modelar ingressos hospitalaris.

Taula 2.2: Mitjanes a llarg termini estimades per un any (48 setmanes)

Setmana	Estimació (DE)	Setmana	Estimació (DE)	Setmana	Estimació (DE)
1	9,00 (0,73)	17	2,30 (0,37)	33	0,87 (0,21)
2	10,27 (0,90)	18	2,05 (0,36)	34	0,96 (0,25)
3	11,43 (1,07)	19	1,82 (0,38)	35	1,04 (0,29)
4	11,99 (1,17)	20	1,71 (0,39)	36	1,07 (0,32)
5	8,90 (0,84)	21	1,42 (0,28)	37	1,86 (0,31)
6	8,14 (0,85)	22	1,32 (0,30)	38	2,09 (0,37)
7	7,35 (0,89)	23	1,24 (0,33)	39	2,30 (0,43)
8	6,99 (0,92)	24	1,20 (0,35)	40	2,40 (0,46)
9	5,82 (0,64)	25	0,70 (0,19)	41	1,88 (0,32)
10	5,42 (0,64)	26	0,55 (0,17)	42	1,75 (0,34)
11	5,09 (0,68)	27	0,42 (0,16)	43	1,62 (0,37)
12	4,93 (0,70)	28	0,35 (0,16)	44	1,56 (0,39)
13	3,87 (0,49)	29	0,49 (0,15)	45	3,41 (0,42)
14	3,55 (0,49)	30	0,52 (0,17)	46	3,92 (0,50)
15	3,26 (0,52)	31	0,55 (0,20)	47	4,40 (0,60)
16	3,12 (0,54)	32	0,57 (0,22)	48	4,63 (0,65)

## 2.4.2 Estimació de la variància

Hem aplicat el *mètode delta* descrit a la secció 0.1.5 per tal d'estimar la variància dels estimadors de les mitjanes a llarg termini que s'acaben de presen-

tar. Per a fer-ho, cal tenir en compte que cada  $P_t = P_t(p_1, p_2, \lambda_1, \dots, \lambda_{12})$ ,  $t = 1, \dots, 48$  és una funció dels paràmetres i és necessari calcular els gradients  $\nabla P_t = (\frac{\partial P_t}{\partial p_1}, \frac{\partial P_t}{\partial p_2}, \frac{\partial P_t}{\partial \lambda_1}, \dots, \frac{\partial P_t}{\partial \lambda_{12}})$ ,  $t = 1, \dots, 48$ . A partir d'aquí, la variància de cada  $\hat{P}_t$  pot ser estimada de la manera habitual, avaluant els gradients en els paràmetres estimats i la matriu de variància-covariància  $\hat{\Sigma}$ , a través de l'expressió  $\hat{V}(\hat{P}_t) = \nabla \hat{P}_t^T \cdot \hat{\Sigma} \cdot \nabla \hat{P}_t$ ,  $t = 1, \dots, 48$ . Els gradients han estat calculats fent servir una aproximació de primer ordre a les derivades parcials:

$$\begin{aligned} \nabla \hat{P}_t \approx \frac{1}{h} & (P_t(\hat{p}_1 + h, \hat{p}_2, \hat{\lambda}_1, \dots, \hat{\lambda}_{12}) - \hat{P}_t, P_t(\hat{p}_1, \hat{p}_2 + h, \hat{\lambda}_1, \dots, \hat{\lambda}_{12}) - \\ & \hat{P}_t, P_t(\hat{p}_1, \hat{p}_2, \hat{\lambda}_1 + h, \dots, \hat{\lambda}_{12}) - \hat{P}_t, \dots, P_t(\hat{p}_1, \hat{p}_2, \hat{\lambda}_1, \dots, \hat{\lambda}_{12} + h) - \hat{P}_t), t = 1, \dots, 48, \end{aligned} \quad (2.14)$$

on  $\hat{P}_t = P_t(\hat{p}_1, \hat{p}_2, \hat{\lambda}_1, \dots, \hat{\lambda}_{12})$ , i  $h$  és un valor petit. Per les nostres dades, després d'algunes proves hem triat  $h = 0.001$ , valor per al qual s'obté una estabilitat satisfactòria dels resultats numèrics. Totes les components dels gradients es poden calcular de la mateixa manera que  $\hat{P}_t$ , és a dir, iterant (2.13) fins que els valors de  $\hat{\mu}_t$  s'estabilitzen en un cicle. Per exemple, per calcular  $P_t(\hat{p}_1, \hat{p}_2 + h, \hat{\lambda}_1, \dots, \hat{\lambda}_{12})$ , hem de substituir  $\hat{p}_2$  per  $\hat{p}_2 + h$  a (2.13) i iterar fins que  $\hat{\mu}_t$  s'estabilitza.

A la figura 2.4 podem veure les mitjanes a llarg termini per un any, amb els intervals de confiança del 95% corresponents. Fent servir l'expressió  $P_t = \hat{P}_t \pm z_{\alpha/2} \sqrt{\hat{V}(\hat{P}_t)}$ ,  $t = 1, \dots, 48$  podríem construir intervals de confiança aproximats de nivell  $100(1 - \alpha)$ . Tal com s'ha comentat a la secció 2.3.1, la precisió d'aquest tipus d'intervals augmenta amb la mida de la sèrie temporal original.

Les mitjanes a llarg termini són interessants per descriure el comportament de les arribades al Servei d'Urgències d'una manera estretament lligada al compor-

### 2.4.3. Regions de predicció a curt termini

---

tament de la grip. En qualsevol cas, l'interès principal a la pràctica serà saber amb antelació quants llits seran necessaris les properes setmanes. L'objectiu de la següent secció serà introduir algunes tècniques per abordar aquesta qüestió.

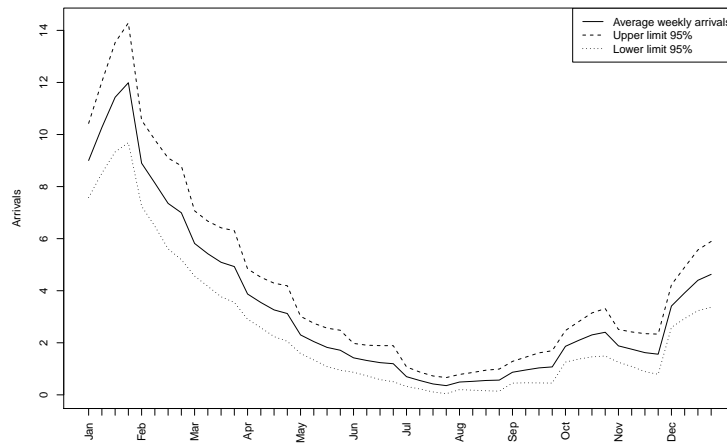


Figura 2.4: Arribades setmanals mitjanes a llarg termini estimades per un any

### 2.4.3 Regions de predicció a curt termini

Donades  $n$  observacions  $x_1, \dots, x_n$  d'arribades al Servei d'Urgències, voldrem conèixer la distribució de  $X_{n+k}$ , això és, la distribució de les arribades  $k$  setmanes en el futur condicionada als valors previs. Per exemple, suposem que  $k = 1$ . Tenint en compte que  $X_{n+1}$ , condicionat a  $X_n = x_n$  i  $X_{n-1} = x_{n-1}$ , és la suma de dues variables binomials independents amb paràmetres  $(x_n, p_1)$  i  $(x_{n-1}, p_2)$ , i una variable independent amb distribució de Poisson amb mitjana  $\lambda_{n+1}$ , podem obtenir la distribució de  $X_{n+1}$  fent servir (2.6).

El càlcul de la distribució de  $X_{n+2}$  és més complicat. El primer pas serà expressar  $X_{n+2}$  en termes dels darrers valors coneguts  $X_n$  i  $X_{n-1}$ , gràcies a les propietats de l'operador *binomial thinning* descrites a la secció 1.2.1 i fent servir una mica d'àlgebra:

$$\begin{aligned} X_{n+2} &= p_1 \circ X_{n+1} + p_2 \circ X_n + W_{n+2} = p_1 \circ (p_1 \circ X_n + p_2 \circ X_{n-1} + W_{n+1}) + \\ &+ p_2 \circ X_n + W_{n+2} = (p_1^2 + p_2) \circ X_n + p_1 p_2 \circ X_{n-1} + p_1 \circ W_{n+1} + W_{n+2}. \end{aligned} \quad (2.15)$$

Cal observar que  $p_1 \circ W_{n+1} + W_{n+2}$  segueix una distribució de Poisson amb mitjana  $p_1 \lambda_{n+1} + \lambda_{n+2}$ , que denotarem per  $W(\lambda_2^*)$ . Per una altra part, condicionant que  $X_n = x_n$  i  $X_{n-1} = x_{n-1}$ , el que tenim a (2.15) és la suma de dues variables binomials independents amb paràmetres  $(x_n, p_1^2 + p_2)$  i  $(x_{n-1}, p_1 p_2)$ , i una variable de Poisson de mitjana  $p_1 \lambda_{n+1} + \lambda_{n+2}$ . De nou, podem fer servir (2.6) per obtenir la distribució de  $X_{n+2}$ , amb els nous paràmetres  $p_1^2 + p_2$ ,  $p_1 p_2$  i  $p_1 \lambda_{n+1} + \lambda_{n+2}$ .

**Proposició 2.4.3** *La distribució de  $X_{n+k}$  condicionada a les dues darreres observacions  $X_n$  i  $X_{n-1}$ , per qualsevol  $k > 1$ , és la suma de dues variables binomials independents i una variable de Poisson, on els paràmetres  $(x_{n+k-1}, f_k)$ ,  $(x_{n+k-2}, g_k)$  i  $\lambda_k^*$  respectivament vénen donats per les següents relacions de recurrència, que es satisfan per  $k > 1$ :*

$$\begin{aligned} f_{k+1} &= p_1 f_k + p_2 f_{k-1}, \\ g_{k+1} &= p_1 g_k + p_2 g_{k-1}, \\ \lambda_{k+1}^* &= p_1 \lambda_k^* + p_2 \lambda_{k-1}^* + \lambda_{n+k+1}, \end{aligned} \quad (2.16)$$

### 2.4.3. Regions de predicció a curt termini

---

amb  $f_1 = p_1$ ,  $g_1 = p_2$ ,  $\lambda_1^* = \lambda_{n+1}$ ,  $f_2 = p_1^2 + p_2$ ,  $g_2 = p_1p_2$  i  $\lambda_2^* = p_1\lambda_{n+1} + \lambda_{n+2}$

**Prova.** Procedirem per inducció sobre  $k$ . Per  $k = 1$ , és clar que  $X_{n+1} = p_1 \circ X_n + p_2 \circ X_{n-1} + W(\lambda_{n+1}^*)$ . Suposem que la relació de recurrència és certa fins a  $X_{n+k}$ , de manera que podem escriure

$$X_{n+k} = f_k \circ X_n + g_k \circ X_{n-1} + W(\lambda_k^*) \quad (2.17)$$

Tenim que

$$\begin{aligned} X_{n+k+1} &= p_1 \circ X_{n+k} + p_2 \circ X_{n+k-1} + W_{\lambda_{n+k+1}} = \\ &= p_1 \circ (f_k \circ X_n + g_k \circ X_{n-1} + W(\lambda_k^*)) + p_2 \circ (f_{k-1} \circ X_n + g_{k-1} \circ X_{n-1} + W(\lambda_{k-1}^*)) = \\ &= (p_1 f_k) \circ X_n + (p_1 g_k) \circ X_{n-1} + W(p_1 \lambda_k^*) + (p_2 f_{k-1}) \circ X_n + \\ &+ (p_2 g_{k-1}) \circ X_{n-1} + W(p_2 \lambda_{k-1}^*) + W(\lambda_{n+k+1}) \end{aligned}$$

Igualant els termes en l'expressió anterior tenim que

$$\begin{aligned} f_{k+1} &= p_1 f_k + p_2 f_{k-1}, \\ g_{k+1} &= p_1 g_k + p_2 g_{k-1}, \\ \lambda_{k+1}^* &= p_1 \lambda_k^* + p_2 \lambda_{k-1}^* + \lambda_{n+k+1}, \end{aligned} \quad (2.18)$$

com volíem ■

Partint de les relacions de recurrència descrites a (2.18), i fent servir de nou (2.6), la distribució de  $X_{n+k}$ ,  $k > 1$  es pot calcular fàcilment. Substituint els paràmetres de (2.18) pels seus estimadors podem calcular regions de predicció

aproximades de mida  $1 - \alpha$ , trobant els límits superior i inferior  $l_1$  i  $l_2$  que satisfan

$$\sum_{j=0}^{l_1} P(X_{n+k} = j) \approx \alpha/2, \quad \sum_{j=l_1}^{l_2} P(X_{n+k} = j) \approx 1 - \alpha. \quad (2.19)$$

A la figura 2.5 podem veure la regió de predicció aproximada de mida 0.95 per cada setmana de l'any 2009, utilitzant com darrers valors coneguts les dues últimes setmanes de 2008. Per a cada  $X_{n+k}$ , pot resultar també adequat calcular la mediana aproximada. Aquest és el valor  $l$  que satisfà

$$\sum_{j=0}^l P(X_{n+k} = j) \approx 0,50. \quad (2.20)$$

Aquestes medianes es mostren també a la figura 2.5.

És interessant observar que aquest mètode utilitzat per calcular la regió de predicció a curt termini, restringida a un model INAR(1) sense tendència estacional, és simplement el mètode de predicció *k-step-ahead* descrit a [16].

En principi, es podrien predir fins i tot valors més allunyats dels darrers valors coneguts continuant amb aquest procediment. En qualsevol cas, per valors grans de  $k$ , un mètode més senzill per obtenir regions de predicció serà introduït a la propera secció.

El mètode descrit pot ser útil també per tal d'avaluar la bondat de l'ajust del model, aplicant-ho sobre un període de temps amb observacions conegudes. Per

#### 2.4.4. Regions de predicció a llarg termini

---

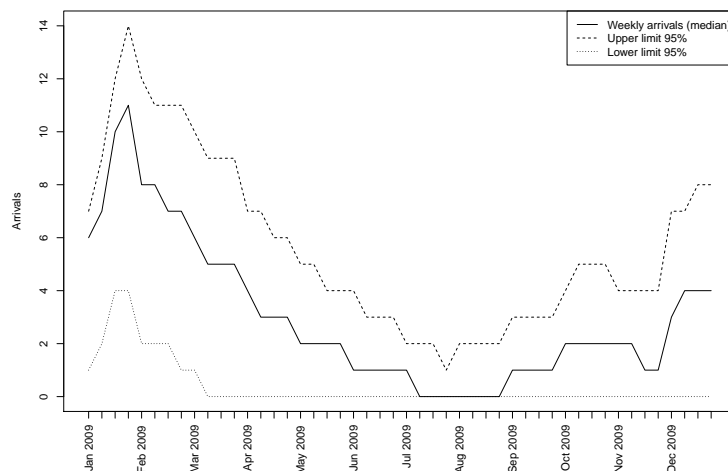


Figura 2.5: Regió de predicció per 2009

exemple, s'ha ajustat el model estimant els paràmetres corresponents utilitzant només les dades de Gener de 2004 fins a Desembre de 2007. Aleshores, utilitzant les arribades al Servei d'Urgències en les dues darreres setmanes de 2007 com a valors inicials, s'han calculat els límits de la regió de predicció a curt termini per 2008. La figura 2.6 mostra que els valors observats reals són dins els límits excepte a l'inici de la sèrie, quan s'ha detectat una major variabilitat.

#### 2.4.4 Regions de predicció a llarg termini

En aquest punt estarem interessats en la distribució de  $X_{n+k}$  quan  $k$  tendeix a infinit. Observem que les dues primeres relacions de recurrència indicades a (2.18) són exactament iguals que a (2.11). Aleshores, l'expressió per  $f_k$  i  $g_k$  es pot obtenir de (2.12), ambdós amb la mateixa estructura però diferents valors



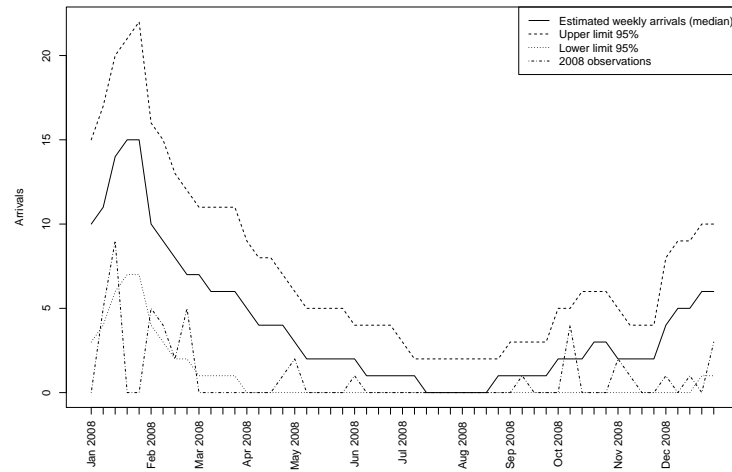


Figura 2.6: Regió de predicció per 2008 i valors observats

per a les constants  $A$  i  $B$ . Això ens porta a enunciar el següent resultat:

**Proposició 2.4.4** *Per valors prou grans de  $k$ ,  $X_{n+k}$  es distribueix asimptòticament com  $W(\lambda_k^*)$ .*

**Prova.** Com que els valors absoluts d'ambdues arrels  $y_1$  i  $y_2$  són menors que 1, podem concloure que  $X_{n+k} \sim W(\lambda_k^*)$ , per valors grans de  $k$ . Per altra banda, la relació de recurrència per  $\lambda_k^*$  que es mostra a (2.18) és exactament la mateixa que apareix a (2.13). En conseqüència, quan  $k$  tendeix a infinit,  $\lambda_k^*$  s'estabilitza en un cicle del mateix període que  $\lambda_t$  ■

Per tant, és clar que la distribució asimptòtica aproximada de  $X_{n+k}$  és un cicle de Poisson amb les mitjanes descrites a la taula 2.2. Les regions de predicció a llarg termini poden ser calculades de la manera descrita en la secció prèvia

#### 2.4.4. Regions de predicció a llarg termini

a (2.19), tenint en compte que ara  $P(X_{n+k})$  segueix una distribució de Poisson amb la mitjana corresponent segons el mes de l'observació futura  $n + k$ . Com abans, podem prendre la mediana de cada  $X_{n+k}$  com el valor central de la regió de predicció. Per tant, és possible expressar (2.17) com

$$X_{n+k} = (A_1(y_1)^k + B_1(y_2)^k) \circ X_n + (A_2(y_1)^k + B_2(y_2)^k) \circ X_{n-1} + W(\lambda_k^*) \quad (2.21)$$

La figura 2.7 mostra la regió de predicció asimptòtica de mida 0,95, i podem apreciar com ràpidament la regió de predicció a curt termini plantejada en la secció anterior convergeix cap a aquesta. De fet, els límits superiors d'ambdues regions de predicció són exactament iguals des de  $k = 9$  i els límits inferiors són iguals des de  $k = 6$ .

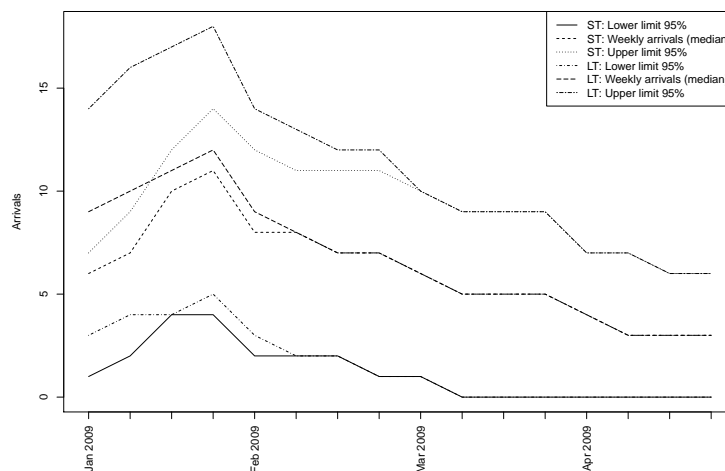


Figura 2.7: Regions de predicció a curt i llarg termini

## 2.5 Conclusions

Diversos autors han utilitzat l'anàlisi convencional de sèries temporals per investigar l'estructura periòdica de la incidència de malalties infeccioses com la grip. A [37], la incidència setmanal de grip a França és estudiada a través de models SARIMA clàssics.

L'anàlisi d'ondetes fou usat a [20] per estudiar la incidència de malalties infeccioses. A [44], els autors utilitzen l'anàlisi espectral de màxima entropia (MEM) per investigar els períodes d'incidència de grip al Japó. Tots aquests mètodes poden resultar apropiats per sèries temporals contínues o bé per sèries de recomptes que prenguin valors grans. En qualsevol cas, per sèries temporals discretes amb valors baixos, els models INAR presenten un rendiment superior.

Hem vist que els models INAR que es presenten en aquest treball són una eina útil per a realitzar prediccions de l'ocupació de llits hospitalaris, basades en les arribades al Servei d'Urgències. Malgrat això, quan fem front a malalties que presenten una important component estacional és necessari incloure aquest impacte en els models. En el cas particular del nostre exemple, hem comprovat que és suficient considerar correlacions de segon ordre, és a dir, un model INAR(2), d'acord amb les funcions d'autocorrelació i autocorrelació parcial que es poden veure a les figures 2.2 i 2.3. En qualsevol cas, tots els mètodes que s'han presentat poden ser directament extesos a models INAR( $p$ ) amb  $p > 2$ .

S'han proposat diversos mètodes de previsió a la secció 2.4. L'estabilitat assimp-tòtica del model s'ha utilitzat per tal d'obtenir una predicció completa per un any arbitrari, sense necessitat de fer servir les darreres observacions disponibles.

## 2.5. Conclusions

---

Aquesta metodologia ens porta als mètodes a llarg termini descrits a les seccions 2.4.1 i 2.4.4. Les regions de predicció proposades a les seccions 2.4.3 i 2.4.4 s'ocupen de donar prediccions del nombre de persones que arribaran al Servei d'Urgències per setmana amb un cert nivell de confiança. Ambdós mètodes són útils per conèixer el nombre de pacients que l'hospital hauria d'esperar en un determinat moment  $n + k$ , coneixent la situació en moments previs de 1 a  $n$ . La regió de predicció a llarg termini comentada a la secció 2.4.4 es pot interpretar com el límit de la regió descrita a 2.4.3 quan  $k$  tendeix a infinit. Altres mètodes per obtenir previsions han estat proposats a la literatura: Per exemple, a [10] els autors mostren una interessant aproximació basada en cadenes de Markov. En qualsevol cas, la metodologia proposada aquí és generalitzable a models d'ordre més alt de forma relativament senzilla. Per altra banda, aquests mètodes suposen només uns pocs minuts de computació per fer tots els càlculs necessaris, per un ordinador personal amb les prestacions habituals.

Enfrontar-se a una epidèmia de la millor manera possible implica la necessitat de fer ús d'un considerable nombre de recursos que en circumstàncies normals no es troben disponibles de manera immediata. En aquesta situació, el sistema de Salut ha de ser capaç de dissenyar estratègies de forma ràpida que permetin solucionar els problemes que l'epidèmia provocarà: Accelerar els processos de descàrrega del sistema sanitari, convertir habitacions individuals en dobles, evitar hospitalitzacions innecessàries, utilitzar àrees comunes com habitacions de tractament o fins i tot tenir disponible un equip de voluntaris entrenats i mobilitzables, com es suggereix a [21]. I es disposa de poc marge de maniobra, ja que el moment en

què apareixerà l'epidèmia és desconegut. El desenvolupament d'eines que permetin preveure, fins i tot de manera aproximada, el moment d'inici d'una epidèmia és crucial per tal d'assegurar l'efectivitat i eficiència del sistema. El llindar que marca l'inici del període epidèmic podria definir-se com el límit superior de la regió de predicció calculada a partir del model ajustat utilitzant les observacions del període no epidèmic previ. En aquest sentit, els mètodes descrits en aquest treball poden ser útils per verificar que en 2009 va aparèixer una nova soca de grip A (H1N1), amb un comportament marcadament diferent del de la grip estacional, com es pot observar a la figura 2.8. Al 2009, la grip estacional es va combinar amb aquesta nova soca i el registre del Servei d'Urgències va quedar notablement distorsionat. La grip és una malaltia amb un comportament molt difícil de predir.

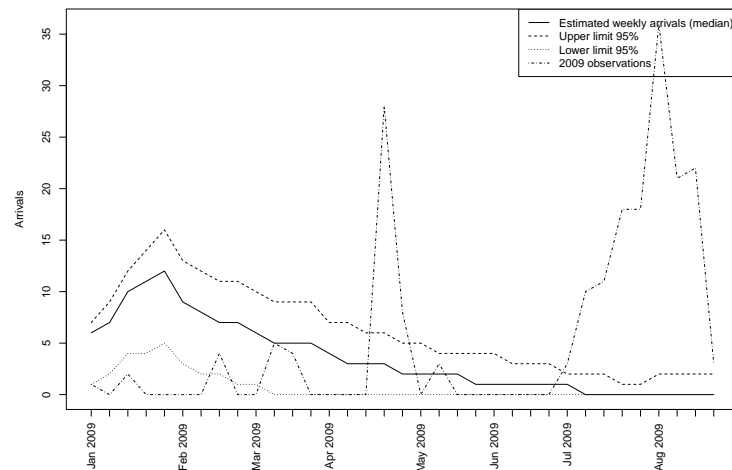


Figura 2.8: Regió de predicció per 2009 i valors observats

Es pot estendre a diferent ritme i velocitat localment, i observem grans diferències d'un lloc a l'altre. Per tant, les prediccions locals, basades en un termini curt de

temps, són de gran interès per tal de planificar de la millor manera possible l'ús dels recursos del sistema de Salut (llits d'hospital, llits ICU, consultes externes i capacitat de reacció) en la situació d'una epidèmia de grip. En qualsevol cas, hem vist que les regions de predicció a curt i llarg termini coincideixen per valors de  $k > 9$ . Per aquest motiu, les regions de predicció a llarg termini són altament recomanables ja que l'aproximació de Poisson és molt senzilla d'implementar.

Molts models epidemiològics s'han desenvolupat seguint l'objectiu d'analitzar el progrés de la majoria de malalties infeccioses. Avui en dia, l'increment en la potència computacional permet el desenvolupament de models basats en individu per poblacions relativament grans, com es proposa a [24]. Tenint en compte les interaccions entre cada persona i l'estructura geogràfica, aquests models podrien explicar acuradament el progrés d'un brot. Un bon model matemàtic per descriure la dinàmica de la grip hauria de ser capaç d'explicar l'estructura correlacional de dues setmanes que hem detectat en el nostre model INAR estacional. Probablement, l'interval de dues setmanes que mostra el nostre treball està relacionat amb paràmetres biològics bàsics del virus de la grip. Aquest fet podria ser investigat desenvolupant models basats en individu adequats.

Altres malalties amb un comportament estacional podrien analitzar-se amb els mètodes proposats en aquest treball, potser amb un ajust de l'ordre del model INAR corresponent. Aquest és el cas de malalties com el dengue, malària, tuberculosi, disenteria bacteriana i moltes al·lèrgies, el comportament estacional de les quals és ben conegut.

---

---

# CAPÍTOL 3

---

## MODELS AUTOREGRESSIUS DE PRIMER ORDRE AMB INNOVACIONS NO GAUSSIANES

### 3.1 Introducció

En aquest capítol, considerarem models autoregressius de primer ordre, definits per l'equació en diferències estocàstica clàssica

$$Y_t = \alpha \cdot Y_{t-1} + W_t, \tag{3.1}$$

### 3.1. Introducció

---

satisfent la condició d'estacionarietat  $|\alpha| < 1$ , on  $W_t$  són les innovacions o soroll blanc, una seqüència de variables aleatòries independents i idènticament distribuïdes amb mitjana  $\mu$  i variància  $\sigma^2$ .

Molts autors han estudiat models autoregressius amb innovacions no gaussianes. La majoria de treballs tenen com a objectiu principal trobar la distribució de les innovacions per marginals determinades. A [17], els autors consideren els casos on la distribució marginal de  $Y_t$  és una exponencial o gamma. També mostren que marginals positives no permeten valors de  $\alpha < 0$  a (3.1), de manera que la condició d'estacionarietat es restringeix a  $0 < \alpha < 1$ .

Una bona revisió de models autoregressius amb innovacions no gaussianes és la presentada a [43], on es considera una metodologia de construcció de models enfocada a l'estimació de paràmetres mitjançant diversos models amb diferents distribucions marginals per  $Y_t$ , com ara l'exponencial, la logística, la secant hiperbòlica i altres. A [1] s'analitza el cas en què la marginal segueix una distribució inversa gaussiana.

Granger i Newbold (1976) construeixen sèries no gaussianes a [19] prenent una transformació instantània  $T(Z_t)$  d'un procés gaussià ARMA  $Z_t$ , com els definits a (1.5). Els autors estudien amb especial detall la transformació  $T(x) = e^x$ , perquè un gran nombre de sèries temporals d'indicadors econòmics són habitualment analitzades en forma logarítmica, tot i que fer inferència sobre la sèrie



original continua essent l'objectiu principal. L'estudi de sèries temporals de termes positius de la forma  $X_t = e^{Y_t}$ , on  $Y_t$  és un model AR(1) amb innovacions no gaussianes, ha estat poc tractat en la literatura, tot i que existeixen algunes referències interessants en aquest sentit com per exemple [31]. En aquest article s'aborda el cas on la distribució marginal de  $X_t$  és gamma, i es prova que la distribució gamma és la única en la qual es compleix que les estructures de correlació de  $X_t$  i  $Y_t$  són la mateixa.

A la propera secció es mostrarà que la funció d'autocorrelació (ACF) de  $X_t$ ,  $\rho_{X_t}(k)$ , caracteritza la distribució de les innovacions  $W_t$  de  $Y_t$  a (3.1). A continuació, veurem com podem utilitzar aquest resultat per tal de construir un contrast de bondat d'ajust per a la distribució de les innovacions, basat en la funció d'autocorrelació empírica de  $X_t$ . En particular, aquest contrast podrà utilitzar-se per al model AR(1) clàssic on el soroll blanc es distribueix normalment.

## 3.2 La funció d'autocorrelació de l'exponencial d'un AR(1)

Considerem el model AR(1) general  $Y_t$  definit a (3.1) i la seva sèrie temporal exponencial  $X_t = e^{Y_t}$ . Considerarem que aquesta operació dóna lloc a una distribució amb moments de tots els ordres. A [31], McKenzie proporciona un mètode senzill per calcular la funció d'autocorrelació de  $X_t$  en termes de la seva distribució marginal, denotada com la variable aleatòria  $X$ . En aquest article, planteja

### 3.2. La funció d'autocorrelació de l'exponencial d'un AR(1)

---

el següent resultat:

**Lema 3.2.1** [31] *La funció d'autocorrelació de la sèrie temporal exponenciada  $X_t$  té l'expressió,*

$$\rho_{X_t}(k) = \frac{\mathbb{E}[X] \left( \mathbb{E}[X^{\alpha^{k+1}}] - \mathbb{E}[X^{\alpha^k}] \mathbb{E}[X] \right)}{\mathbb{E}[X^{\alpha^k}] \text{Var}[X]}, k = 0, 1, 2, \dots \quad (3.2)$$

**Prova.** Per qualsevol  $\alpha \in (0, 1)$  podem trobar una seqüència de variables aleatòries independents i idènticament distribuïdes  $W_t$  complint

$$Y_t = \alpha Y_{t-1} + W_t$$

Tenint en compte que  $Y_t = \log X_t$ , podem escriure  $X_t = X_{t-1}^\alpha e^{W_t} = \left( \prod_{i=0}^{k-1} V_{t-i}^{\alpha^i} \right) X_{t-k}^{\alpha^k}$ , on  $V_t = e^{W_t}$ . I d'aquí tenim que

$$\mathbb{E}[X_t X_{t-k}] = \left( \prod_{i=0}^{k-1} \mathbb{E}[V^{\alpha^i}] \right) \mathbb{E}[X^{\alpha^{k+1}}].$$

Però  $\mathbb{E}[X^s] = \mathbb{E}[X^{\alpha^s}] \mathbb{E}[V^s]$ , i per tant

$$\mathbb{E}[X_t X_{t-k}] = \prod_{i=0}^{k-1} \left( \frac{\mathbb{E}[X^{\alpha^i}] \mathbb{E}[V^{\alpha^i}]}{\mathbb{E}[X^{\alpha^{i+1}}]} \right) \mathbb{E}[X^{\alpha^{k+1}}] = \frac{\mathbb{E}[X] \mathbb{E}[X^{\alpha^{k+1}}]}{\mathbb{E}[X^{\alpha^k}]}.$$

Tenint en compte que  $\gamma_X(k) = \mathbb{E}[(X_t - \mathbb{E}[X_t])(X_{t-k} - \mathbb{E}[X_{t-k}])]$  i que  $\rho_{X_t}(k) = \frac{\gamma_X(k)}{\gamma_X(0)}$ , tenim el resultat ■

El resultat següent estableix una important relació amb la distribució marginal de  $Y_t$ , denotada com la variable aleatòria  $Y$ :

**Lema 3.2.2** *La funció d'autocorrelació de  $X_t$  es pot escriure com,*

$$\rho_{X_t}(k) = \frac{\psi_Y(1)\psi_Y(\alpha^k + 1) - \psi_Y(1)^2\psi_Y(\alpha^k)}{\psi_Y(\alpha^k)(\psi_Y(2) - \psi_Y(1)^2)}, k = 0, 1, 2, \dots, \quad (3.3)$$

on  $\psi_Y(z)$  és la funció generatriu de moments de  $Y$ .

**Prova.** L'expressió (3.3) és una conseqüència directa de (3.2) i del fet que  $\psi_Y(z) = \mathbb{E}[e^{zY}] = \mathbb{E}[e^{z \log(X)}] = \mathbb{E}[X^z]$  ■

**Exemple 3.2.3** *Considerem per exemple la situació clàssica, en que la distribució de les innovacions és normal, és a dir,  $W_t \sim N(\mu, \sigma^2)$ . En aquest cas, la distribució marginal de  $Y_t = \alpha Y_{t-1} + W_t$  serà també normal, amb mitjana  $\frac{\mu}{1-\alpha}$  i variància  $\frac{\sigma^2}{(1-\alpha^2)}$ . Per tant, la funció generatriu de moments de  $Y$  serà  $\psi_Y(s) = e^{\frac{\mu}{1-\alpha} + \frac{1}{2} \frac{\sigma^2}{(1-\alpha^2)} t^2}$ , de manera que d'acord amb el lemma 3.2.2, tindrem que*

$$\rho_{X_t}(k) = \frac{e^{\frac{\sigma^2 \alpha^k}{1-\alpha^2} - 1}}{e^{\frac{\sigma^2}{1-\alpha^2} - 1}}, k = 0, 1, 2, \dots,$$

on  $X_t$  és la sèrie exponenciada  $X_t = e^{Y_t}$ .

És important remarcar que, donada la distribució marginal de  $Y_t$ , el lemma 3.2.2 proporciona una manera senzilla de calcular la funció d'autocorrelació de la sèrie exponenciada, com s'ha vist a l'exemple 3.2.3 per al cas d'innovacions

### 3.2. La funció d'autocorrelació de l'exponencial d'un AR(1)

---

gaussianes.

La relació entre la distribució de  $Y$  i la distribució de les innovacions  $W_t$  de (3.1) es pot establir a través de les seves respectives funcions característiques. La funció característica  $\phi_{Y_t}(z)$  de la distribució de  $Y_t$  és:

$$\phi_{Y_t}(z) = \mathbb{E}[e^{ziY_t}] = \mathbb{E}[e^{zi(\alpha Y_{t-1} + W_t)}] = \phi_{Y_{t-1}}(\alpha z)\phi_{W_t}(z).$$

D'aquí, com que  $Y_t$  és estacionari, obtenim,  $\phi_Y(z) = \phi_Y(\alpha z)\phi_W(z)$ , i en conseqüència,

$$\phi_W(z) = \frac{\phi_Y(z)}{\phi_Y(\alpha z)}. \quad (3.4)$$

Cal tenir en compte que l'expressió (3.4) no és sempre la funció característica d'una distribució apropiada per a les innovacions. Les variables aleatòries  $Y$  per a les quals l'expressió (3.4) és una funció característica apropiada han de tenir necessàriament distribucions autodescomposables, una classe de distribucions molt àmplia, definida a la secció 0.5. Durant la resta del capítol treballarem amb algunes d'aquestes distribucions, que són el logaritme de variables aleatòries positives  $X$  ben conegudes, la distribució marginal de  $X_t$ . La taula 3.1 mostra les marginals que considerarem, la funció generatriu de moments de  $Y = \log(X)$  i la funció d'autocorrelació de la sèrie temporal  $X_t$ , expressada com  $\rho_X(s)$  on  $s = \alpha^k$ .

Com que en el procés AR(1) definit a (3.1) les innovacions satisfan  $\mathbb{E}[W_t] = \mu$  i  $\mathbb{V}[W_t] = \sigma^2$ , és immediat veure que l'esperança i la variància de la distribució

Taula 3.1: Distribucions marginals de  $X_t$  amb la funció generatriu de moments de la distribució marginal de  $Y_t$ ,  $\psi_Y(z)$  ( $Y = \log(X)$ ), i la ACF estandarditzada  $\rho_X(s)$  ( $s = \alpha^k$ )

<b>X</b>	$\psi_Y(z)$	$\rho_X(s)$
log-Normal	$e^{\frac{\mu}{1-\alpha}z + \frac{\sigma^2}{2(1-\alpha^2)}z^2}$	$\frac{e^s - 1}{e - 1}$
log-Uniforme	$\frac{e^{\frac{\mu}{1-\alpha}z} \sinh(\sqrt{\frac{3}{1-\alpha^2}}\sigma z)}{\sqrt{\frac{3}{1-\alpha^2}}\sigma z}$	$\frac{s-c}{s+1} + \frac{c\sqrt{3}s \cosh(\sqrt{3}s)}{(s+1) \sinh(\sqrt{3}s)}$ $c = \frac{\sinh(\sqrt{3})}{\sqrt{3} \cosh(\sqrt{3}) - \sinh(\sqrt{3})}$
Gamma	$\frac{\Gamma(\nu+z)}{\Gamma(\nu)a^z}$	s
Inv. Gaussian	$\frac{K_{z-1/2}(ab)}{K_{-1/2}(ab)} \left(\frac{a}{b}\right)^z$	$c \left( \frac{K_{s+1/2}(c)}{K_{s-1/2}(c)} - 1 \right)$ c=0.608545

marginal són  $\mathbb{E}[Y] = \frac{\mu}{1-\alpha}$  i  $\mathbb{V}[Y] = \frac{\sigma^2}{1-\alpha^2}$ . Totes les distribucions marginals considerades a la taula 3.1 tenen dos paràmetres. Les dues primeres funcions generatrius de moments s'han expressat en termes de l'esperança i la variància. Per a la distribució gamma, això es pot fer resolent les equacions  $\Psi'(\nu) = \frac{\sigma^2}{1-\alpha^2}$  i  $\Psi(\nu) - \log(a) = \frac{\mu}{1-\alpha}$ , on  $\Psi(\cdot)$  és la funció digamma i  $\Psi'(\cdot)$  és la seva primera derivada. Pel que fa al cas de la distribució inversa gaussiana, obtenir una expressió en termes de  $\frac{\mu}{1-\alpha}$  i  $\frac{\sigma^2}{1-\alpha^2}$  és més complicat a causa de la funció de Bessel modificada de segona classe que apareix a  $\psi_Y(z)$ . De totes maneres, la qüestió es pot resoldre numèricament fent servir un programari apropiat com *Maple*.

**Exemple 3.2.4** Considerem el cas en que les innovacions segueixen una distribució uniforme,  $W_t \sim U(a, b)$ . Per tal que es compleixi  $\mathbb{E}[Y] = 0$  i  $\mathbb{V}[Y] = 1$ , tindrem que  $W_t \sim U(-\sqrt{3(1-\alpha^2)}, \sqrt{3(1-\alpha^2)})$ . Aleshores, la seva funció gene-

### 3.2. La funció d'autocorrelació de l'exponencial d'un AR(1)

---

ratriu de moments serà

$$\psi_W(z) = \frac{e^{-z\sqrt{3(1-\alpha^2)}} - e^{z\sqrt{3(1-\alpha^2)}}}{2z\sqrt{3(1-\alpha^2)}}$$

La funció generatriu de moments de  $Y$ ,  $\psi_Y(z) = \prod_{k=0}^{\infty} \psi_W(\alpha^k z)$  no és fàcilment expressable explícitament, però podem calcular numèricament la transformada de Laplace inversa per tal d'estimar la funció de densitat de  $Y$ :

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \psi_Y(t) dt \quad (3.5)$$

Per valors de  $\alpha$  entre  $-0,8$  i  $0,8$  podem graficar les corresponents funcions de densitat, representades a la figura 3.1. Observem que el resultat és el mateix per  $\alpha$  que per  $-\alpha$ , en tots els casos considerats.

I d'acord amb (3.2), podem calcular també en aquest cas la funció d'autocorrelació de la sèrie exponenciada  $X = e^Y$ , representada pels mateixos valors de  $\alpha$  a la figura 3.2.

La majoria de les funcions generatrius de moments considerades a la taula 3.1 estan ben definides per  $z \in (-1, 1)$ . Aquest fet és important perquè permet considerar models AR(1) com els definits a (3.1) amb  $\alpha \in (-1, 1)$ . En el cas de la distribució gamma, això implica una restricció en el domini dels paràmetres, concretament, per assegurar aquesta bona definició, necessitem que  $\nu > 1$ .

És important tenir en compte que la funció d'autocorrelació del model AR(1) definit a (3.1) és sempre la mateixa,  $\rho_Y(k) = \alpha^k$ , independentment de la distri-

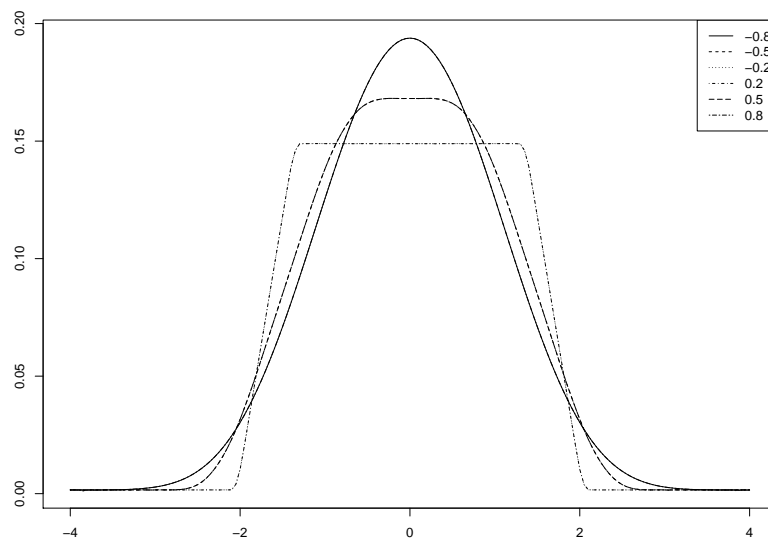


Figura 3.1: Funcions de densitat de  $Y$  en el cas d'innovacions uniformes

bució marginal o de la distribució de les innovacions. Al contrari del que succeeix per  $Y_t$ , la funció d'autocorrelació de  $X_t$  depèn de la distribució marginal de  $Y_t$ . La figura 3.3 mostra les gràfiques de  $\rho_X(s)$  per a les marginals indicades, prèviament estandarditzades ( $\mathbb{E}[Y] = 0$  i  $\mathbb{V}[Y] = 1$ ).

En la propera secció veurem que la funció d'autocorrelació de  $X_t$  caracteritza la distribució marginal de  $Y_t$ , i en conseqüència, la distribució de les innovacions.

### 3.3. Caracterització de la distribució de les innovacions

---

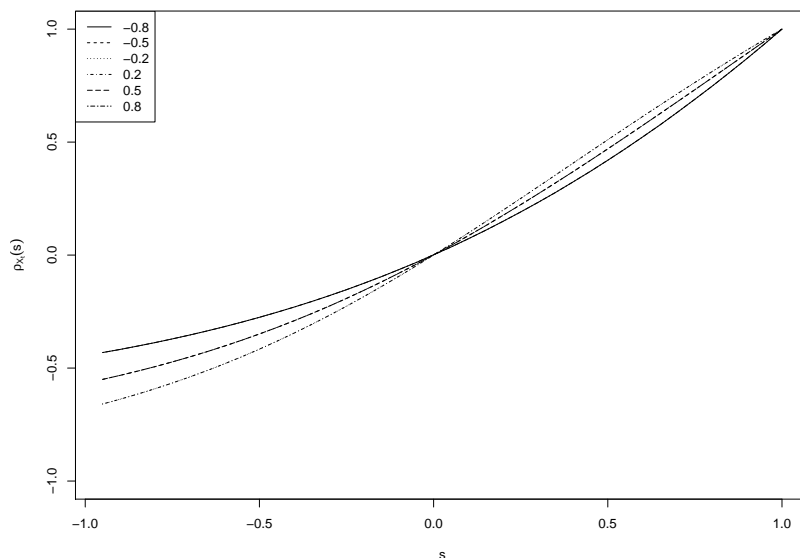


Figura 3.2: Funcions d'autocorrelació de  $X$  en el cas d'innovacions uniformes

### 3.3 Caracterització de la distribució de les innovacions

Prenent logaritmes a l'expressió (3.3) del lemma 3.2.2 i amb una mica d'àlgebra obtenim,

$$\kappa_Y(s+1) - \kappa_Y(s) = \log \left( \frac{\psi_Y(2) - \psi_Y(1)^2}{\psi_Y(1)} \cdot \rho_{X_t}(s) + \psi_Y(1) \right), \quad (3.6)$$

on  $s = \alpha^k$  i  $\kappa_Y(s) = \log(\psi_Y(s))$  és la funció generatriu de cumulants de  $Y$ . Denotant per  $h(s)$  la part dreta de (3.6), podem veure que  $\kappa_Y(s)$  és una solució



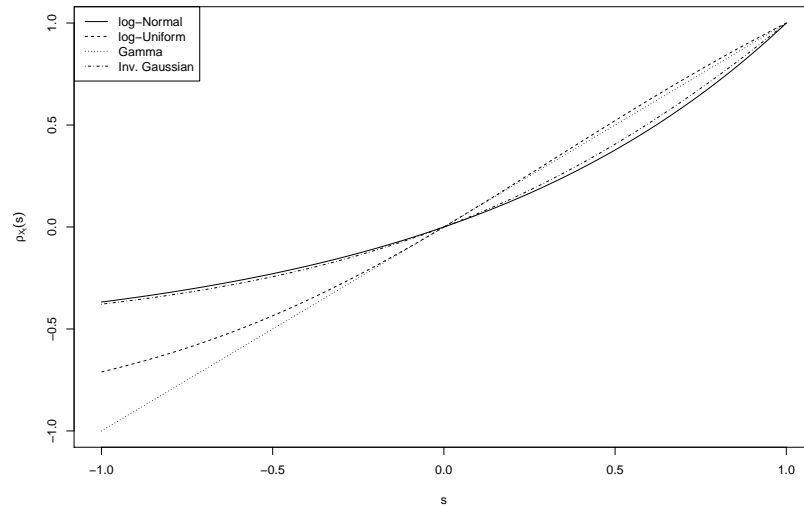


Figura 3.3:  $\rho_X(s)$  per algunes marginals estandarditzades

de l'equació funcional de primer ordre de la forma,

$$f(s+1) - f(s) = h(s), \quad (3.7)$$

amb la condició inicial  $f(0) = 0$ , on  $h(s)$  és la part dreta de (3.6), com s'ha indicat a la secció 0.3. Aquest tipus d'equacions funcionals ha estat estudiat àmpliament a la literatura, per exemple a [25]. Una lleugera modificació del seu teorema 5.11 ens permet enunciar el següent resultat:

**Proposició 3.3.1** [25] *Considerem una funció  $h : I \rightarrow \mathbb{R}$  a (3.7),  $I = (a, \infty)$ ,  $-\infty \leq a < \infty$ , satisfent la condició*

$$\lim_{s \rightarrow \infty} [h(s+1) - h(s)] = 0.$$

### 3.3. Caracterització de la distribució de les innovacions

---

Suposem que tenim una solució convexa de (3.7),  $\varphi(s)$  definida a  $I$ , complint que  $\varphi(s_0) = \eta_0$ , per  $s_0 \in I$  i  $\eta_0 \in \mathbb{R}$ . Aleshores,  $\varphi(s)$  és l'única solució convexa de l'equació funcional (3.7), i es pot expressar com,

$$\varphi(s) = \eta_0 + (s - s_0)h(s_0) + \sum_{n=0}^{\infty} (s - s_0)(h(s_0 + n + 1) - h(s_0 + n)) - (h(s + n) - h(s_0 + n)). \quad (3.8)$$

Partint d'aquesta proposició, estem en condicions d'enunciar el següent teorema, que caracteritza la distribució de les innovacions.

**Teorema 3.3.2** *Suposem que la distribució de les innovacions  $W_t$  de (3.1) és tal que,*

1. *La distribució marginal de  $Y_t$  (indicada per la variable aleatòria  $Y$ ) té funció generatriu de moments  $\psi_Y(s)$  definida a  $s \in (a, \infty)$ ,  $a < 0$ .*
2. *La distribució de  $X = e^Y$  està ben definida, amb ACF de la sèrie temporal exponenciada  $X_t = e^{Y_t}$  donada per  $\rho_{X_t}(s)$ , i  $\lim_{s \rightarrow \infty} h(s + 1) - h(s) = b$ ,  $0 \leq b < \infty$ , on*

$$h(s) = \log \left( \frac{\psi_Y(2) - \psi_Y(1)^2}{\psi_Y(1)} \cdot \rho_{X_t}(s) + \psi_Y(1) \right).$$

3. *La funció  $\log(\psi_Y(s)) - \frac{bs(s-1)}{2}$  és convexa.*

*Aleshores, la distribució de  $W_t$  és l'única que té com a funció d'autocorrelació de la sèrie exponenciada a  $\rho_{X_t}(s)$ .*

**Prova.** Sigui  $\kappa_Y(s) = \log(\psi_Y(s))$ , la funció generatriu de cumulants de  $Y$ . Tenint present (3.6), és clar que  $f(s) = \kappa_Y(s)$  satisfà l'equació funcional (3.7) amb la condició inicial  $f(0) = 0$ . Com que qualsevol funció generatriu de cumulants és convexa, si el límit  $b$  és zero, la proposició 3.3.1 mostra directament que (3.7) té només una solució. Si el límit  $b$  és més gran que 0, definim  $h^*(s) = h(s) - bs$ , i és immediat veure que  $f(s) = \kappa_Y(s) - bs(s-1)/2$  és una solució de l'equació funcional  $f(s+1) - f(s) = h^*(s)$ . Com que  $\lim_{s \rightarrow \infty} h^*(s+1) - h^*(s) = 0$  i per hipòtesi  $\kappa_Y(s) - \frac{bs(s-1)}{2}$  és convexa, altre cop la proposició 3.3.1 prova que aquesta solució és única ■

Els valors del límit  $b = \lim_{s \rightarrow \infty} h(s+1) - h(s)$  per als exemples contemplats a la taula 3.1 són 0 excepte per al cas log-Normal (innovacions gaussianes) on es té que  $b = 1$ .

**Exemple 3.3.3** *Suposem que  $X$  té una funció d'autocorrelació amb forma de polinomi de segon grau  $\rho_X(s) = as^2 + bs + c$ . Imposant les condicions  $\rho_X(0) = 0$ ,  $\rho_X(1) = 1$  i  $\rho_X(-1) = \frac{e^{-1}-1}{e-1}$ , com en el cas en què  $X$  segueix una distribució log-Normal (veure taula 3.1), tenim que aquesta funció d'autocorrelació ha de ser  $\rho_X(s) = (1-b)s^2 + bs$ , on  $b = \frac{1}{2} \cdot \left(1 - \frac{e^{-1}-1}{e-1}\right)$ . Expressant  $h(s) = \log\left(\frac{\mathbb{V}[X]}{\mathbb{E}[X]} \cdot \rho_X(s) + \mathbb{E}[X]\right)$  i resolent l'equació en diferències (3.7) podem obtenir la funció generatriu de cumulants de  $Y = \log X$ , i exponenciant aquesta tindrem la funció generatriu de moments de  $Y$ . Avaluant aquesta a  $z = is$  tindrem la funció característica  $\Phi_Y(z)$ , i fent servir que  $\Phi_W(z) = \frac{\Phi_Y(z)}{\Phi_Y(\alpha z)}$  podrem obtenir la funció de densitat  $f(x)$  de les innovacions  $W_t$ , calculant la transformada de*

### 3.3. Caracterització de la distribució de les innovacions

---

*Laplace inversa*

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \frac{\Phi_Y(t)}{\Phi_Y(\alpha t)} dt.$$

*Els resultats obtinguts per diferents valors de  $\alpha$  de -0,8, -0,5, -0,2, 0,2, 0,5 i 0,8 es poden veure a la figura 3.4.*

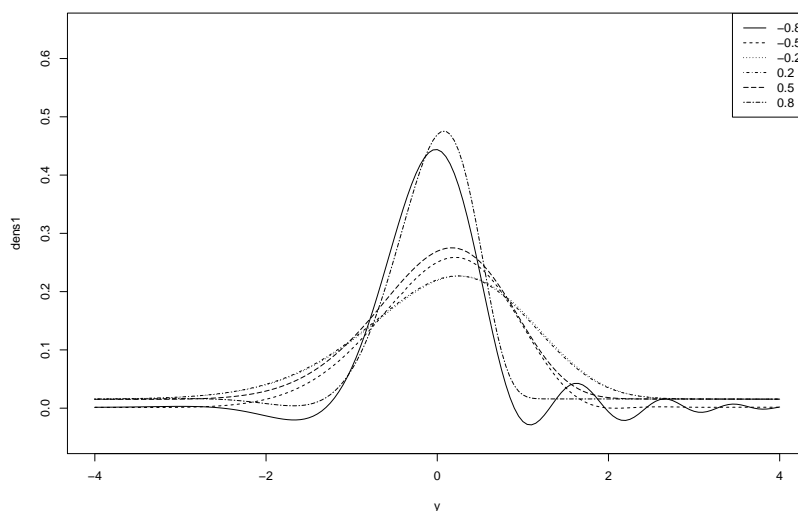


Figura 3.4: Funcions de densitat de les innovacions en el cas  $\rho_X(s) = as^2 + bs + c$

*Observem que per la majoria de valors de  $\alpha$ , la funció de densitat de les innovacions és similar a una normal estàndard, com podríem esperar en haver considerat una funció d'autocorrelació  $\rho_X$  de  $X$  similar a la de la distribució log-Normal. Veiem que per  $\alpha = -0,8$ , la funció d'autocorrelació no dóna lloc a una distribució pròpia per  $W_t$ . Això dóna lloc a una qüestió interessant, com és la caracterització de les funcions d'autocorrelació  $\rho_X$  que donen lloc a innovacions pròpies.*

En la propera secció veurem amb detall com es pot aplicar aquest teorema de caracterització per tal de construir contrastos de bondat de l'ajust en aquest context.

### 3.4 Un contrast de bondat de l'ajust per a les innovacions

Els resultats de la secció anterior es poden fer servir per tal de construir un contrast de bondat de l'ajust per a la distribució de les innovacions. A partir d'aquí centrarem la nostra atenció en el procés clàssic AR(1) gaussià, malgrat que aquesta metodologia es pot generalitzar a altres models AR(1). Per tant, les hipòtesis nul·la i alternativa que considerarem són les següents:

$$H_0 : W_t \sim N(\mu, \sigma^2) \quad H_1 : W_t \approx N(\mu, \sigma^2) \quad (3.9)$$

Cal tenir present que el teorema 3.3.2 prova que aquestes hipòtesis són equivalents a les següents:

$$H_0 : \rho_{X_t}(k) = \frac{e^{\frac{\sigma^2 \alpha^k}{1-\alpha^2}} - 1}{e^{\frac{\sigma^2}{1-\alpha^2}} - 1} \quad H_1 : \rho_{X_t}(k) \neq \frac{e^{\frac{\sigma^2 \alpha^k}{1-\alpha^2}} - 1}{e^{\frac{\sigma^2}{1-\alpha^2}} - 1}, \quad (3.10)$$

on  $X_t$  és la sèrie exponenciada  $X_t = e^{Y_t}$ , d'acord amb l'expressió obtinguda per  $\rho_{X_t}(k)$  en el cas d'innovacions gaussianes a l'exemple 3.2.3.

Per tant, una idea senzilla per construir un contrast de bondat d'ajust és

### 3.4. Un contrast de bondat de l'ajust per a les innovacions

---

comparar la funció d'autocorrelació empírica de  $X_t$  i la funció d'autocorrelació teòrica sota la hipòtesi d'innovacions gaussianes. En comptes de la funció d'autocorrelació directament, preferim emprar la transformada de Fourier discreta de les autocorrelacions, això és, la densitat espectral estandarditzada, ja que presenta alguns avantatges en termes d'estabilitat. Tot i que ja s'ha comentat al primer capítol, no està de més recordar aquí que el coneixement de la densitat espectral estandarditzada és equivalent al coneixement de la funció d'autocorrelació.

Donat un procés estacionari amb funció d'autocorrelació denotada per  $\rho_k$ ,  $k = 0, 1, 2, \dots$ , la seva densitat espectral estandarditzada es pot definir de la manera següent:

**Definició 3.4.1** *La densitat espectral estandarditzada d'un procés estacionari amb funció d'autocorrelació  $\rho_k$ ,  $k = 0, 1, 2, \dots$  és una funció  $f$  definida per*

$$f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \rho_k \cos(\omega k) = \frac{1}{2\pi} \left( 1 + 2 \sum_{k=1}^{\infty} \rho_k \cos(\omega k) \right), \quad -\pi \leq \omega \leq \pi.$$

Observem que  $f(\omega) = f(-\omega)$ . A partir d'aquí podem definir la distribució espectral estandarditzada de la manera següent:

**Definició 3.4.2** *La distribució espectral estandarditzada d'un procés estacionari amb funció d'autocorrelació  $\rho_k$ ,  $k = 0, 1, 2, \dots$  és una funció  $F$  definida per*

$$F(\lambda) = \int_{-\lambda}^{\lambda} f(\omega) d\omega = 2 \int_0^{\lambda} f(\omega) d\omega = \frac{1}{\pi} \left( \lambda + 2 \sum_{k=1}^{\infty} \rho_k \frac{\sin(\lambda k)}{k} \right), \quad 0 \leq \lambda \leq \pi.$$

**Definició 3.4.3** Donada una sèrie temporal  $x_1, x_2, \dots, x_n$ , els coeficients d'autocorrelació mostral  $r_k$ ,  $k = 1, \dots, n - 1$  es defineixen com  $r_k = c_k/c_0$  on

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

són les covariàncies mostrals.

Donada una sèrie temporal amb  $n$  observacions, una estimació de la distribució espectral estandarditzada es pot obtenir directament utilitzant la funció d'autocorrelació mostral  $r_k$ ,  $k = 0, 1, 2, \dots$  ([15]), obtenint

$$\hat{F}_n(\lambda) = \frac{1}{\pi} \left( \lambda + 2 \sum_{k=1}^{n-1} r_k \frac{\sin(\lambda k)}{k} \right), \quad 0 \leq \lambda \leq \pi. \quad (3.11)$$

Anderson suggereix a [3] contrastar la hipòtesi que una sèrie temporal donada prové d'un procés amb una densitat espectral estandarditzada específica  $f_0(\lambda)$ , mitjançant un estadístic de Cramér-von Mises de la forma,

$$W_n^2 = \frac{n}{2\pi G^2(\pi)} \int_0^\pi \left[ \hat{F}_n(\lambda) - F_0(\lambda) \right]^2 f_0^2(\lambda) d\lambda, \quad (3.12)$$

on  $G(\pi) = 2 \int_0^\pi f_0^2(\omega) d\omega$ . Aquest estadístic mesura la discrepància entre la distribució espectral estandarditzada teòrica i la seva estimació mostral. En aquest mateix treball es proposa una expressió molt útil des d'un punt de vista pràctic que permet obtenir  $W_n^2$ , observant que si deixem de banda el terme  $\frac{2\sqrt{n}}{\pi} \sum_{h=n}^\infty \frac{\sin \lambda h}{h}$ ,

### 3.4. Un contrast de bondat de l'ajust per a les innovacions

---

podem escriure

$$W_n^2 = \frac{n}{2\pi G^2(\pi)} \cdot \int_0^\pi \left[ \frac{2}{\pi} \sum_{h=1}^{n-1} \frac{\sin \lambda h}{h} (r_h - \rho_h) \right]^2 \left[ \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} \rho_r e^{i\lambda r} \right]^2 d\lambda = \quad (3.13)$$

$$\frac{n}{4\pi^5 G^2(\pi)} \sum_{g,h=1}^{n-1} \frac{(r_g - \rho_g)(r_h - \rho_h)}{gh} \cdot \sum_{r,s=-\infty}^{\infty} \rho_r \rho_s \int_{-\pi}^{\pi} \sin \lambda g \sin \lambda h e^{i\lambda(r-s)} d\lambda$$

Integrant aquesta expressió, obtenim

$$W_n^2 = \frac{n}{4\pi^4 G^2(\pi)} \sum_{j=1}^{\infty} \left[ \sum_{k=1}^{n-1} \frac{(r_k - \rho_k)(\rho_{j+k} - \rho_{j-k})}{k} \right]^2 \quad (3.14)$$

Pensant en la implementació pràctica, aquesta suma infinita es pot truncar a partir d'un cert valor de  $j$ , quan els termes següents siguin negligibles.

A [5], els autors utilitzen l'estadístic  $W_n^2$  per contrastar la hipòtesi que una sèrie temporal donada prové d'un procés amb estructura AR(1). També podem utilitzar  $W_n^2$  per contrastar si la sèrie exponenciada  $X_t$  té una funció d'autocorrelació específica, és a dir, contrastar les hipòtesis (3.10) o equivalentment (3.9).

Convé remarcar que la situació més habitual en la pràctica serà aquella en la qual  $\rho_k$  són desconeguts perquè depenen de  $\alpha$  i  $\sigma^2$ , que seran normalment desconeguts.

En qualsevol cas, poden ser estimats a partir de la sèrie original (3.1), utilitzant el primer coeficient d'autocorrelació mostral  $\hat{\alpha}$  i  $\hat{\sigma}^2 = s_y^2(1 - \hat{\alpha}^2)$ , on  $s_y^2$  és la



variància mostral. Per tant, definim

$$\hat{\rho}_k = \frac{e^{\frac{\hat{\sigma}^2 \hat{\alpha}^k}{1-\hat{\alpha}^2}} - 1}{e^{\frac{\hat{\sigma}^2}{1-\hat{\alpha}^2}} - 1} = \frac{e^{s_y^2 \hat{\alpha}^k} - 1}{e^{s_y^2} - 1}, \quad (3.15)$$

i anomenarem  $\hat{W}_n^2$  l'estadístic que considerarem, construït substituint  $\rho_s$  per  $\hat{\rho}_s$  i  $G^2(\pi)$  per  $\hat{G}^2(\pi) = [2 \int_0^\pi \hat{f}_0^2(\omega) d\omega]^2$  a (3.14), on  $\hat{f}_0(\omega) = \frac{1}{2\pi} (1 + 2 \sum_{k=1}^\infty \hat{\rho}_k \cos(\omega k))$ .

Per tant, donada una sèrie  $Y_t$  amb estructura AR(1), els passos que cal realitzar per aplicar el contrast proposat són els que s'enumeren a continuació:

1. Estandarditzar la sèrie, si és necessari.
2. Calcular la sèrie exponencial  $X_t$ .
3. Estimar els paràmetres de les sèries, per exemple mitjançant les equacions de Yule-Walker presentades a (1.2), i calcular els coeficients d'autocorrelació mostrals per  $X_t$ .
4. Obtenir l'estadístic  $\hat{W}_n^2$  del contrast d'acord amb l'expressió 3.14.
5. Fixant els paràmetres  $\alpha$  i  $\sigma^2$  obtinguts al tercer pas, simular noves sèries i calcular l'estadístic del contrast  $\hat{W}_n^2$  per a cadascuna. Per a construir cada sèrie simulada es generen innovacions gaussianes de mitjana 0 i variància  $1 - \hat{\alpha}^2$ , i es simula una sèrie partint de  $\hat{\alpha}$  i  $\hat{\sigma}^2$ , amb estructura AR(1) i aquestes innovacions. Aquesta sèrie simulada es modifica fins que els seus paràmetres són exactament els mateixos  $\hat{\alpha}$  i  $\hat{\sigma}^2$ .

### 3.5. Dos exemples

---

6. El p-valor del contrast es pot obtenir en base a la distribució de  $\hat{W}_n^2$  obtinguda.

A la propera secció veurem alguns exemples d'aplicació d'aquest contrast de bondat de l'ajust.

## 3.5 Dos exemples

Per als exemples que descriurem a continuació, s'han realitzat 10000 simulacions, seguint l'esquema descrit a l'apartat anterior.

### 3.5.1 Captures de peix a l'Atlàntic i al golf del riu St. Lawrence

Considerem les captures mensuals de peix a l'Atlàntic i al golf del riu St. Lawrence des del 1990 fins al 1996 com el nostre procés  $Y_t$ . Aquest exemple va ser tractat prèviament a [5], on els autors consideren que un model AR(1) és adequat per ajustar aquestes dades, després d'eliminar l'estacionalitat, que té un efecte notable. S'estandarditza la sèrie per tal de ser capaç de calcular la sèrie exponencial sense dificultats numèriques, i els autors mostren que l'estructura de model autoregressiu de primer ordre no es pot rebutjar (p-valor de 0,83). Utilitzant les equacions de Yule-Walker podem estimar els paràmetres del model, obtenint  $\hat{\alpha} = 0,39$  i  $\hat{\sigma}^2 = 0,86$ . En aquest cas, obtenim un valor de l'estadístic de  $W_n^2 = 0,067$ , corresponent a un p-valor de 0,69. Per tant, la hipòtesi nul·la de normalitat dels residus  $W_t$  no pot ser rebutjada. En la figura 3.6 podem veure

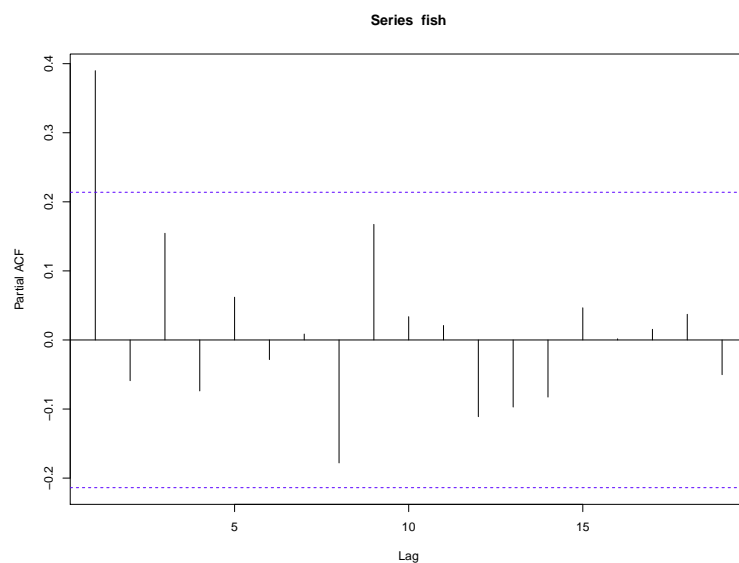


Figura 3.5: Funció d'autocorrelació parcial (captures de peix)

com la funció de distribució espectral teòrica sota la hipòtesi nul·la i la funció de distribució espectral mostral tenen un comportament similar, com esperàvem pel resultat del contrast de bondat de l'ajust.

La taula 3.2 mostra els p-valors per un rang de coeficients d'autocorrelació  $\alpha$  al voltant del valor empíric de  $\hat{\alpha} = 0,39$ . Observem que la hipòtesi nul·la de normalitat de les innovacions no es pot rebutjar per cap valor de  $\alpha$ .

Taula 3.2: p-valors corresponents a l'exemple de les captures de peix

$\alpha$	0,1	0,2	0,3	0,4	0,5	0,6
p-valor	0,58	0,64	0,67	0,69	0,69	0,68

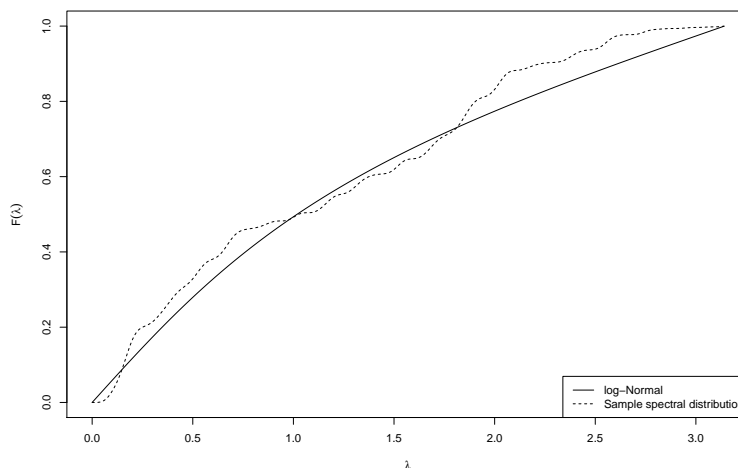


Figura 3.6: Comparació de les distribucions espectrals mostral i teòrica (captures de peix)

### 3.5.2 Deflactor del producte interior brut (PIB) espanyol

Com a segon exemple, considerem una sèrie del Banc Mundial consistent en els valors del deflactor del producte interior brut espanyol anual des del 1962 fins al 2011. Per tal de treballar amb valors més baixos, s'han pres logaritmes sobre aquesta sèrie i s'ha estandarditzat.

De nou mitjançant les equacions de Yule-Walker estimem els paràmetres del model, obtenint uns valors de  $\hat{\alpha} = 0,79$  i  $\hat{\sigma}^2 = 0,35$ . Com podem veure en la figura 3.7, un model autoregressiu de primer ordre sembla apropiat, i això s'ha confirmat a través d'un contrast com el proposat a [5]. En aquest cas, obtenim un p-valor de 0,80, de manera que l'estructura AR(1) no es pot rebutjar amb un nivell de confiança del 95%.

Contrastant la normalitat de les innovacions de la manera descrita, obtenim un valor de  $W_n^2 = 0,24$ , corresponent a un p-valor de 0,023, i per tant, amb un nivell de confiança del 95%, podem rebutjar la hipòtesi nul·la que les innovacions  $W_t$  segueixen una distribució normal. L'estadístic definit a (3.14) pot utilitzar-se

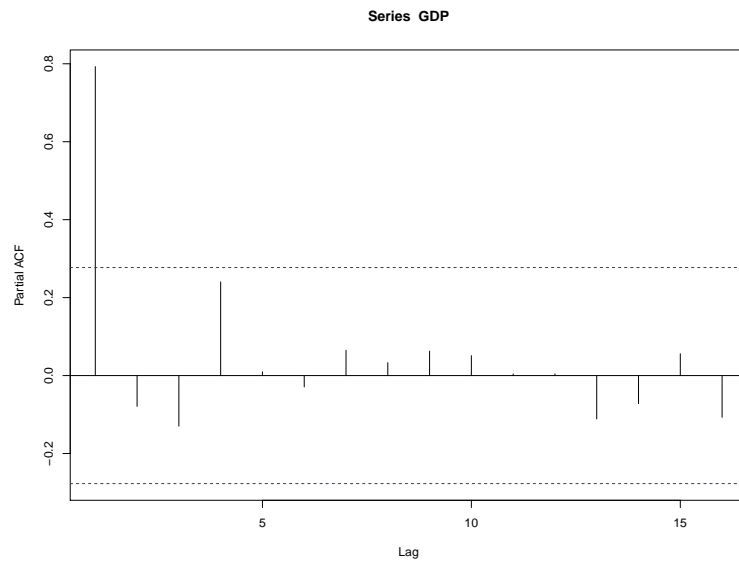


Figura 3.7: Funció d'autocorrelació parcial (PIB espanyol)

també per tal de triar una distribució marginal per  $X_t$  més apropiada comparant els valors de l'estadístic considerant les distribucions descrites a la taula 3.1 per a la distribució marginal de  $X_t$ , o en altres situacions que puguin resultar d'interès en cada cas.

En aquest exemple, el valor més baix de l'estadístic s'assoleix sota la hipòtesi que  $X$  segueix una distribució log-Uniforme, obtenint un valor de  $W_n^2 = 0,065$ . A la figura 3.8 podem veure que la funció de distribució espectral log-Normal i

### 3.5.2. Deflactor del producte interior brut (PIB) espanyol

---

Taula 3.3: p-valors corresponents a l'exemple del PIB espanyol

$\alpha$	0,5	0,6	0,7	0,8	0,9
p-valor	0,08	0,07	0,04	0,02	0,002

la funció de distribució espectral mostral estan considerablement allunyades, de forma coherent amb el resultat del contrast, mentre que la funció de distribució espectral log-Uniforme es troba més propera. A la taula 3.3 podem veure els

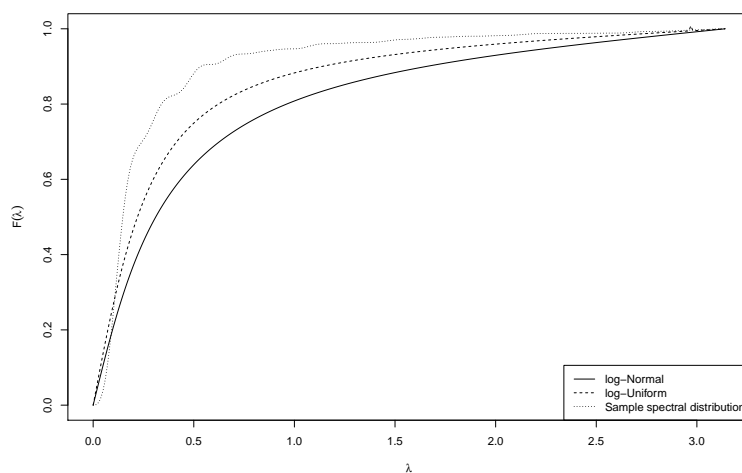


Figura 3.8: Comparació de les distribucions espectrals mostrals i teòriques (PIB espanyol)

p-valors del contrast per aquest exemple, per diversos valors de  $\alpha$  propers al valor estimat per a la sèrie. La hipòtesi nul·la de normalitat de les innovacions pot ser rebutjada per valors de  $\alpha = 0,7$ ,  $\alpha = 0,8$  and  $\alpha = 0,9$ .

## 3.6 Estudi de la potència del contrast

L'objectiu d'aquesta secció és discutir la potència del contrast en algunes situacions particulars, concretament, quan les innovacions segueixen una distribució uniforme, Laplace, logística o gamma.

La taula 3.4 s'ha obtingut simulant 5000 sèries amb innovacions seguint les distribucions especificades en cada cas i calculant el valor  $W_n^2$  corresponent, i considerant el nombre de casos en que aquest valor està per sobre del valor crític corresponent per al cas normal. La primera fila de la taula 3.4 correspon al cas uniforme, la segona fila correspon al cas Laplace, la tercera fila correspon a la distribució logística, i la darrera fila correspon a innovacions seguint una distribució gamma. S'han considerat escenaris amb diverses mides mostrals (des de  $n = 20$  fins a  $n = 500$ ), i diferents valors del primer coeficient d'autocorrelació (des de  $\alpha = -0,8$  fins a  $\alpha = 0,8$ ).

El cas de les innovacions gamma és especial en el sentit que per tal d'estandarditzar  $Y$  s'ha hagut de considerar  $W = Z - k$ , on  $Z$  segueix una distribució gamma i  $k$  és una constant. En particular, si  $\theta$  és el paràmetre d'escala de la distribució gamma, tenim que el paràmetre de forma és  $\frac{1-\alpha^2}{\theta^2}$  i la constant pren el valor  $k = \frac{1-\alpha^2}{\theta}$ .

És interessant destacar que els valors de potència del contrast que es mostren a la taula 3.4 varien si aixequem  $X$  a un determinat exponent, o equivalentment, si considerem una estandardització alternativa. El càlcul de les potències descrites s'ha realitzat en tots els casos considerant la sèrie  $X$  original, fixant els paràmetres  $\alpha$  i  $\sigma$  obtinguts en les 5000 sèries simulades i contrastant amb 5000 sèries amb

### 3.7. Conclusions

---

Taula 3.4: Anàlisi de la potència per innovacions uniformes, Laplace, logística i gamma

Distribució	$\alpha / n$	-0,8	-0,5	-0,2	0,2	0,5	0,8
Uniforme	20	4	7	8	6	5	6
Laplace		4	5	4	5	7	6
Logística		5	5	5	6	5	5
Gamma		2	1	2	1	7	4
Uniforme	50	7	12	10	8	6	4
Laplace		4	3	5	4	12	7
Logística		5	4	5	5	7	5
Gamma		1	1	4	1	20	11
Uniforme	100	6	20	17	8	9	4
Laplace		5	8	5	4	16	8
Logística		4	5	5	4	9	6
Gamma		6	27	5	2	40	20
Uniforme	200	7	34	34	11	11	5
Laplace		7	25	7	6	25	12
Logística		6	10	5	6	11	8
Gamma		20	73	12	10	63	31
Uniforme	500	10	84	82	22	17	6
Laplace		19	58	50	38	48	19
Logística		8	25	21	16	22	8
Gamma		60	99	93	74	87	62

innovacions normals fixant també els valors dels paràmetres  $\alpha$  i  $\sigma$  obtinguts, d'acord amb els passos descrits en l'apartat anterior.

## 3.7 Conclusions

Els resultats presentats en la primera part d'aquest capítol permeten caracteritzar la distribució de les innovacions d'un procés autoregressiu de primer ordre en unes



certes condicions prou generals, de manera que donada una sèrie temporal amb estructura AR(1)  $Y_t = \alpha Y_{t-1} + W_t$  podrem triar una distribució adequada per  $W_t$ .

A més, s'ha desenvolupat un contrast de bondat de l'ajust que pot utilitzar-se per contrastar la normalitat de les innovacions o en un context més general, per contrastar si les innovacions d'un procés autoregressiu de primer ordre segueixen una determinada distribució.

Els resultats obtinguts en els exemples mostren que de vegades no és massa realista considerar que les innovacions d'un procés AR(1) segueixen una distribució gaussiana, i resulta més adequat buscar altres distribucions alternatives que s'adaptin millor al procés.

En la part final del capítol es presenta l'estudi de la potència del contrast de bondat de l'ajust, per a innovacions seguint diferents distribucions i per diferents valors de  $\alpha$  i mides mostrals. En aquest estudi de la potència veiem que en general, els millors resultats s'obtenen per mides mostrals grans, el que és raonable, i per valors de  $\alpha$  allunyats de -1 i 1. Quant a les distribucions de les innovacions, veiem que les innovacions amb distribució uniforme són les que presenten un rendiment més discret, mentre que els millors resultats s'obtenen per les distribucions gamma i Laplace.

### 3.7. Conclusions

---

---

## BIBLIOGRAFIA

- [1] B. Abraham and N. Balakrishna. Inverse gaussian autoregressive models. *Journal of time series analysis*, 20(6):605–618, 1999.
- [2] T. W. Anderson. Goodness-of-fit tests for autoregressive processes. *Journal of time series analysis*, 18(4):321–339, 1997.
- [3] T. W. Anderson. Goodness-of-fit tests for spectral distributions. *Annals of Statistics*, 21:830–847, 1997.
- [4] T. W. Anderson, R. A. Lockhart, and M. A. Stephens. Goodness-of-fit tests for the time series models AR(1) and MA(1). *Research report, Simon Fraser University*, 97–05, 1995.
- [5] T. W. Anderson, R. A. Lockhart, and M. A. Stephens. An omnibus test for the time series model AR(1). *Journal of econometrics*, 118:111–127, 2004.

- [6] E. Artin. *The gamma function*. Holt, Rinehart, Winston, 1964.
- [7] P. M. Batchelder. *An introduction to linear difference equations*. Dover publications, 1967.
- [8] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer-Verlag, 1991.
- [9] R. Bu, K. Hadri, and B. McCabe. Conditional maximum likelihood estimation of higher-order integer-valued autoregressive processes. *Journal of time series analysis*, 29:973–994, 2008.
- [10] R. Bu and B. McCabe. Model selection, estimation and forecasting in INAR(p) models: A likelihood-based markov chain approach. *International journal of forecasting*, 24:151–162, 2008.
- [11] M. Cardinal, R. Roy, and J. Lambert. On the application of integer-valued time series models for the analysis of disease incidence. *Statistics in medicine*, 18(15):2025–2039, Aug 15 1999. LR: 20061115; CI: Copyright 1999; JID: 8215016; ppublish.
- [12] G. Casella and R. L. Berger. *Statistical inference*. Duxbury Thompson Learning, 2002.
- [13] R. C. Chambers. Variation in and among early life history traits of laboratory-reared winter flounder *pseudopleuronectes americanus*. *Marine ecology progress series*, 47:1–15, 1988.

- [14] S. Csörgö and J. J. Faraway. The exact and asymptotic distributions of Cramér-von Mises statistics. *J. R. Statist. Soc. B*, 58(1):221–234, 1996.
- [15] M. Falch, F. Marohn, R. Michel, D. Hofmann, M. Macke, C. Spachmann, and Englert S. *A first course on time series analysis - Examples with SAS*. Chair of statistics, University of Würzburg, 2011.
- [16] R. K. Freeland and B. McCabe. Forecasting discrete valued low count time series. *International Journal of Forecasting*, 20:427–434, 2004.
- [17] D. P. Gaver and P. A. W. Lewis. First-order autoregressive gamma sequences and point processes. *Advances in applied probability*, 12:727–745, 1980.
- [18] D. Gomes and L. Canto e Castro. Generalized integer-valued random coefficient for a first order structure autoregressive (RCINAR) process. *J. Statist. Plann. Inference*, 139(12):4088–4097, 2009.
- [19] C. W. J. Granger and P. Newbold. Forecasting transformed series. *Journal of the Royal Statistical Society B*, 2:189–203, 1976.
- [20] B. T. Grenfell, O. N. Bjørnstad, and J. Kappey. Travelling waves and spatial hierarchies in measles epidemics. *Nature*, 414:716–723, 2001.
- [21] T. Hampton. Pandemic flu planning falls short many vulnerabilities in health care system. *Journal of American Medical Association*, 297:1177–1178, 2007.
- [22] J. L. Hay and A. N. Pettitt. Bayesian analysis of a time series of counts with covariates: an application to the control of an infectious disease. *Biostatistics*, 2(4):433–444, 2001.

- [23] R. C. Jung and A. R. Tremayne. Binomial thinning models for integer time series. *Stat. Model.*, 6(2):81–96, 2006.
- [24] M. J. Keeling and L. Danon. Mathematical modelling of infectious diseases. *British Medical Bulletin*, 92:33–42, 2009.
- [25] M. Kuczma. *Functional equations in a single variable*. Polish scientific publishers, 1968.
- [26] M. Kuczma. Problems of uniqueness in the theory of functional equations in a single variable. *Prace matematyczne zeszyt*, 14(223), 1969.
- [27] M. Kuczma, B. Choczewski, and R. Ger. *Iterative functional equations*. Cambridge University press, 1990.
- [28] F. Laio. Cramer-von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters. *Water Resour. Res.*, 40, 2004.
- [29] E. L. Lehmann and G. Casella. *Theory of point estimation*. Springer-Verlag, 1998.
- [30] E. Lukacs. *Characteristic functions*. Hafner publishing company, 1970.
- [31] E. McKenzie. Product autoregression: A time-series characterization of the gamma distribution. *Journal of applied probability*, 19:463–468, 1982.

- [32] E. McKenzie. Discrete variate time series. In *Stochastic processes: modelling and simulation*, volume 21 of *Handbook of Statist.*, pages 573–606. North-Holland, Amsterdam, 2003.
- [33] K. S. Miller. *Linear difference equations*. W. A. Benjamin, 1968.
- [34] M. Monteiro, M. G. Scotto, and I. Pereira. Integer-valued autoregressive processes with periodic structure. *J. Statist. Plann. Inference*, 140(6):1529–1541, 2010.
- [35] D. Moriña, P. Puig, J. Rios, A. Vilella, and A. Trilla. A statistical model for hospital admissions caused by seasonal diseases. *Statistics in medicine*, 30(26):3125–3136, Nov 20 2011.
- [36] P. Puig and J. Valero. Characterization of count data distributions involving additivity and binomial subsampling. *Bernoulli*, 13(2):544–555, 2007.
- [37] P. Quénel and W. Dab. Influenza A and B epidemic criteria based on time-series analysis of health services surveillance data. *European Journal of Epidemiology*, 5:285–293, 1998.
- [38] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.
- [39] C. R. Rao. *Linear statistical inference and its applications*. John Wiley & sons, 1973.
- [40] R. J. Serfling. *Approximation Theorems of Mathematical Statistics*. John Wiley & sons, 1980.

- [41] D. N. Shanbhag and M. Sreehari. On certain self-decomposable distributions. *Probability theory and related fields*, 38:217–222, 1977.
- [42] R. H. Shumway and D. S. Stoffer. *Time Series analysis and its applications: With R examples*. Springer-Verlag, 2010.
- [43] C. H. Sim. Modelling non-normal first-order autoregressive time series. *Journal of forecasting*, 13:369–381, 1994.
- [44] A. Sumi, K. Kamo, N. Ohtomo, K. Mise, and N. Kobayashi. Time series analysis of incidence data of influenza in japan. *Journal of Epidemiology*, 21:21–29, 2011.
- [45] C. H. Weiß. Thinning operations for modeling time series of counts - a survey. *Advances in statistical analysis*, 92:319–341, 2008.
- [46] C. Wiuf and P. H. Stumpf. Binomial subsampling. *Proceedings of the Royal Society A*, 462:1181–1195, 2006.
- [47] R. Zhu and H. Joe. Modelling count data time series with Markov processes based on binomial thinning. *J. Time Ser. Anal.*, 27(5):725–738, 2006.