

Series de Tiempo

Germán Aneiros Pérez

Departamento de Matemáticas
Universidade da Coruña

Máster Universitario en Técnicas Estadísticas
Curso 2022-23

*“Time series are analysed to **understand** the past and to **predict** the future, enabling managers or policy markets to make properly informed decisions”*

Paul S.P. Cowpertwait y Andrew V. Metcalfe (2009)

Bibliografía básica



COWPERTWAIT, P.S.P. Y METCALFE, A.V. (2009).
Introductory Time Series with R. Springer.



CRYER, J.D. Y CHAN, K-S (2008).
Time Series Analysis. With Applications in R. Springer (2ª edición).



HYNDMAN R.J. Y ATHANASOPOULOS, G. (2018).
Forecasting: principles and practice. O Texts: Melbourne, Australia (2ª edición). Accesible online en <https://otexts.com/fpp2/>

Bibliografía complementaria



BROCKWELL, P.J. Y DAVIS, R.A. (2002).
Introduction to Time Series and Forecasting. Springer (2ª edición).



PEÑA, D. (2005).
Análisis de Series Temporales. Alianza Editorial.



SHUMWAY, R.H. Y STOFFER, D.S. (2017).
Time Series Analysis and Its Applications. With R Examples. Springer (4ª edición).

R

- Paquete **fpp2** (CRAN).
- Paquete **TSA** (CRAN).
- Paquete **tseries** (CRAN).
- Módulo **dynamic-arimax** (GitHub).
(<https://github.com/anaezquerro/dynamic-arimax>)

Parte I

Series de tiempo y procesos estocásticos

Serie de tiempo

Una **serie de tiempo** es una colección de observaciones de una variable, X , tomadas secuencialmente a lo largo del tiempo.

Por ejemplo:

- X : “Nº de turistas que visitan España”.
Frecuencia de observación: mensual.
Serie de tiempo: 2698170, 2821540, 3310270, ...
- X : (“Euríbor”, “Cotización del barril de crudo”).
Frecuencia de observación: diaria.
Serie de tiempo: (4.32 , 87.95), (4.33 , 88.40), ...

A lo largo de la mayor parte de este curso nos centraremos en variables unidimensionales; por lo tanto, trataremos con series de tiempo univariantes.

Las observaciones las tomaremos en intervalos regulares de tiempo (cada hora, cada día, cada mes, ...).

Sin pérdida de generalidad, supondremos que la variable X ha sido observada en los instantes $1, 2, \dots, T$.

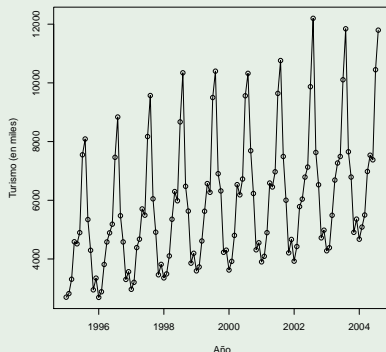
La serie de tiempo observada de la variable X será representada por x_1, x_2, \dots, x_T .

Así, en el ejemplo referente a la variable X : “Nº de turistas que visitan España”, tendríamos que:

$$x_1 = 2698170, x_2 = 2821540, x_3 = 3310270, \dots$$

Para analizar una serie de tiempo, lo primero que se debe hacer es representar su **gráfico secuencial**; esto es, representar gráficamente cada observación, x_t , frente al instante, t , en que se observa, y luego unir con segmentos cada uno de los T puntos (t, x_t) .

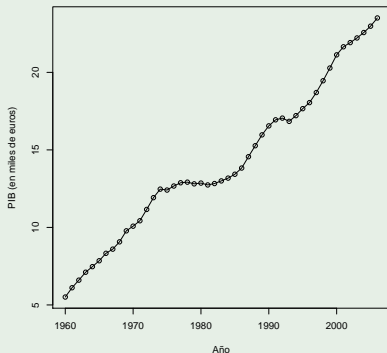
Turismo: gráfico secuencial



El gráfico secuencial permite observar cómo evoluciona la serie a lo largo del tiempo; específicamente, muestra las principales características de la serie de tiempo, que son:

- Posible presencia de **tendencia**: comportamiento a largo plazo de la serie (si hay tendencia, el nivel de la serie no es constante; esto es, cambia).
- Posible presencia de **estacionalidad**: comportamiento “periódico” de la serie (si hay estacionalidad, la serie presenta un patrón “repetitivo”; dicho patrón surge cuando la serie de tiempo se ve afectada por factores estacionales, como por ejemplo el mes del año o el día de la semana. La estacionalidad es siempre de una frecuencia fija y conocida).
- Posible presencia de **heterocedasticidad**: la variabilidad de la serie no es constante (por ejemplo, depende de su nivel).

Producto interior bruto per capita (PIB)



Información sobre la serie

Lugar: España

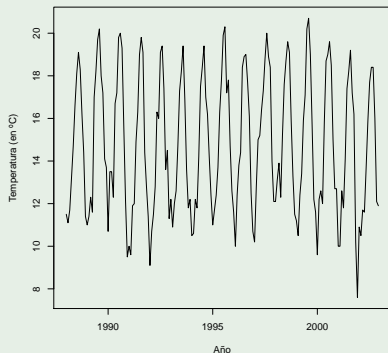
Frecuencia de observ.: Anual

Cantidad de observaciones: 47
(1960-2006)

La serie presenta **tendencia** (su nivel cambia, no es constante).

`plot(...)`

Temperatura media



Información sobre la serie

Lugar: A Coruña

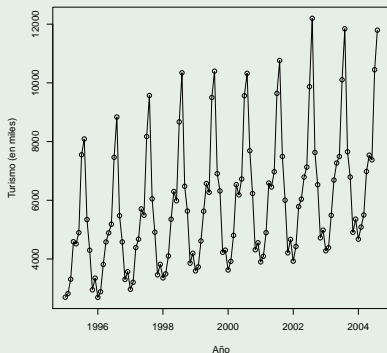
Frecuencia de observ.: Mensual

Cantidad de observaciones: 180
(enero 1988-diciembre 2002)

La serie presenta **componente estacional** (su evolución sigue un patrón “repetitivo”).

`plot(...)`

Turismo



Información sobre la serie

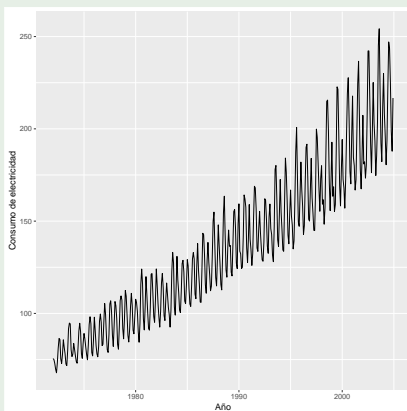
Lugar: España

Frecuencia de observ.: Mensual

Cantidad de observaciones: 116
(enero 1995-agosto 2004)

La serie presenta tendencia y componente estacional.

Consumo de electricidad



Información sobre la serie

Lugar: EE.UU.

Frecuencia de observ.: Mensual

Cantidad de observaciones: 396
(enero 1972-diciembre 2004)

La serie presenta tendencia, componente estacional y **heterocedasticidad** (su variabilidad cambia; en este ejemplo, la variabilidad aumenta al aumentar el nivel).

```
autoplot(...)
```

El objetivo principal de este curso es la **inferencia estadística** en el contexto de las series de tiempo.

En concreto, dada una serie de tiempo se pretende **predecir** sus observaciones futuras.

Para poder hacer inferencia estadística, supondremos que la serie de tiempo, x_1, \dots, x_T (las observaciones fijas de que disponemos), ha sido generada por un **modelo estocástico**.

La conjunción de las propiedades del modelo estocástico y de la información dada por la serie de tiempo nos permitirá cumplir nuestro objetivo (predicción).

Las etapas a seguir serán:

- 1 Identificar el modelo estocástico.
- 2 Estimar el modelo estocástico.
- 3 Chequear el modelo estimado.
- 4 Predecir el comportamiento del modelo estocástico.

Proceso estocástico

Un **proceso estocástico** es un conjunto de **variables aleatorias**, $\{X_t\}_{t \in C}$, definidas sobre el mismo espacio de probabilidad.

Nos centraremos en el proceso estocástico $\{X_t\}_{t \in \mathbb{Z}}$, siendo \mathbb{Z} el conjunto de los números enteros; esto es, nuestro proceso estocástico será

$$\dots, X_{-2}, X_{-1}, X_0, X_1, X_2, \dots$$

El subíndice t de cada variable aleatoria representa el instante de tiempo en que es observada.

Los procesos estocásticos forman el **marco teórico** en el que nos situaremos en lo que resta de curso.

Una observación del proceso estocástico será denotada por

$$\dots, X_{-2}, X_{-1}, X_0, X_1, X_2, \dots$$

y se conoce como una **realización** o **trayectoria** del mismo.

La serie de tiempo

$$X_1, X_2, \dots, X_T$$

es, por tanto, una **realización** o **trayectoria parcial** de un proceso estocástico.

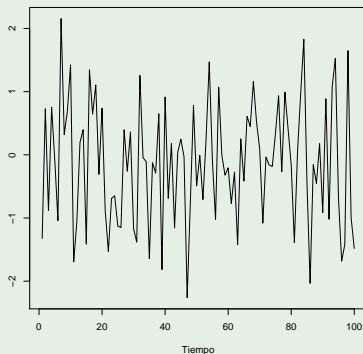
Ejemplo: Ruido blanco

Es una colección de v.a. incorreladas, con media 0 y varianza finita σ_a^2 .

Se denotará por $\{a_t\}_t$.

Nota: Si el ruido blanco es **gaussiano**, entonces las v.a. que lo conforman son **i.i.d.**

Serie de tiempo generada por un proceso de ruido blanco gaussiano con $\sigma_a^2 = 1$



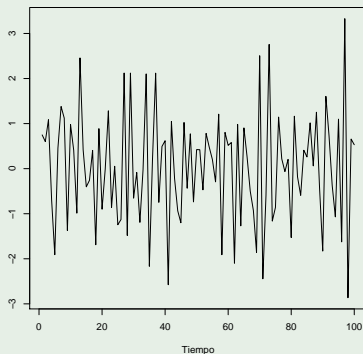
Ejemplo: MA(1)

$$X_t = c + a_t + \theta_1 a_{t-1},$$

donde $\{a_t\}_t$ es ruido blanco, y c y θ_1 ($\theta_1 \neq 0$) son parámetros.

Nota: Existe correlación entre X_t y X_{t-1} , pues a ambas les afecta (linealmente) a_{t-1} . Sin embargo, no existe correlación entre X_t y X_{t-i} , siendo $|i| > 1$.

Serie de tiempo generada por un MA(1) ($c = 0$, $\theta_1 = -0.6$ y $\{a_t\}_t$ gaussiano con $\sigma_a^2 = 1$)



Ejemplo: Paseo aleatorio

$$X_t = c + X_{t-1} + a_t,$$

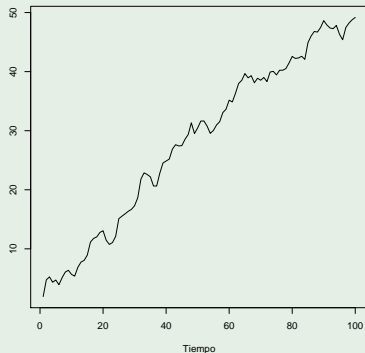
donde $\{a_t\}_t$ es ruido blanco.

Nota 1: Existe correlación entre X_t y X_{t-i} para cualquier valor de i .

Nota 2: Si el paseo aleatorio comienza en $t = 0$ y se establece la condición inicial $X_0 = 0$, entonces la anterior expresión puede escribirse como

$$X_t = ct + \sum_{j=1}^t a_j, \quad \forall t \geq 1.$$

Serie de tiempo generada por un paseo aleatorio ($X_0 = 0$, $c = 0.5$ y $\{a_t\}_t$ gaussiano con $\sigma_a^2 = 1$)



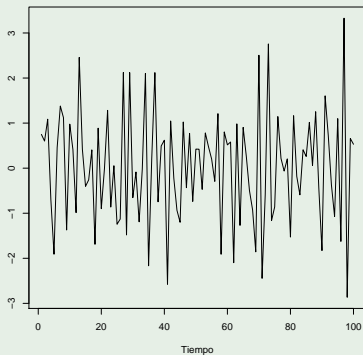
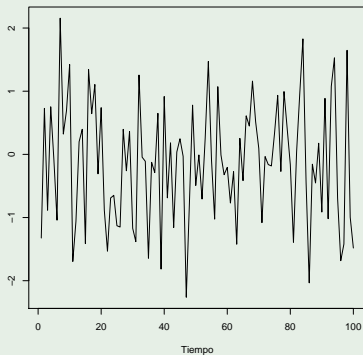
En los 3 ejemplos anteriores partimos del **conocimiento** de un **proceso estocástico**:

- Ruido blanco con $\sigma_a^2 = 1$,
- MA(1) con $c = 0$, $\theta_1 = -0.6$ y $\sigma_a^2 = 1$,
- Paseo aleatorio con $c = 0.5$ y $\sigma_a^2 = 1$,

y generamos una serie temporal a partir de él.

En la práctica, la situación es inversa. Partimos del **conocimiento** de una **serie temporal** y tratamos de “**descubrir**” el proceso estocástico que la ha generado (posteriormente, basándonos en tal proceso, realizaremos predicciones).

Por ejemplo, a partir de las siguientes series de tiempo, habría que descubrir qué proceso estocástico ha generado a cada una: ¡ruido blanco gaussiano con $\sigma_a^2 = 1$ (izquierda) y MA(1) con $c = 0, \theta_1 = -0.6$ y $\{a_t\}_t$ gaussiano con $\sigma_a^2 = 1$ (derecha)!



“Descubrir” el proceso estocástico que ha generado a una serie de tiempo supone:

- 1 **Identificarlo**. Por ejemplo, indicar que estamos ante una serie generada por un MA(1): $X_t = c + a_t + \theta_1 a_{t-1}$.
- 2 **Estimar** sus parámetros (en el caso del MA(1), estimar los valores de c , θ_1 y σ_a^2).
- 3 **Chequear** el modelo estimado.

Matizaciones:

- La pretensión de “descubrir el proceso estocástico que ha generado a una serie de tiempo” debe ser considerada en su justa medida.
- Una serie de tiempo de datos reales habrá sido generada por la Naturaleza, por distintos factores económicos,... y por tanto no existe un proceso estocástico que la haya generado (la Naturaleza no se rige por una ecuación...).
- Lo que se pretende es construir un proceso estocástico sencillo que de manera razonable pueda generar una serie de tiempo con las características de la nuestra.
- En general, hay más de un proceso estocástico susceptible de haber generado a una serie temporal.

Observaciones:

- La tarea de construir un proceso estocástico que razonablemente haya podido generar a la serie de tiempo en estudio es, en principio, una tarea ardua.
- Basta pensar que para ello únicamente disponemos de la información suministrada por la serie: x_1, \dots, x_T ; esto es, únicamente disponemos de un valor de cada variable aleatoria X_1, \dots, X_T del proceso $\{X_t\}_{t \in \mathbb{Z}}$.

A continuación, definiremos varias medidas características y propiedades de interés de un proceso estocástico. La posesión de algunas de dichas propiedades facilitará/posibilitará nuestra tarea.

Algunas definiciones

Dado un proceso estocástico $\{X_t\}_t$, se definen:

- **Función de medias:** $\mu_t = E(X_t)$

Es una medida de posición de carácter central de X_t .

- **Función de varianzas:** $\sigma_t^2 = Var(X_t) = E((X_t - \mu_t)^2)$

Es una medida del grado de variabilidad de X_t .

- **Función de autocovarianzas:**

$$\gamma(s, t) = Cov(X_s, X_t) = E((X_s - \mu_s)(X_t - \mu_t))$$

Es una medida del grado de dependencia lineal existente entre X_s y X_t .

- Función de autocorrelaciones simples (fas):

$$\rho(s, t) = \frac{\gamma(s, t)}{\sigma_s \sigma_t}$$

Es una medida del grado de dependencia lineal existente entre X_s y X_t . Toma valores en $[-1, 1]$.

- Función de autocorrelaciones parciales (fap):

$$\alpha(s, t) = \frac{\text{Cov} \left(X_s - \hat{X}_s^{(s,t)}, X_t - \hat{X}_t^{(s,t)} \right)}{\sqrt{\text{Var} \left(X_s - \hat{X}_s^{(s,t)} \right) \text{Var} \left(X_t - \hat{X}_t^{(s,t)} \right)}},$$

donde $\hat{X}_j^{(s,t)}$ denota al mejor predictor lineal de X_j construido a partir de las variables medidas en los instantes comprendidos entre s y t (sin considerar ni s ni t).

$\alpha(s, t)$ es una medida del grado de **dependencia lineal** existente entre X_s y X_t , una vez que se les ha **sustraído el efecto lineal** que sobre cada una de ellas ejercen las variables medidas en los instantes comprendidos entre s y t . Toma valores en $[-1, 1]$.

Ejemplo: Autocorrelación simple y parcial

Supongamos que X_1, X_2 y X_3 miden la temperatura de ayer, hoy y mañana, respectivamente.

- $\rho(1, 3)$ nos orienta acerca de la relación lineal existente entre la temperatura de ayer y la de mañana.
- Es posible que exista dependencia lineal entre la temperatura de hoy y la ayer, y también entre la temperatura de hoy y la de mañana.
 $\alpha(1, 3)$ nos orienta acerca de la relación lineal existente entre las temperaturas de ayer y mañana, una vez que se ha extraído de ambas la influencia lineal de la temperatura de hoy.

Las funciones que acabamos de definir son de interés porque sus valores nos informan sobre distintas características del proceso estocástico que estamos tratando de construir (aquel susceptible de haber generado a la serie de tiempo).

Sin embargo, nos enfrentamos a varias limitaciones:

- Dichas funciones dependen del proceso estocástico, que desconocemos. Por tanto, debemos estimarlas.
- Únicamente disponemos de una observación x_t de cada v.a. X_t ($t = 1, \dots, T$), lo que dificulta dicha estimación.

Por tanto, debemos imponer condiciones/restricciones que nos permitan estimar dichas características a partir de la serie de tiempo.

Proceso estacionario

Un proceso estocástico $\{X_t\}_t$ se dice que es estacionario si se verifican:

- 1 $\mu_t = \mu, \forall t.$
- 2 $\sigma_t^2 = \sigma^2, \forall t.$
- 3 $\gamma(t, t + k) = \gamma_k, \forall t, k.$

La estacionariedad dota al proceso estocástico de [propiedades de estabilidad](#) en la media, en la varianza y en las autocovarianzas. Estas propiedades nos permitirán estimar distintas características del proceso a partir de la serie de tiempo x_1, \dots, x_T .

A partir de ahora, ρ_k y α_k denotarán a las autocorrelaciones simple y parcial, respectivamente, existentes entre dos variables de un [proceso estacionario](#) separadas k instantes de tiempo.

Estimaciones asociadas a la serie x_1, \dots, x_T :

- **Media muestral:** $\bar{x} = \frac{\sum_{t=1}^T x_t}{T}$

- **Función de autocovarianzas muestrales:**

$$\hat{\gamma}_k = \frac{\sum_{t=1}^{T-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{T}$$

con $\hat{\gamma}_{-k} = \hat{\gamma}_k$ para $k = 0, 1, \dots, T-1$.

- **Función de autocorrelaciones simples muestrales:**

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}$$

- **Función de autocorrelaciones parciales muestrales:** $\hat{\alpha}_k = \hat{\alpha}_{kk}$, donde $\hat{\alpha}_{kk}$ es el estimador mínimo cuadrático de α_{kk} en la regresión

$$x_t = \alpha_{k0} + \alpha_{k1}x_{t-1} + \dots + \alpha_{kk}x_{t-k} + \text{error}.$$

Ejemplo: Ruido blanco

$\{a_t\}_t$ v.a. incorreladas, con media 0 y varianza finita σ_a^2 .

Se tiene que:

❶ $\mu_t = 0, \forall t.$

❷ $\sigma_t^2 = \sigma_a^2, \forall t.$

❸ $\gamma(t, t+k) = E(a_t a_{t+k}) = \begin{cases} \sigma_a^2, & \text{si } k = 0 \\ 0, & \text{si } k \neq 0 \end{cases}$

Conclusión: El ruido blanco es un proceso estacionario en el que:

• $\mu = 0.$

• $\sigma^2 = \sigma_a^2.$

• $\gamma_k = \begin{cases} \sigma_a^2, & \text{si } k = 0 \\ 0, & \text{si } k \neq 0 \end{cases}$

Ejemplo: Paseo aleatorio con $X_0 = 0$ y comienzo en $t = 0$

$X_t = ct + \sum_{j=1}^t a_j$, donde $\{a_t\}_t$ es ruido blanco y c es un parámetro.

Se tiene que:

① $\mu_t = ct, t = 1, 2, \dots$

② $\sigma_t^2 = \sigma_a^2 t, t = 1, 2, \dots$

③ $\gamma(t, t+k) = E\left(\sum_{i=1}^t a_i \sum_{j=1}^{t+k} a_j\right) = \min\{t, t+k\}\sigma_a^2,$
 $\forall t \geq 1, \forall k \geq 1 - t.$

Conclusión: El paseo aleatorio no es un proceso estacionario. Aunque para el caso particular $c = 0$ es estable en la media, nunca lo es en la varianza (su varianza es explosiva; esto es, tiende al infinito con t) y tampoco en las autocovarianzas.

Más definiciones

Sean $\{a_t\}_t$ un proceso de ruido blanco y $\{X_t\}_t$ un proceso estocástico. Se definen:

- **Proceso lineal:** Aquél que admite una representación del tipo

$$X_t = \cdots + \psi_{-1}a_{t+1} + c + \psi_0a_t + \psi_1a_{t-1} + \cdots$$

$$\text{con } \sum_{i=-\infty}^{\infty} |\psi_i| < \infty.$$

Nota: Los procesos lineales son estacionarios, siendo su función de autocovarianzas:

$$\gamma_k = \sigma_a^2 \sum_{i=-\infty}^{\infty} \psi_i \psi_{i+k}.$$

- **Proceso causal** (o $MA(\infty)$): Aquél que admite una representación del tipo

$$X_t = c + \psi_0 a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots$$

$$\text{con } \sum_{i=0}^{\infty} |\psi_i| < \infty.$$

- **Proceso invertible** (o $AR(\infty)$): Aquél que admite una representación del tipo

$$X_t = c + a_t + \pi_1 X_{t-1} + \pi_2 X_{t-2} + \cdots$$

$$\text{con } \sum_{i=1}^{\infty} |\pi_i| < \infty.$$

A continuación presentamos un importante resultado:

Descomposición de Wold

Si el proceso estocástico $\{X_t\}_t$ es estacionario y no contiene componentes deterministas (una constante, una función del tiempo, etc), entonces admite una representación del tipo

$$X_t = \psi_0 a_t + \psi_1 a_{t-1} + \cdots$$

$$\text{con } \psi_0 = 1 \text{ y } \sum_{i=0}^{\infty} \psi_i^2 < \infty.$$

Nota: El resultado anterior garantiza que cualquier proceso estacionario o bien es lineal o bien puede ser transformado para que lo sea (sin más que sustraerle la componente determinista). Tenemos por tanto que la clase de los procesos lineales forma un marco general para el estudio de los procesos estacionarios.

A lo largo de este tema:

- Se ha establecido el marco teórico en el que nos situaremos a lo largo del presente curso: los **procesos estocásticos**.
- Se ha definido un proceso estocástico que será fundamental en nuestros estudios: el **ruido blanco**.
- Se han definido e interpretado distintas funciones asociadas a un proceso estocástico: **fas**, **fap**,...
- Se ha definido el concepto de **proceso estacionario**.
- Se han propuesto estimadores de distintas funciones asociadas a un proceso estocástico: **fas y fap muestrales**,...
- Se han definido propiedades de interés de un proceso estocástico: **linealidad**, **causalidad** e **invertibilidad**.
- Se ha enunciado e interpretado la **descomposición de Wold**.

Parte II

Modelos Box-Jenkins

Comenzamos recordando la notación general:

- x_1, x_2, \dots, x_T : serie de tiempo.
- $\{X_t\}_t$: proceso generador de la serie de tiempo.
- $\{a_t\}_t$: proceso de ruido blanco (**innovaciones**).

Además, supondremos que:

- a_t es **independiente** de X_{t-1}, X_{t-2}, \dots .

Esto implica que el conocimiento de los valores de X_{t-1}, X_{t-2}, \dots no aporta información acerca de a_t .

El primer objetivo de este tema es la **construcción** de un modelo estocástico sencillo que, de forma razonable, haya podido generar a la serie de tiempo de que disponemos.

Entonces, basándonos en dicho modelo, podremos:

- Comprender la **dinámica/evolución** de la serie de tiempo.
- Efectuar **predicciones** de valores futuros de la serie de tiempo.

La planificación del presente tema puede resumirse en las siguientes etapas:

- 1 Presentación de una **clase de modelos estocásticos** paramétricos pero flexibles (han demostrado su utilidad como posibles generadores de series de tiempo reales).
- 2 Exposición de técnicas que nos permitan **identificar** alguno de dichos modelos como posible generador de nuestra serie de tiempo.
- 3 Construcción de **estimadores** para los parámetros del modelo.
- 4 **Diagnosis** o chequeo del modelo.
- 5 **Predicción** de futuros valores de la serie en base al modelo estimado/ajustado.

Proceso autorregresivo de orden 1

Un proceso **estacionario**, $\{X_t\}_t$, que admite la representación

$$X_t = c + \phi_1 X_{t-1} + a_t,$$

donde c y ϕ_1 ($\phi_1 \neq 0$) son constantes, se conoce como un “proceso autorregresivo de orden 1”, y se denota por **AR(1)**.

Se verifica que:

- El proceso AR(1) explica el valor actual (X_t) como una función lineal de 1 valor pasado (X_{t-1}).
- El proceso AR(1) siempre es invertible.

Ejemplo

Imaginemos que deseamos construir un modelo para la cantidad de agua (X_t) que hay al final del mes t en un embalse. Puesto que dicha cantidad es aleatoria, el modelo ha de ser un modelo estocástico. Haremos las siguientes suposiciones:

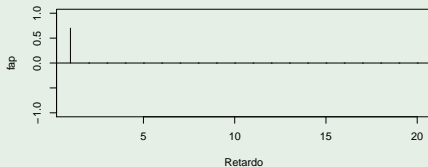
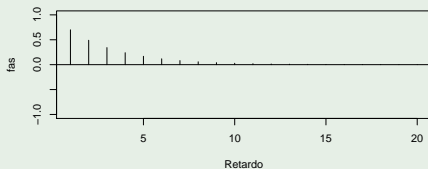
- Durante el mes t llega al embalse una cantidad de agua $c + a_t$, siendo c la cantidad media de agua que llega y a_t una v.a. de media cero y varianza constante que hace que la entrada de agua varíe de un mes a otro.
- Cada mes se gasta una proporción fija $1 - \phi_1$ de las existencias iniciales.

Se tiene que el modelo buscado es:

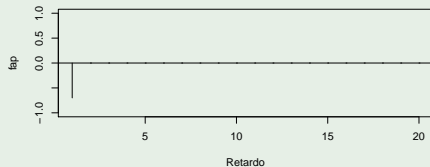
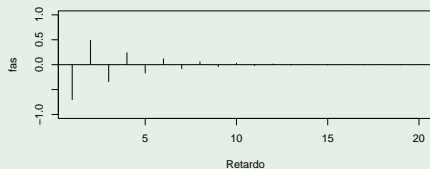
$$X_t = c + \phi_1 X_{t-1} + a_t.$$

Ejemplos de la fas y la fap de procesos AR(1):

AR(1): $\phi_1 > 0$



AR(1): $\phi_1 < 0$



Proceso autorregresivo de orden 2

Un proceso **estacionario** $\{X_t\}_t$ que admite la representación

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + a_t,$$

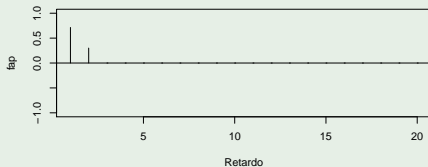
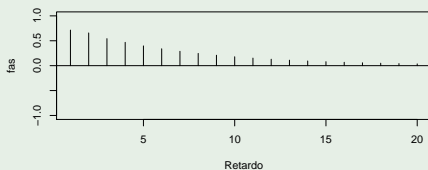
donde c, ϕ_1 y ϕ_2 ($\phi_2 \neq 0$) son constantes, se conoce como un “proceso autorregresivo de orden 2”, y se denota por **AR(2)**.

Se verifica que:

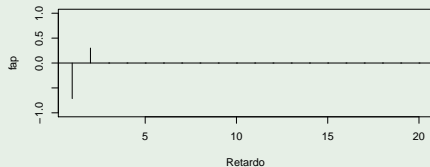
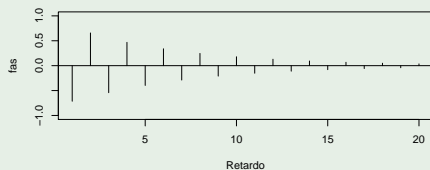
- El proceso AR(2) explica el valor actual (X_t) como una función lineal de 2 valores pasados (X_{t-1} y X_{t-2}).
- El proceso AR(2) siempre es invertible.

Ejemplos de la fas y la fap de procesos AR(2):

AR(2): $\phi_1 > 0, \phi_2 > 0$

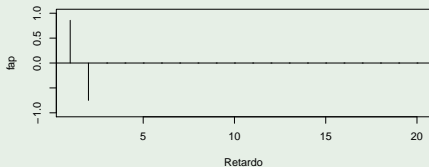
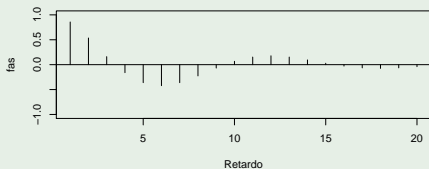


AR(2): $\phi_1 < 0, \phi_2 > 0$

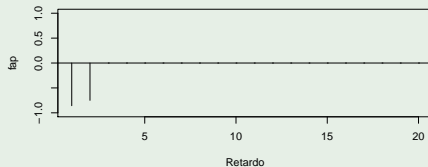
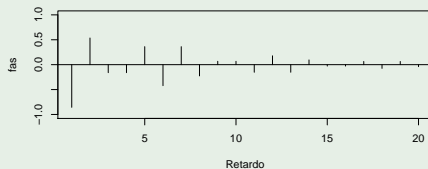


Ejemplos de la fas y la fap de procesos AR(2):

AR(2): $\phi_1 > 0, \phi_2 < 0$



AR(2): $\phi_1 < 0, \phi_2 < 0$



Proceso autorregresivo de orden p

Un proceso **estacionario** $\{X_t\}_t$ que admite la representación

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + a_t,$$

donde c, ϕ_1, \dots, ϕ_p ($\phi_p \neq 0$) son constantes, se conoce como un “proceso autorregresivo de orden p ” (**AR(p)**).

Se verifica que:

- La representación anterior da lugar a un proceso estacionario (equivalentemente, a un AR(p)) $\Leftrightarrow 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p \neq 0$
 $\forall z$ con $|z| = 1$.
- Causal $\Leftrightarrow 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p \neq 0 \forall z$ con $|z| \leq 1$.
- Siempre es invertible.
- La fap se anula para todo retardo mayor que p , pero no en el retardo p .

Proceso de medias móviles de orden 1

Un proceso $\{X_t\}_t$ que admite la representación

$$X_t = c + a_t + \theta_1 a_{t-1},$$

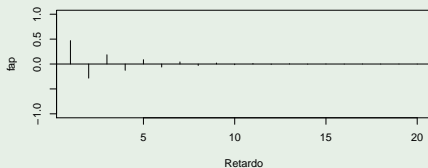
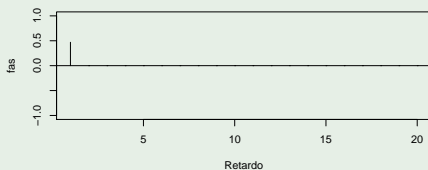
donde c y θ_1 ($\theta_1 \neq 0$) son constantes, se conoce como un “proceso de medias móviles de orden 1”, y se denota por **MA(1)**.

Se verifica que:

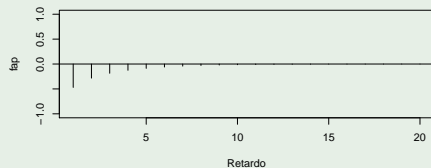
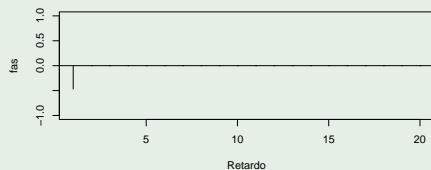
- El proceso MA(1) explica el valor actual (X_t) como una función lineal de 1 valor pasado de un proceso de ruido blanco (a_{t-1}).
- El proceso MA(1) siempre es estacionario y causal.

Ejemplos de la fas y la fap de procesos MA(1):

MA(1): $\theta_1 > 0$



MA(1): $\theta_1 < 0$



Proceso de medias móviles de orden 2

Un proceso $\{X_t\}_t$ que admite la representación

$$X_t = c + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2},$$

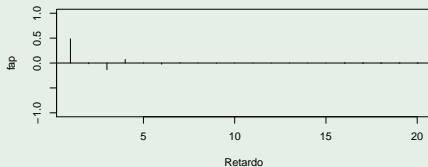
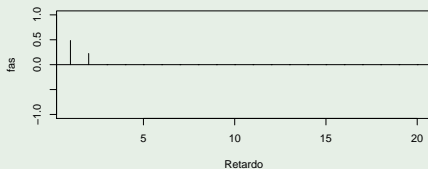
donde c, θ_1 y θ_2 ($\theta_2 \neq 0$) son constantes, se conoce como un “proceso de medias móviles de orden 2”, y se denota por **MA(2)**.

Se verifica que:

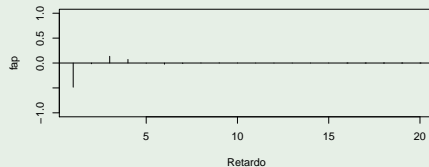
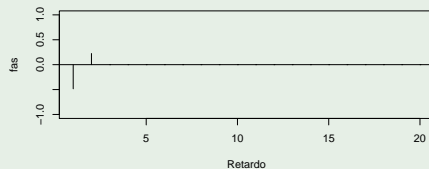
- El proceso MA(2) explica el valor actual (X_t) como una función lineal de 2 valores pasados de un proceso de ruido blanco (a_{t-1} y a_{t-2}).
- El proceso MA(2) siempre es estacionario y causal.

Ejemplos de la fas y la fap de procesos MA(2):

MA(2): $\theta_1 > 0, \theta_2 > 0$

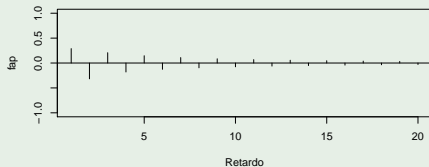
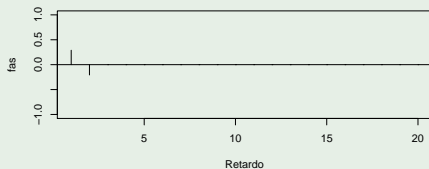


MA(2): $\theta_1 < 0, \theta_2 > 0$

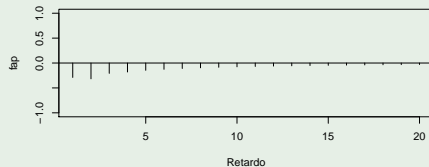
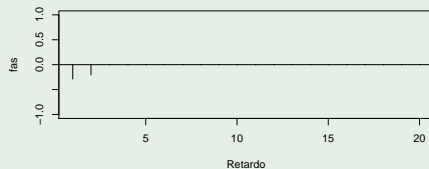


Ejemplos de la fas y la fap de procesos MA(2):

MA(2): $\theta_1 > 0, \theta_2 < 0$



MA(2): $\theta_1 < 0, \theta_2 < 0$



Proceso de medias móviles de orden q

Un proceso $\{X_t\}_t$ que admite la representación

$$X_t = c + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-q},$$

donde $c, \theta_1, \dots, \theta_q$ ($\theta_q \neq 0$) son constantes, se conoce como un “proceso de medias móviles de orden q ” (MA(q)).

Se verifica que:

- Siempre es estacionario y causal.
- Invertible $\Leftrightarrow 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q \neq 0 \ \forall z$ con $|z| \leq 1$.
- La fas se anula para todo retardo mayor que q , pero no en el retardo q .

La introducción en un mismo proceso de estructura autorregresiva (AR) y de medias móviles (MA) da lugar a los procesos ARMA. Por ejemplo:

Proceso ARMA(1,1)

Un proceso **estacionario** $\{X_t\}_t$ que admite la representación

$$X_t = c + \phi_1 X_{t-1} + a_t + \theta_1 a_{t-1},$$

donde c, ϕ_1 y θ_1 ($\phi_1 \neq 0 \neq \theta_1$) son constantes, se conoce como un proceso **ARMA(1,1)**.

Proceso ARMA(2,1)

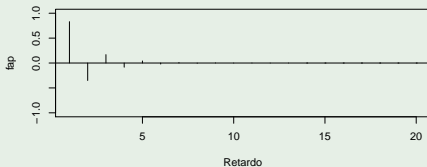
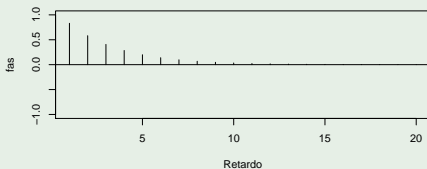
Un proceso **estacionario** $\{X_t\}_t$ que admite la representación

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + a_t + \theta_1 a_{t-1},$$

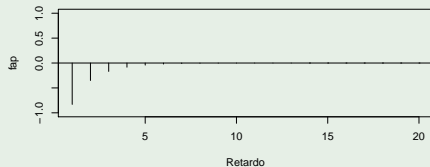
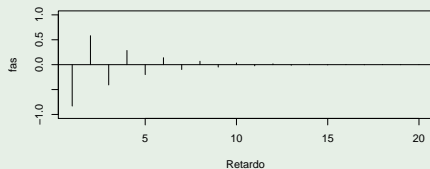
donde c, ϕ_1, ϕ_2 y θ_1 ($\phi_2 \neq 0 \neq \theta_1$) son constantes, se conoce como un proceso **ARMA(2,1)**.

Ejemplos de la fas y la fap de procesos ARMA(1,1):

ARMA(1,1): $\phi_1 > 0, \theta_1 > 0$

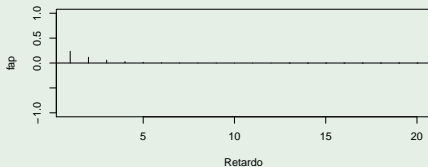
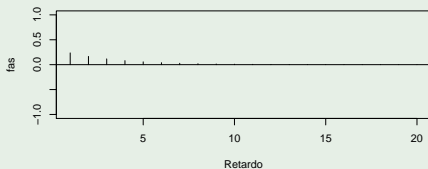


ARMA(1,1): $\phi_1 < 0, \theta_1 < 0$

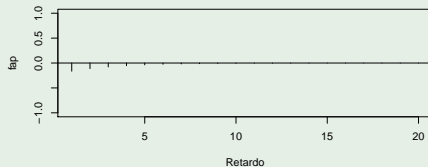
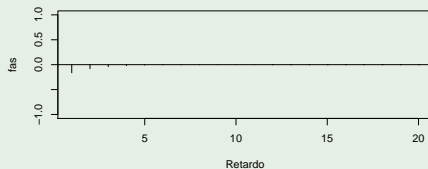


Ejemplos de la fas y la fap de procesos ARMA(1,1):

ARMA(1,1): $\phi_1 > 0, \theta_1 < 0$

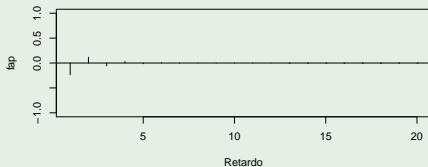
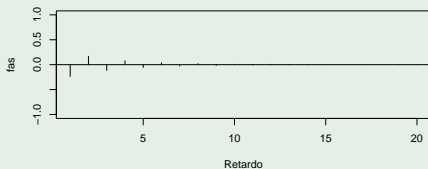


ARMA(1,1): $\phi_1 > 0, \theta_1 < 0$

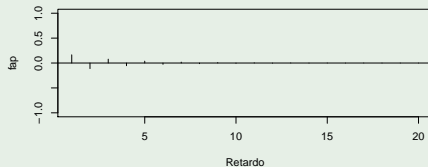
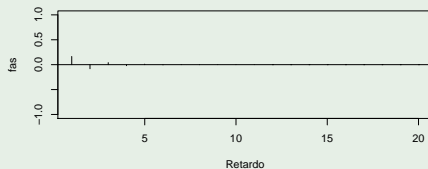


Ejemplos de la fas y la fap de procesos ARMA(1,1):

ARMA(1,1): $\phi_1 < 0, \theta_1 > 0$



ARMA(1,1): $\phi_1 < 0, \theta_1 > 0$



Procesos ARMA(p,q)

Un proceso **estacionario** $\{X_t\}_t$ que admite la representación

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} \\ + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-q},$$

donde $c, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ ($\phi_p \neq 0 \neq \theta_q$) son constantes, se conoce como un proceso **ARMA(p,q)**.

Se verifica que:

- Para que la representación anterior dé lugar a un proceso estacionario (equivalentemente, a un $\text{ARMA}(p, q)$) es necesario y suficiente que

$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p \neq 0 \quad \forall z \text{ con } |z| = 1.$$

- Causal $\Leftrightarrow 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p \neq 0 \quad \forall z \text{ con } |z| \leq 1.$
- Invertible $\Leftrightarrow 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q \neq 0 \quad \forall z \text{ con } |z| \leq 1.$
- $c = \mu(1 - \phi_1 - \cdots - \phi_p).$
- $\text{ARMA}(p, 0) \Leftrightarrow \text{AR}(p); \text{ARMA}(0, q) \Leftrightarrow \text{MA}(q).$

La ecuación que define al proceso ARMA(p,q)

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} \\ + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-q},$$

se puede escribir en la forma compacta

$$\phi(B) X_t = c + \theta(B) a_t,$$

donde

$$\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p), \\ \theta(B) = (1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q)$$

y B denota al **operador retardo**, definido por $BX_t = X_{t-1}$.

La siguiente tabla resume la estructura general de las correlaciones simples y parciales de los modelos anteriormente presentados:

	f_{AS}^*	f_{AP}^*
AR(p)	Muchos coeficientes no nulos*	El último retardo en el que no se anula es p
MA(q)	El último retardo en el que no se anula es q	Muchos coeficientes no nulos*
ARMA(p,q)	Muchos coeficientes no nulos*	Muchos coeficientes no nulos*

*A partir de los primeros retardos, convergen muy rápidamente a cero, como suma de funciones exponenciales y/o sinusoidales.

La clase de procesos ARMA(p,q) que acabamos de presentar:

- Es una familia de procesos estacionarios.
- Es muy flexible: si $\{Y_t\}$ es un proceso estacionario tal que $\gamma_{Y,h} \rightarrow 0$ cuando $h \rightarrow \infty$, entonces, dado cualquier número entero $k > 0$, existe un proceso ARMA $\{X_t\}$ tal que $\gamma_{X,h} = \gamma_{Y,h} \forall h = 0, 1, \dots, k$. Por ello, puede modelizar gran variedad de series generadas por procesos estacionarios.

Dada una “serie de tiempo estacionaria”, trataremos de modelizarla a través de un proceso ARMA(p,q).

Las primeras cuestiones a resolver son:

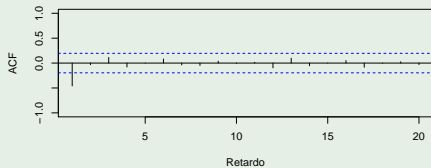
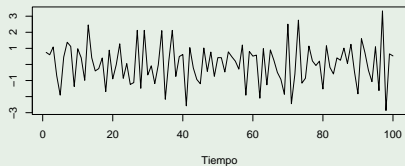
- 1 ¿Cómo [asesorarnos](#) acerca de si una [serie de tiempo](#) ha sido generada por un [proceso ARMA](#)?
- 2 Si efectivamente ha sido generada por un proceso ARMA, ¿cómo [identificar](#) los valores de [p](#) y [q](#)?

La serie de tiempo... ¿ha sido generada por un proceso ARMA?

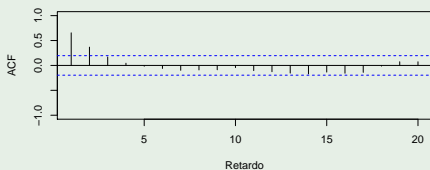
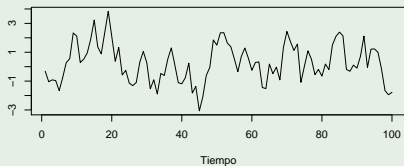
- El gráfico secuencial de la serie de tiempo debe mostrar:
 - Nivel constante.
 - Variabilidad constante.
- La fas muestral $\hat{\rho}_k$ debe converger a cero muy rápidamente a medida que el retardo k crece.

```
plot(...); acf(...)
```

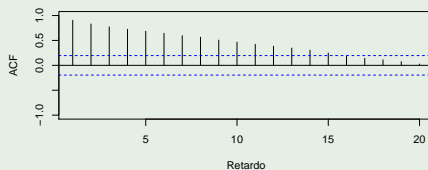
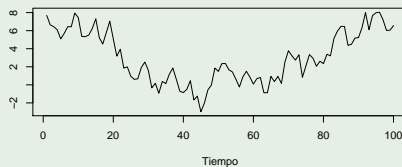
ARMA



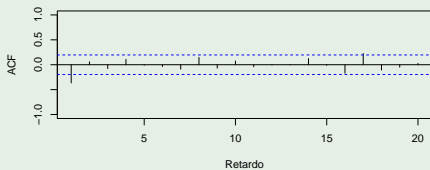
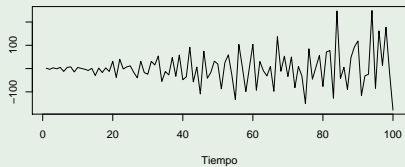
ARMA



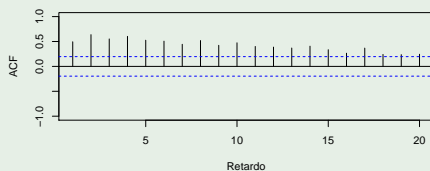
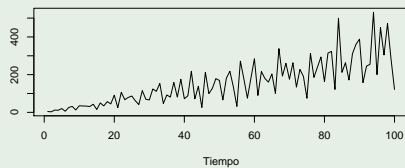
Nivel no constante (tendencia)



Varianza no constante



Nivel y varianza no constantes



Dada una serie de tiempo generada por un proceso ARMA...
¿cuáles son los órdenes p y q correspondientes? (Identificación)

Para responder a esta pregunta nos basaremos, en un principio, en la información que nos suministran la $\hat{\rho}_k$ y la $\hat{\alpha}_k$ (respectivamente).

Tanto $\hat{\rho}_k$ como $\hat{\alpha}_k$ dependen de la serie de tiempo observada, y sus valores cambiarán con ella (son aleatorios).

Por ello, las estimaciones de ρ_k y α_k obtenidas a través de nuestra (única) muestra no aportan suficiente información. Necesitamos utilizar la **distribución muestral** de $\hat{\rho}_k$ y $\hat{\alpha}_k$.

Distribución muestral de $\hat{\rho}_k$ y de $\hat{\alpha}_k$. Inferencia

- Si las variables del proceso $\{X_t\}$ tienen varianza finita y el tamaño (T) de la serie es “grande”, entonces se verifica que:

$$X_t \text{ i.i.d.} \Rightarrow \begin{cases} \hat{\rho}_k \approx N\left(0, \frac{1}{\sqrt{T}}\right), & \forall k \\ \hat{\alpha}_k \approx N\left(0, \frac{1}{\sqrt{T}}\right), & \forall k \end{cases}$$

Conclusión: Si la serie ha sido generada por dicho proceso (i.i.d) debería cumplirse que, para cada $k = 1, 2, \dots$ (con un nivel de significación individual aproximado del 5 %):

$$\hat{\rho}_k \in \left(-\frac{1.96}{\sqrt{T}}, \frac{1.96}{\sqrt{T}}\right) \text{ y } \hat{\alpha}_k \in \left(-\frac{1.96}{\sqrt{T}}, \frac{1.96}{\sqrt{T}}\right).$$

Nota: Estas regiones de no rechazo serán utilizadas para contrastar la **incorrelación** de variables cuya media y varianza son constantes (equivalentemente, estructura ARMA(0,0); esto es, “ruido blanco + c”).

- Bajo condiciones generales (incluyendo T “grande”) se verifica que:

$$\text{AR}(p) \Rightarrow \hat{\alpha}_k \approx N\left(0, \frac{1}{\sqrt{T}}\right), \forall k > p.$$

Conclusión: Si la serie ha sido generada por un proceso $\text{AR}(p)$ debería cumplirse que, para cada $k > p$ (con un nivel de significación individual aproximado del 5 %):

$$\hat{\alpha}_k \in \left(-\frac{1.96}{\sqrt{T}}, \frac{1.96}{\sqrt{T}}\right).$$

Nota: Estas regiones de no rechazo serán utilizadas para contrastar estructura $\text{AR}(p)$.

- Bajo condiciones generales (incluyendo T “grande”) se verifica que:

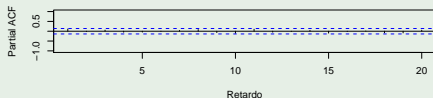
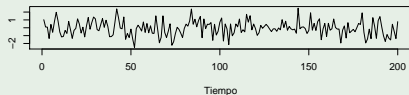
$$\text{MA}(q) \Rightarrow \hat{\rho}_k \approx N \left(0, \frac{\sqrt{1 + 2(\rho_1^2 + \dots + \rho_q^2)}}{\sqrt{T}} \right), \forall k > q.$$

Conclusión: Si la serie ha sido generada por un proceso MA(q) debería cumplirse que, para cada $k > q$ (con un nivel de significación individual aproximado del 5 %):

$$\hat{\rho}_k \in \left(\pm 1.96 \sqrt{\frac{1 + 2(\hat{\rho}_1^2 + \dots + \hat{\rho}_q^2)}{T}} \right)$$

Nota: Estas regiones de no rechazo (en general, considerando en ellas $\hat{\rho}_j = 0, j = 1, \dots, q$) serán utilizadas para contrastar estructura MA(q).

Serie, fas y fap



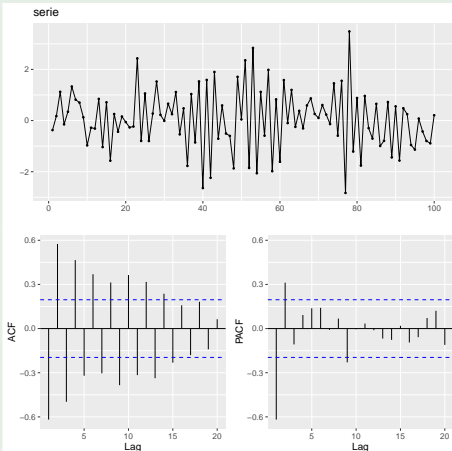
Conclusión

Los gráficos de la izquierda **sugieren** que la serie:

- 1 Es **estacionaria**.
- 2 Ha sido generada por un proceso **ARMA(0,0)** (equivalentemente, por un proceso de “ruido blanco + c ”).

```
plot(...); acf(...); pacf(...)
```

Serie, fas y fap



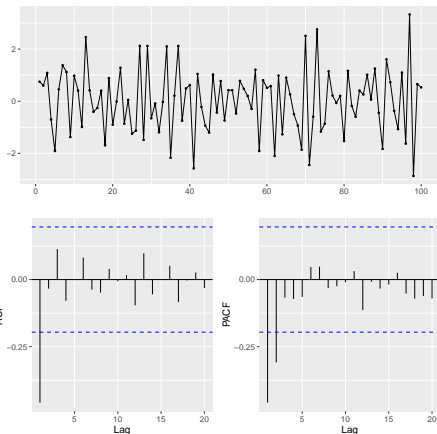
Conclusión

Los gráficos de la izquierda **sugieren** que la serie:

- 1 Es **estacionaria**.
- 2 Ha sido generada por un proceso **AR(2)**.

`ggtsdisplay(...)`

Serie, fas y fap



Conclusión

Los gráficos de la izquierda **sugieren** que la serie:

- 1 Es **estacionaria**.
- 2 Ha sido generada por un proceso **MA(1)** o por un **AR(2)**.

Las tres técnicas que acabamos de exponer permiten la identificación de procesos ARMA(0,0), AR(p) o MA(q), pero no la identificación de ARMAAs mixtos (esto es, de modelos ARMA(p,q) siendo $p \neq 0 \neq q$). En realidad, la identificación de ARMAAs mixtos a partir del estudio de la fas ($\hat{\rho}_k$) y la fap ($\hat{\alpha}_k$) muestrales es una tarea muy complicada.

Si bien existen distintas propuestas para identificar ARMAAs mixtos (el método de la esquina (Becuín y otros, 1980); el método de la función de autocorrelación simple extendida (Tsay y Tiao, 1984); el método de la menor correlación canónica (Tsay y Tiao, 1985); etc), no son demasiado populares.

Como ya hemos justificado anteriormente, la clase de procesos ARMA que acabamos de estudiar es una familia muy flexible de procesos estacionarios.

Sin embargo, **no abundan** series reales generadas por procesos estacionarios:

- Las series reales suelen presentar tendencia y/o patrones repetitivos.

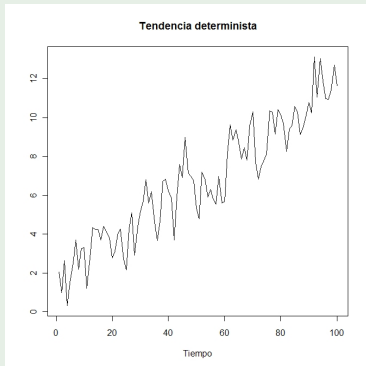
Por tanto:

- Necesitamos ampliar la clase de procesos ARMA, de modo que la nueva clase permita incorporar dichas características.

Motivación (ejemplo I):

Sea $\{V_t\}_t$ un proceso estacionario.

El proceso $X_t = \beta_0 + \beta_1 t + V_t$, donde β_0 y $\beta_1 \neq 0$ son parámetros, no es estacionario (tiene tendencia determinista).



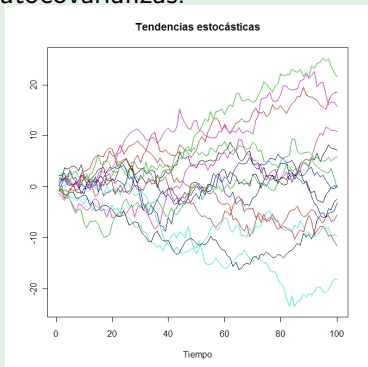
El proceso **diferenciado regularmente**

$$X_t - X_{t-1} = \beta_1 + (V_t - V_{t-1})$$

es estacionario.

Motivación (ejemplo II):

El paseo aleatorio $X_t = X_{t-1} + a_t$ no es estacionario. Por ejemplo, si comienza en $t = 0$ y suponemos que $X_0 = 0$, aunque su media es estable (no tiene tendencia determinista), no lo son ni sus varianzas ni sus autocovarianzas.



El paseo aleatorio diferenciado regularmente

$$X_t - X_{t-1} = a_t$$

es estacionario.

Los ejemplos anteriores muestran situaciones en las que la aplicación de 1 diferencia regular consigue eliminar la tendencia y transformar un proceso no estacionario en otro estacionario.

En base a esto, ante una serie con tendencia, sugerimos:

- **Eliminar la tendencia** de la serie aplicando sucesivamente d diferencias regulares (en general, $d \leq 3$). Esto es, si después de diferenciar regularmente la serie persiste la existencia de tendencia, diferenciaremos la serie diferenciada, y así sucesivamente hasta obtener una serie sin tendencia.
- Si la serie obtenida es estacionaria, modelizarla a través de un proceso ARMA (recuérdese la gran capacidad que tiene la clase ARMA para modelizar procesos estacionarios).

Procesos ARIMA(p,d,q)

Un proceso ARIMA(p,d,q) es aquél que, después de aplicarle d diferencias regulares, se convierte en un proceso ARMA(p,q). Es decir:

$$\{X_t\}_t \text{ es ARIMA}(p,d,q) \Leftrightarrow (1 - B)^d X_t \text{ es ARMA}(p,q).$$

Equivalentemente:

$\{X_t\}_t$ es un proceso ARIMA(p,d,q) si admite una representación del tipo:

$$\phi(B)(1 - B)^d X_t = c + \theta(B) a_t,$$

donde el polinomio $\phi(z)$ no tiene raíces de módulo 1.

La expresión del ARIMA(1,1,1) es

$$\begin{array}{ccc}
 \text{AR} & & \text{MA} \\
 \downarrow & & \downarrow \\
 (1 - \phi_1 B)(1 - B)X_t = c + (1 + \theta_1 B)a_t. \\
 \uparrow \\
 \text{Dif.}
 \end{array}$$

Operando en dicha expresión, se obtiene la representación:

$$X_t = c + (1 + \phi_1)X_{t-1} - \phi_1 X_{t-2} + a_t + \theta_1 a_{t-1},$$

que muestra de una manera explícita la relación existente entre las variables del proceso en el instante actual t y en los instantes pasados $t - 1$ y $t - 2$.

En la práctica, ante una **serie real**...

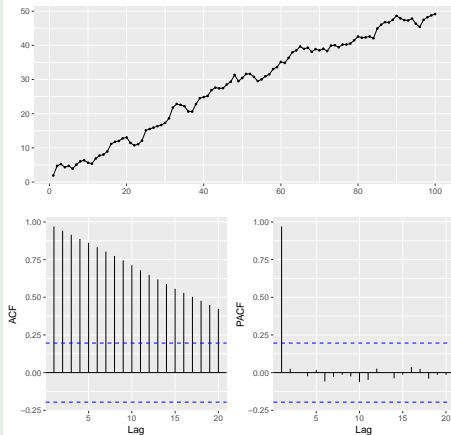
¿cuándo propondremos un **ARIMA(p,d,q)** como su posible generador?

Cuando muestre no estacionariedad provocada exclusivamente por la presencia de tendencia.

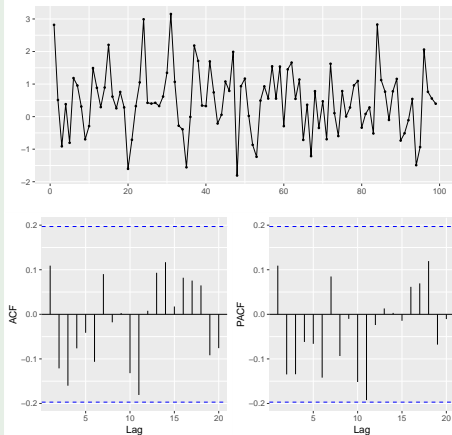
La presencia de tendencia (necesidad de diferenciar la serie para convertirla en estacionaria) suele ser delatada por:

- El gráfico secuencial de la serie.
- La fas muestral:
 - Toma **valores positivos** (a menudo próximos a 1 en los primeros retardos, aunque no necesariamente).
 - Decae lentamente a cero (**decrecimiento lineal**) a medida que el retardo crece.

Serie original (tendencia)

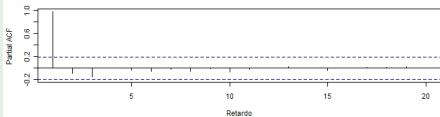
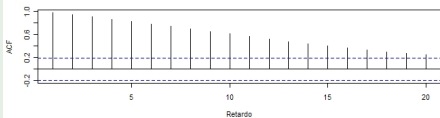
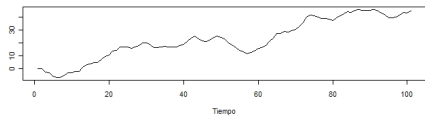


Serie diferenciada (estacionaria)

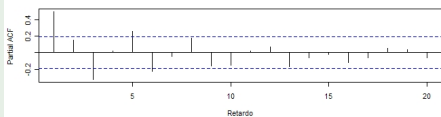
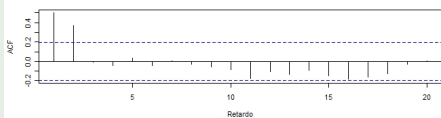
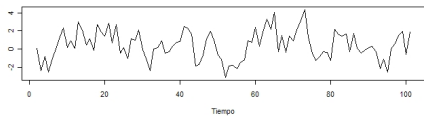


Sugerencia: Serie original no estacionaria, y generada por un proceso $ARIMA(0,1,0)$.

Serie original (tendencia)



Serie diferenciada (estacionaria)



Sugerencia: Serie original no estacionaria, y generada por un proceso ARIMA(0,1,2).

Caso particular: Estimación en procesos ARMA(p,q)

Supongamos que la serie de tiempo x_1, \dots, x_T ha sido generada por un proceso ARMA(p,q):

- Causal, invertible y gaussiano.
- Cuyos órdenes p y q son conocidos.

Representemos dicho proceso a través de

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} \\ + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q}.$$

Los parámetros $c, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ y σ_a^2 son desconocidos y deben ser estimados a partir de la información disponible (la serie de tiempo).

Los dos métodos de estimación que presentaremos en primer lugar requieren de la noción de residuo.

Los **residuos** asociados a las estimaciones

$$\tilde{c}, \tilde{\phi}_1, \dots, \tilde{\phi}_p, \tilde{\theta}_1, \dots, \tilde{\theta}_q$$

se definen, para $t = 1, \dots, T$, como la diferencia entre los valores observados y los correspondientes valores ajustados; esto es,

$$\hat{a}_t = x_t - (\tilde{c} + \tilde{\phi}_1 x_{t-1} + \dots + \tilde{\phi}_p x_{t-p} + \tilde{\theta}_1 \hat{a}_{t-1} + \dots + \tilde{\theta}_q \hat{a}_{t-q}),$$

siendo la suma de sus cuadrados

$$S(\tilde{c}, \tilde{\phi}_1, \dots, \tilde{\phi}_p, \tilde{\theta}_1, \dots, \tilde{\theta}_q) = \sum_{t=1}^T \hat{a}_t^2.$$

Estimación por mínimos cuadrados

La estimación de los parámetros $(c, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)$ por medio del **método de mínimos cuadrados** se obtiene a través de los valores $(\hat{c}, \hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q)$ que minimizan a la función S ; esto es,

$$\begin{aligned} (\hat{c}, \hat{\phi}_1, \dots, \hat{\theta}_q) &= \arg \min_{\tilde{c}, \tilde{\phi}_1, \dots, \tilde{\theta}_q} S(\tilde{c}, \tilde{\phi}_1, \dots, \tilde{\theta}_q) \\ &= \arg \min_{\tilde{c}, \tilde{\phi}_1, \dots, \tilde{\theta}_q} \sum_{t=1}^T \hat{a}_t^2, \end{aligned}$$

siendo

$$\hat{a}_t = x_t - (\tilde{c} + \tilde{\phi}_1 x_{t-1} + \dots + \tilde{\phi}_p x_{t-p} + \tilde{\theta}_1 \hat{a}_{t-1} + \dots + \tilde{\theta}_q \hat{a}_{t-q}).$$

Estimación por mínimos cuadrados condicionados

Si $p > 0$, una dificultad asociada a la estimación por mínimos cuadrados está en la obtención de

$$\hat{a}_1, \hat{a}_2, \dots, \hat{a}_p,$$

pues dependen de los valores no observados de

$$X_0, X_{-1}, \dots, X_{1-p}.$$

Esta dificultad puede ser solventada minimizando la función

$$S_C \left(\tilde{c}, \tilde{\phi}_1, \dots, \tilde{\phi}_p, \tilde{\theta}_1, \dots, \tilde{\theta}_q \right) = \sum_{t=p+1}^T \hat{a}_t^2$$

en lugar de la función S .

A su vez, \hat{a}_{p+1} depende de los valores de

$$\hat{a}_p, \hat{a}_{p-1}, \dots, \hat{a}_{p+1-q},$$

los cuales dependen a su vez de valores no observados de X_t .

Es importante notar que si, dada la serie x_1, \dots, x_T , fijamos los valores de

$$\hat{a}_p, \hat{a}_{p-1}, \dots, \hat{a}_{p+1-q},$$

podemos construir iterativamente los valores de

$$\hat{a}_{p+1}, \hat{a}_{p+2}, \dots, \hat{a}_T.$$

La estimación de los parámetros $(c, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)$ por medio del **método de mínimos cuadrados condicionados** se obtiene a través de los valores $(\hat{c}, \hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q)$ que minimizan a la función

$$S_C \left(\tilde{c}, \tilde{\phi}_1, \dots, \tilde{\phi}_p, \tilde{\theta}_1, \dots, \tilde{\theta}_q \right) = \sum_{t=p+1}^T \hat{a}_t^2$$

condicionada a que

$$\hat{a}_p = \hat{a}_{p-1} = \dots = \hat{a}_{p+1-q} = 0.$$

Estimación de máxima verosimilitud

La estimación de máxima verosimilitud de los parámetros $c, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ y σ_a^2 se obtiene a través de los valores que dan **mayor credibilidad** a la serie observada x_1, \dots, x_T .

La credibilidad que los valores $\tilde{c}, \tilde{\phi}_1, \dots, \tilde{\phi}_p, \tilde{\theta}_1, \dots, \tilde{\theta}_q, \tilde{\sigma}_a^2$ dan a la serie observada x_1, \dots, x_T se mide a través de la **función de verosimilitud**:

$$L_{x_1, \dots, x_T} \left(\tilde{c}, \tilde{\phi}_1, \dots, \tilde{\phi}_p, \tilde{\theta}_1, \dots, \tilde{\theta}_q, \tilde{\sigma}_a^2 \right) = f_{\tilde{c}, \tilde{\phi}_1, \dots, \tilde{\phi}_p, \tilde{\theta}_1, \dots, \tilde{\theta}_q, \tilde{\sigma}_a^2} (x_1, \dots, x_T),$$

donde $f_{\tilde{c}, \tilde{\phi}_1, \dots, \tilde{\phi}_p, \tilde{\theta}_1, \dots, \tilde{\theta}_q, \tilde{\sigma}_a^2}$ denota a la función de densidad conjunta de un vector aleatorio $(\tilde{X}_1, \dots, \tilde{X}_T)'$ procedente de un proceso ARMA con parámetros $\tilde{c}, \tilde{\phi}_1, \dots, \tilde{\phi}_p, \tilde{\theta}_1, \dots, \tilde{\theta}_q, \tilde{\sigma}_a^2$.

Por tanto, la estimación de los parámetros $c, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ y σ_a^2 por medio del **método de máxima verosimilitud** se obtiene a través de los valores $\hat{c}, \hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q$ y $\hat{\sigma}_a^2$ que maximizan a la función de verosimilitud L ; esto es,

$$(\hat{c}, \hat{\phi}_1, \dots, \hat{\theta}_q, \hat{\sigma}_a^2) = \arg \max_{\tilde{c}, \tilde{\phi}_1, \dots, \tilde{\theta}_q, \tilde{\sigma}_a^2} L(\tilde{c}, \tilde{\phi}_1, \dots, \tilde{\theta}_q, \tilde{\sigma}_a^2).$$

Nota: Hemos suprimido los subíndices en la función de verosimilitud; esto es, a partir de ahora utilizaremos simplemente la letra L para denotar a la función de verosimilitud L_{x_1, \dots, x_T} .

Ejemplo: Proceso ARMA(p,q) gaussiano

Su función de verosimilitud es:

$$L\left(\tilde{c}, \tilde{\phi}_1, \dots, \tilde{\phi}_p, \tilde{\theta}_1, \dots, \tilde{\theta}_q, \tilde{\sigma}_a^2\right) = \frac{1}{\sqrt{(2\pi)^T |\tilde{\mathbf{V}}_T|}} \exp\left(-\frac{(\mathbf{x}_T - \tilde{\boldsymbol{\mu}}_T)' \tilde{\mathbf{V}}_T^{-1} (\mathbf{x}_T - \tilde{\boldsymbol{\mu}}_T)}{2}\right),$$

donde $\mathbf{x}_T = (x_1, \dots, x_T)'$, $\tilde{\boldsymbol{\mu}}_T = (\tilde{\mu}, \dots, \tilde{\mu})'$ con

$$\tilde{\mu} = E\left(\tilde{X}_t\right) = \tilde{c} / \left(1 - \tilde{\phi}_1 - \dots - \tilde{\phi}_p\right)$$

y $\tilde{\mathbf{V}}_T$ es la matriz de varianzas-covarianzas del vector aleatorio $(\tilde{X}_1, \dots, \tilde{X}_T)'$ procedente de un proceso ARMA(p,q) con parámetros $\tilde{c}, \tilde{\phi}_1, \dots, \tilde{\phi}_p, \tilde{\theta}_1, \dots, \tilde{\theta}_q$ y $\tilde{\sigma}_a^2$.

Ejemplo: Proceso AR(1) gaussiano ($X_t = c + \phi_1 X_{t-1} + a_t$)

Su función de verosimilitud es:

$$L(\tilde{c}, \tilde{\phi}_1, \tilde{\sigma}_a^2) = (2\pi\tilde{\sigma}_a^2)^{-T/2} (1 - \tilde{\phi}_1^2)^{1/2} \exp\left(-\frac{S(\tilde{c}, \tilde{\phi}_1)}{2\tilde{\sigma}_a^2}\right),$$

donde

$$S(\tilde{c}, \tilde{\phi}_1) = (1 - \tilde{\phi}_1^2) \left(x_1 - \frac{\tilde{c}}{1 - \tilde{\phi}_1}\right)^2 + \sum_{t=2}^T (x_t - \tilde{c} - \tilde{\phi}_1 x_{t-1})^2.$$

Los valores de \tilde{c} , $\tilde{\phi}_1$ y $\tilde{\sigma}_a^2$ que maximizan a la función L dan lugar a los estimadores de máxima verosimilitud de los parámetros del AR(1) gaussiano.

Bajo ciertas condiciones de regularidad generales, se tiene que:

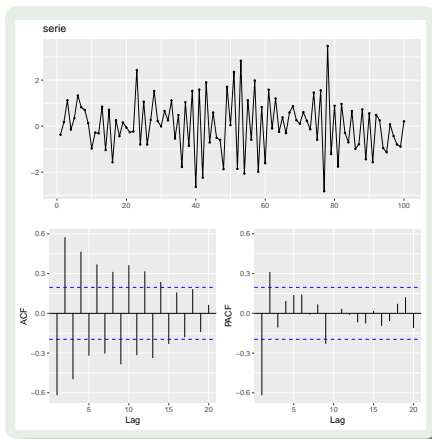
- Los estimadores de máxima verosimilitud de los parámetros (salvo σ_a^2) de un ARMA(p,q) son **asintóticamente óptimos**; esto es, si el tamaño T de la serie es **grande**, se puede considerar que:
 - 1 Son centrados (o insesgados).
 - 2 Son eficientes.
 - 3 Su distribución es normal.
- El estimador de máxima verosimilitud de σ_a^2 es consistente.

Nota: Las propiedades 1 y 3 nos permiten construir intervalos de confianza y contrastes de hipótesis referentes a los parámetros. Así, si denotamos por γ a cualquiera de los anteriores parámetros (salvo σ_a^2), por $\hat{\gamma}$ a su estimación máximo verosímil y por $\hat{\sigma}_{\hat{\gamma}}$ a la estimación de la desviación típica del estimador correspondiente, entonces para resolver el contraste de hipótesis $H_0 : \gamma = 0$ frente a $H_1 : \gamma \neq 0$, tendremos que:

$$\hat{\gamma} \text{ es significativamente distinta de } 0 \text{ al } 5\% \Leftrightarrow |\hat{\gamma}| \geq 1.96 \times \hat{\sigma}_{\hat{\gamma}}.$$

Las propiedades 1 y 3 se mantienen para estimadores basados en la verosimilitud gaussiana, aunque el proceso no sea gaussiano.

En un ejemplo anterior, habíamos hecho un análisis básico de cierta serie (gráfico de la derecha). Dicho análisis sugería que la serie podría haber sido generada por un proceso AR(2).



Si estimamos sus parámetros por máxima verosimilitud resulta:

$$\hat{\phi}_1 = -0.4201 (0.0942), \hat{\phi}_2 = 0.3096 (0.0943),$$

$$\hat{\mu} = 0.0473 (0.0794) \text{ y } \hat{\sigma}_a^2 = 0.804.$$

Observamos que, al 5 %, $\hat{\mu}$ no es significativamente distinta de cero. Pasamos por tanto a estimar un AR(2) con $\mu = 0$.

Nota: Estimar un ARMA(p,q) con $\mu = 0$ es lo mismo que estimar un ARMA(p,q) con $c = 0$, pues se tiene que

$$c = \mu(1 - \phi_1 - \cdots - \phi_p).$$

`Arima(...)`

Bajo la restricción $\mu = 0$, se obtienen las estimaciones:

$$\hat{\phi}_1 = -0.4149 \text{ (0.0939)}, \hat{\phi}_2 = 0.315 \text{ (0.094)} \text{ y } \hat{\sigma}_a^2 = 0.7986,$$

resultando tanto $\hat{\phi}_1$ como $\hat{\phi}_2$ significativamente distintos de cero (al 5 %).

Por tanto, el AR(2) estimado es:

$$X_t = -0.4149X_{t-1} + 0.315X_{t-2} + a_t,$$

siendo 0.7986 la varianza del ruido blanco.

```
Arima(..., include.mean=FALSE)
```

Caso particular: Diagnósis en procesos ARMA(p,q)

Una vez que un modelo ARMA ha sido ajustado, la siguiente etapa es la “comprobación” de que las hipótesis básicas realizadas sobre él se verifican. Esto se conoce como la **diagnósis** o **chequeo** del modelo.

- La hipótesis más importante es la que exige que las innovaciones $\{a_t\}_t$ sean ruido blanco; esto es:
 - Tengan media cero.
 - Tengan varianza constante.
 - Estén incorreladas.

Su no verificación invalida al modelo ajustado como posible generador de la serie de tiempo en estudio.

- La hipótesis de normalidad es conveniente por tres motivos:
 - 1 Bajo normalidad, el ruido blanco equivale a la independencia. Esto garantiza que no estamos dejando información por modelizar.
 - 2 Bajo normalidad, los estimadores que utilizamos (máxima verosimilitud gaussiana) son asintóticamente eficientes.
 - 3 Próximamente realizaremos predicciones de valores futuros del proceso, resultando conveniente que vayan acompañadas de intervalos de predicción. Si no tenemos normalidad, no podremos “garantizar” el nivel de confianza de los intervalos que construiremos en base a la distribución asintótica del error de predicción.

A continuación, presentamos algunos [gráficos](#) que nos pueden asesorar acerca de si una muestra y_1, \dots, y_T (por ejemplo, una muestra de las innovaciones a_1, \dots, a_T) es o no una realización de un conjunto de variables aleatorias procedentes de un proceso de ruido blanco gaussiano.

El gráfico secuencial

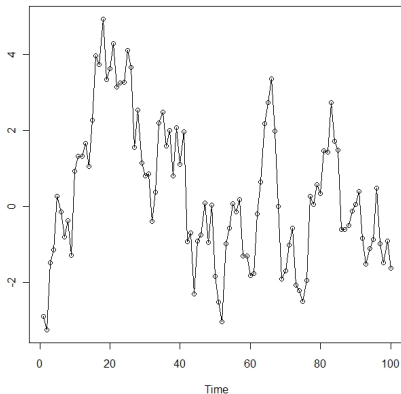
La representación gráfica de la muestra frente al tiempo (esto es, el gráfico secuencial) puede ayudarnos a detectar de manera visual y rápida la presencia de:

- Tendencia.
- Componente estacional.
- Variabilidad no constante.
- Dependencia lineal
 - Positiva: tendencias que desaparecen a corto plazo (tendencias locales).
 - Negativa: valores altos son seguidos con frecuencia por valores bajos, y viceversa (zig-zag).

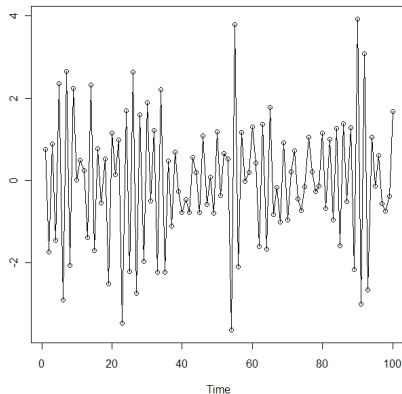
Cualquiera de estas situaciones invalidaría la hipótesis de ruido blanco.

Ejemplo: Dependencia positiva vs Dependencia negativa

Dependencia positiva



Dependencia negativa



El gráfico Q-Q normal

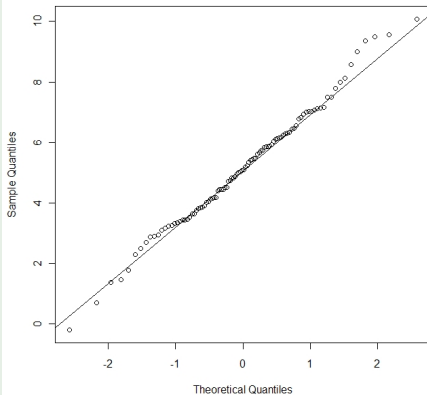
El gráfico Q-Q (Cuantil-Cuantil) normal representa a los cuantiles muestrales frente a los cuantiles de una distribución $N(0, 1)$.

Si la muestra y_1, \dots, y_T es i.i.d. con distribución normal, su gráfico Q-Q normal debería ser aproximadamente lineal.

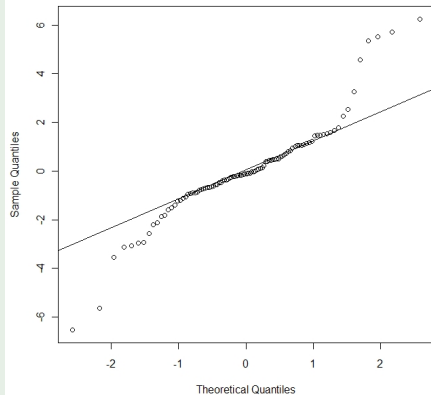
Por tanto, la no linealidad del gráfico Q-Q normal sugiere ausencia de normalidad.

Ejemplo: Normalidad vs Falta de normalidad

Normalidad



Falta de normalidad



`qqnorm(...); qqline(...)`

A continuación, presentamos varios **contrastes de hipótesis** diseñados para contrastar si una muestra y_1, \dots, y_T es una realización de un conjunto de variables aleatorias:

- Independientes.
- Con media cero.
- Con distribución común gaussiana.

Contraste de independencia

Denotemos por $\hat{\rho}_k$ ($k = 1, 2, \dots$) a la fas de la muestra en estudio, y_1, \dots, y_T .

Bajo la hipótesis nula de que la muestra proviene de variables aleatorias i.i.d. con varianza finita, y asumiendo que el tamaño muestral T es “grande”, se tiene que:

$$\hat{\rho}_k \approx N\left(0, \frac{1}{\sqrt{T}}\right).$$

Por tanto, rechazaremos la independencia (al 5 %) si

$$|\hat{\rho}_k| \geq \frac{1.96}{\sqrt{T}}.$$

`acf(...)`, o también `ggAcf(...)`

Otro contraste de independencia: Ljung-Box

Utilicemos ahora además la notación

$$Q_H = T(T+2) \sum_{k=1}^H \frac{\hat{\rho}_k^2}{T-k}.$$

Bajo la hipótesis nula de que la muestra proviene de variables aleatorias i.i.d. con varianza finita, y asumiendo que el tamaño muestral T es “grande”, se tiene que:

$$Q_H \approx \chi_H^2.$$

Por tanto, rechazaremos la independencia (al 5%) si el valor de Q_H es mayor o igual que el percentil 0.95 de la distribución χ_H^2 .

Este contraste se conoce como contraste de [Ljung-Box](#).

```
Box.test(..., type="Ljung-Box"), o también tsdiag(...)
```

Contraste de media cero

Utilicemos ahora los símbolos \bar{y} y \hat{s}_y^2 para denotar a la media y a la cuasivarianza muestrales, respectivamente.

Bajo la hipótesis nula de que la muestra y_1, \dots, y_T proviene de variables aleatorias i.i.d. con media cero y varianza finita, y asumiendo que el tamaño muestral T es “grande”, se tiene que:

$$\frac{\bar{y}}{\hat{s}_y/\sqrt{T}} \approx t_{T-1} \approx N(0, 1).$$

Por tanto, rechazaremos que la media μ_y es cero (al 5 %) si

$$|\bar{y}| \geq 1.96 \frac{\hat{s}_y}{\sqrt{T}}.$$

```
t.test(..., mu=0)
```

Contraste de normalidad: Jarque-Bera

Utilicemos ahora la notación

$$G_1 = \frac{\sum_{t=1}^T (y_t - \bar{y})^3}{T s_y^3} \text{ y } G_2 = \frac{\sum_{t=1}^T (y_t - \bar{y})^4}{T s_y^4} - 3,$$

donde s_y^2 denota a la varianza muestral.

Bajo la hipótesis nula de que la muestra y_1, \dots, y_T proviene de variables aleatorias i.i.d. con distribución gaussiana, y asumiendo que el tamaño muestral T es “grande”, se tiene que:

$$T \left(\frac{G_1^2}{6} + \frac{G_2^2}{24} \right) \approx \chi_2^2.$$

Rechazaremos la normalidad (al 5 %) si el valor del estadístico es mayor o igual que el percentil 0.95 de la distribución χ_2^2 . Este contraste se conoce como contraste de [Jarque-Bera](#).

```
jarque.bera.test(...)
```

Otro contraste de normalidad: Shapiro-Wilk

Sea $\omega = \frac{\left(\sum_{t=1}^{\lfloor T/2 \rfloor} b_{t,T} (y_{(T-t+1)} - y_{(t)})\right)^2}{Ts_y^2}$, donde $y_{(t)}$ denota al estadístico ordenado de orden t y las constantes $b_{t,T}$ vienen dadas a partir de la inversa de la distribución normal estándar.

El estadístico ω puede interpretarse como el cuadrado del coeficiente de correlación lineal de los puntos muestrales dibujados sobre papel probabilístico normal. Puesto que bajo la hipótesis nula de que y_1, \dots, y_T son i.i.d. con distribución gaussiana dicho gráfico debería ser aproximadamente lineal, se rechaza la normalidad para valores pequeños de ω . [Shapiro y Wilk](#) tabularon los valores de $b_{t,T}$, y la distribución de ω bajo la hipótesis nula.

```
shapiro.test(...)
```

Los gráficos y contrastes de hipótesis que acabamos de presentar nos ayudarán en la verificación de si el modelo ARMA propuesto es o no adecuado como generador de nuestra serie de tiempo (etapa de chequeo o diagnosis).

Concretamente, nos asesorarán en la toma de la decisión referente a si las innovaciones a_t del modelo ARMA son o no **ruido blanco** con distribución **gaussiana**.

Puesto que las innovaciones a_t no son observables, lo que se hace es “estimarlas” y realizar el chequeo sobre dichas estimaciones (esto es, sobre los residuos \hat{a}_t del modelo estimado o ajustado).

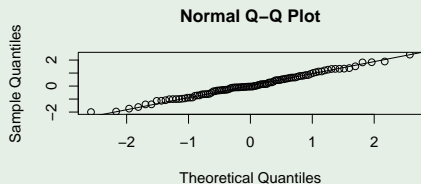
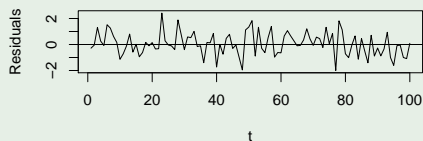
Los contrastes de independencia aplicados a los residuos \hat{a}_t sufren las siguientes **modificaciones** (con respecto a su aplicación a las innovaciones a_t):

- Contraste basado en la distribución de cada autocorrelación muestral $\hat{\rho}_k$: La varianza asintótica de $\hat{\rho}_k$, para retardos k “pequeños”, deja de ser $1/T$ (es menor que $1/T$).
- Contraste de **Ljung-Box**: Los grados de libertad de la distribución asintótica de Q_H pasan a ser $H - p - q - 1$ o $H - p - q$, dependiendo de que el ARMA tenga o no constante, respectivamente (necesitamos por tanto que $H > p + q + 1$ o $H > p + q$, respectivamente).

En ambos casos, la región de rechazo resulta modificada.

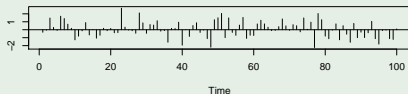
A continuación, pasamos a realizar la diagnosis del modelo AR(2) que habíamos estimado para cierta serie. Para ello, utilizaremos los gráficos y contrastes que acabamos de proponer.

Gráficos de residuos y Q-Q normal

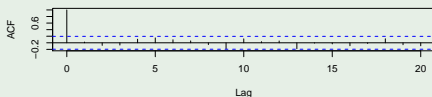


Contrastes de independencia

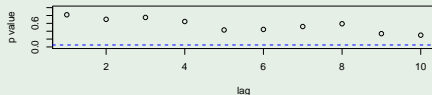
Standardized Residuals



ACF of Residuals



p values for Ljung-Box statistic



`tsdiag(...)`

Contrastes de media cero y normalidad

- $\mu_a = 0$: $p - \text{valor} = 0.5637$
- Normalidad:
 - Jarque-Bera: $p - \text{valor} = 0.8715$
 - Shapiro-Wilk: $p - \text{valor} = 0.9016$

Conclusión: Un AR(2) gaussiano es un modelo apropiado como generador de la serie analizada.

A continuación se muestra un resumen del análisis de residuos recién presentado obtenido a través de la ejecución de una única orden.

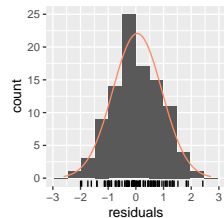
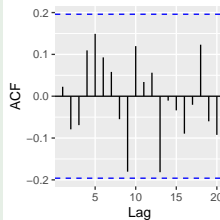
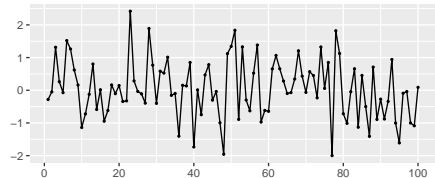
Ljung-Box test

data: Residuals from ARIMA(2,0,0)
with zero mean

$Q^* = 11.781$, $df = 8$, $p\text{-value} = 0.1613$

Model df: 2. Total lags used: 10

Residuals from ARIMA(2,0,0) with zero mean



`checkresiduals(...)`

Caso general: Estimación y diagnóstico en procesos ARIMA(p,d,q)

Como consecuencia de la estrecha relación existente entre los procesos ARIMA(p,d,q) y los procesos ARMA(p,q):

$$\{X_t\}_t \text{ es ARIMA}(p,d,q) \Leftrightarrow (1 - B)^d X_t \text{ es ARMA}(p,q),$$

se tiene que para estimar y realizar la diagnóstico de un modelo ARIMA(p,d,q) es suficiente:

- Identificar el orden de diferenciación regular, d.
- Saber estimar y realizar la diagnóstico de un modelo ARMA(p,q).

Caso particular: Selección del modelo en procesos ARMA(p,q)

Las etapas que hemos seguido para proponer un modelo ARMA como posible generador de una serie de tiempo han sido:

- 1 Asesorarnos (gráficamente) acerca de la “estacionariedad de la serie”. Si es estacionaria, pasar a la etapa 2.
- 2 Identificar los órdenes p y q del ARMA: estudio de sus fas y fap muestrales.
- 3 Estimar el modelo cuyos órdenes se identificaron en la etapa 2: mín. cuadr. condicionados, máx. verosimilitud.
- 4 Chequear el modelo estimado: análisis de residuos.

Si los residuos pueden ser considerados como procedentes de un proceso de ruido blanco (preferiblemente gaussiano), el modelo estimado es propuesto como posible generador de la serie de tiempo.

Como acabamos de recordar, hemos basado la etapa correspondiente a la identificación de los órdenes p y q del ARMA en el estudio de las fas y fap muestrales.

Sin embargo:

- Es posible que a partir de la fas y fap muestrales no identifiquemos ningún proceso como posible generador de la serie de tiempo: convendría disponer de algún método que sugiera modelos de forma automática.
- Las fas y fap muestrales pueden llevarnos a identificar varios procesos como posibles generadores de la serie de tiempo: parece entonces necesario establecer criterios que nos permitan determinar, entre varios modelos, cuál es preferible.
- Es posible que haya algún proceso distinto de los identificados a través de la fas y fap muestrales que sea preferible a ellos: convendría disponer de algún método que sugiera modelos de forma automática.

Sea k la cantidad de parámetros, excluyendo a la varianza de las innovaciones (σ_a^2), de un modelo ARMA(p,q) (esto es, $k = p + q + 1$ o $k = p + q$ si el ARMA tiene o no constante, respectivamente), y sea $\hat{\varphi}_{(p,q)}$ el vector formado por las estimaciones de máxima verosimilitud de dichos k parámetros y de σ_a^2 .

Se propone seleccionar aquel modelo ARMA que minimice el valor de una de las funciones siguientes:

- $AIC(p, q) = -2 \log (L(\hat{\varphi}_{(p,q)})) + 2k$
(Criterio de Información de Akaike).
- $AICc(p, q) = -2 \log (L(\hat{\varphi}_{(p,q)})) + 2(kT + k + 2) / (T - k - 2)$
(Criterio de Información de Akaike corregido).
- $BIC(p, q) = -2 \log (L(\hat{\varphi}_{(p,q)})) + k \log (T)$
(Criterio de Información Bayesiano).

Nota: Se tiene que

$$AICc(p, q) = AIC(p, q) + 2(k + 1)(k + 2) / (T - k - 2).$$

Como se puede ver en la transparencia anterior, la estructura de las funciones AIC, AICc y BIC es similar:

- **Primer sumando:** mide tanto la calidad del ajuste como la credibilidad que le da a la serie de tiempo. Cuanto menor es su valor mejor es el ajuste y mayor la credibilidad que le da a la serie. Su valor disminuye al aumentar el valor de p y/o q (que a su vez implica aumentar k).
- **Segundo sumando:** penaliza el aumento en la cantidad de coeficientes del ARMA. Su valor disminuye al disminuir k .

El modelo que minimiza a una de estas 3 funciones consigue un equilibrio entre ambos sumandos (ambos serán “pequeños”); esto es, un buen ajuste sin demasiados parámetros (que darían problemas tanto a la hora de estimar como de predecir).

Comparativa de los 3 criterios de información

- El criterio BIC es consistente: Si realmente la serie ha sido generada por un ARMA, el BIC selecciona los órdenes correctos con probabilidad 1 (esto no ocurre con los criterios AIC y AICc).
- Los criterios AIC y AICc son asintóticamente eficientes: Si realmente la serie ha sido generada por un AR (posiblemente de orden ∞), el AIC y el AICc seleccionan el modelo que da lugar al menor error de predicción esperado (esto no ocurre con el criterio BIC).

Continuando con el análisis de la serie (un proceso AR(2) resultaba adecuado para haberla generado), buscamos el par (p, q) que minimice el valor de la función AICc. Dicho par daría lugar al modelo óptimo desde el punto de vista AICc (si bien para que fuese válido debiera superar el análisis de residuos). El modelo ARMA óptimo (AICc) resulta ser un ARMA(1,1) con media nula (equivalentemente, sin constante c), siendo 265.25 el valor de su AICc.

También hay otros modelos cuyos valores AICc distan del mínimo menos de 2 unidades (modelos equivalentes al óptimo). Entre ellos, se encuentra el modelo AR(2) con media nula (AICc=266.20) que habíamos propuesto en base al estudio de las fas y fap (y que superaba el análisis de residuos).

Dicho AR(2) no es más complejo que el óptimo (ambos tienen la misma cantidad de parámetros) y es más fácil de interpretar que éste. Por tanto, podríamos quedarnos con el AR(2) en lugar de con el ARMA(1,1).

```
auto.arima(...)
```

Caso general: Selección del modelo en procesos ARIMA(p,d,q)

Como consecuencia de la estrecha relación existente entre los procesos ARIMA y los procesos ARMA:

$$\{X_t\}_t \text{ es ARIMA}(p,d,q) \Leftrightarrow (1 - B)^d X_t \text{ es ARMA}(p,q),$$

se tiene que para seleccionar un modelo ARIMA es suficiente:

- Identificar el orden de diferenciación regular, d .
- Saber seleccionar un modelo ARMA.

Predicción puntual en procesos ARIMA(p,d,q)

Una vez que se ha ajustado a una serie de tiempo un modelo ARIMA(p,d,q) y se ha comprobado que supera la diagnosis, la última etapa consiste en predecir futuros valores del proceso. Como es de esperar, esta etapa se basa en la expresión del modelo ajustado. Consta de 3 pasos:

- 1 Expandir la representación del ARIMA ajustado, de modo que en el lado izquierdo de la igualdad aparezca únicamente X_t .
- 2 Reescribir la representación obtenida en el Paso 1, pero cambiando t por $T + h$.
- 3 En el lado derecho de la representación obtenida en el Paso 2, reemplazar observaciones pasadas y actual (X_t con $t \leq T$) por sus valores, futuras observaciones (X_t con $t > T$) por sus predicciones, innovaciones pasadas y actual (a_t con $t \leq T$) por los residuos correspondientes (\hat{a}_t con $t \leq T$), y futuras innovaciones (a_t con $t > T$) por cero.

La predicción así obtenida se denomina predicción con origen en T y horizonte h , y la denotaremos por $\hat{x}_T(h)$.

A continuación, utilizamos el siguiente modelo ARIMA(2,1,1) ajustado para ilustrar la construcción de predicciones a horizontes $h = 1$ y $h = 2$:

$$(1 - \hat{\phi}_1 B - \hat{\phi}_2 B^2)(1 - B)X_t = \hat{c} + (1 + \hat{\theta}_1 B)a_t.$$

- ❶ Expandiendo la representación anterior, se tiene que:

$$X_t = (1 + \hat{\phi}_1)X_{t-1} - (\hat{\phi}_1 - \hat{\phi}_2)X_{t-2} - \hat{\phi}_2 X_{t-3} + \hat{c} + a_t + \hat{\theta}_1 a_{t-1}.$$

- ❷ Considerando $t = T + h$ en la representación obtenida en el Paso 1:

$$\begin{aligned} X_{T+h} = & (1 + \hat{\phi}_1)X_{T+h-1} - (\hat{\phi}_1 - \hat{\phi}_2)X_{T+h-2} - \hat{\phi}_2 X_{T+h-3} \\ & + \hat{c} + a_{T+h} + \hat{\theta}_1 a_{T+h-1}. \end{aligned}$$

- ② En particular, considerando $h = 1$ y $h = 2$, se tienen, respectivamente:

$$X_{T+1} = (1 + \hat{\phi}_1)X_T - (\hat{\phi}_1 - \hat{\phi}_2)X_{T-1} - \hat{\phi}_2X_{T-2} + \hat{c} + a_{T+1} + \hat{\theta}_1a_T$$

y

$$X_{T+2} = (1 + \hat{\phi}_1)X_{T+1} - (\hat{\phi}_1 - \hat{\phi}_2)X_T - \hat{\phi}_2X_{T-1} + \hat{c} + a_{T+2} + \hat{\theta}_1a_{T+1}.$$

- ③ Reemplazando observaciones e innovaciones como se indica en el Paso 3:

$$\hat{x}_T(1) = (1 + \hat{\phi}_1)x_T - (\hat{\phi}_1 - \hat{\phi}_2)x_{T-1} - \hat{\phi}_2x_{T-2} + \hat{c} + \hat{\theta}_1\hat{a}_T$$

y

$$\hat{x}_T(2) = (1 + \hat{\phi}_1)\hat{x}_T(1) - (\hat{\phi}_1 - \hat{\phi}_2)x_T - \hat{\phi}_2x_{T-1} + \hat{c}.$$

La constante, c , y el orden de diferenciación, d , juegan papeles muy importantes en las predicciones a largo plazo (esto es, a medida que aumenta el horizonte de predicción h). Así:

- Si $d = 0$ (proceso ARMA), las predicciones puntuales a largo plazo convergerán a la media del proceso.
- Si $d = 1$ (proceso no estacionario), las predicciones puntuales a largo plazo convergerán, si $c = 0$, a una constante, mientras que si $c \neq 0$ tenderán a seguir una línea recta.
- Si $d = 2$ (proceso no estacionario), las predicciones puntuales a largo plazo tenderán, si $c = 0$, a seguir una línea recta, mientras que si $c \neq 0$ su tendencia será cuadrática.

Nota: Si $d > 1$, lo usual es no incluir la constante c en el modelo puesto que una tendencia cuadrática (o de grado mayor que 2) resulta muy peligrosa a la hora de hacer predicción.

Intervalos de predicción en procesos ARIMA(p,d,q) (I)

Además de obtener predicciones puntuales de valores futuros del proceso, resulta fundamental disponer de algún conocimiento acerca de la precisión de dichas predicciones. Esto puede tenerse a través de los intervalos de predicción. La teoría a seguir para su construcción es:

- 1 Se parte de la expresión

$$X_t = c + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots$$

donde los ψ_i son parámetros desconocidos (función de θ_j y ϕ_k).

- 2 Se asume que “T es grande” y $\{a_t\}_t$ **gaussiano**. Bajo estas condiciones, se tiene que el error de predicción a horizonte h ,

$$e_T(h) = X_{T+h} - \hat{X}_T(h),$$

verifica:

$$e_T(h) \approx N(0, \sigma_a^2 (1 + \psi_1^2 + \cdots + \psi_{h-1}^2)).$$

- 3 Por tanto, un intervalo de predicción (al 95 %) para el valor de X_{T+k} será

$$\left(\hat{X}_T(h) \pm 1.96 \sqrt{\sigma_a^2 (1 + \psi_1^2 + \cdots + \psi_{h-1}^2)} \right).$$

Observación 1

Es importante notar que la incertidumbre existente en las predicciones depende mucho de si el modelo en base al cual se predice es estacionario ($d = 0$) o no estacionario ($d \neq 0$).

- En modelos estacionarios, se tiene que la serie $\sum_{i=1}^{\infty} \psi_i^2$ es convergente (esto es, $\sum_{i=1}^{\infty} \psi_i^2 = \lim_{h \rightarrow \infty} \sum_{i=1}^h \psi_i^2 < \infty$). Por tanto, la varianza del error de predicción a largo plazo (esto es, a medida que aumenta el horizonte de predicción h) converge a una constante (concretamente, a la varianza marginal del proceso). Como consecuencia, “la incertidumbre de la predicción puntual es finita” (el radio máximo de los intervalos de predicción es finito).
- En modelos no estacionarios, se tiene que la serie $\sum_{i=1}^{\infty} \psi_i^2$ no es convergente. Por tanto, la varianza del error de predicción a largo plazo crece sin límite hacia infinito, y lo mismo le ocurre a la incertidumbre de la predicción puntual (a largo plazo, no se puede prever el comportamiento de un proceso no estacionario).

Observación 2

Para poder llevar a la práctica la teoría que acabamos de mostrar para construir los intervalos de predicción, es necesario conocer los órdenes del ARIMA (p, d y q), sus coeficientes (c, θ_j y ϕ_k ; equivalentemente, c y ψ_i) y la varianza de las innovaciones (σ_a^2).

Puesto que ni los órdenes ni los parámetros del ARIMA son conocidos, en la práctica lo que se hace es estimarlos (con lo que dejan de ser fijos, pues pasan a depender de la serie) y utilizar dichas estimaciones en la expresión del intervalo de predicción en lugar de sus valores reales (desconocidos).

Esto provoca un cambio (aumento) en la varianza del error de predicción y, por tanto, en la amplitud de los intervalos de predicción (pues en el cálculo de la distribución asintótica del error de predicción únicamente se tuvo en cuenta la variación de las innovaciones, considerando por tanto todo lo demás fijo).

Este cambio se puede considerar despreciable si el tamaño de la serie es “grande”.

Intervalos de predicción en procesos ARIMA(p,d,q) (II)

Como hemos notado anteriormente, los intervalos de predicción que hemos presentado se basan en la hipótesis de gaussianidad de las innovaciones; por tanto, si dicha hipótesis falla esos intervalos carecen de validez.

En esta transparencia y en la siguiente, presentaremos intervalos de predicción que **no necesitan de la hipótesis de gaussianidad**. La idea se basa en:

- 1 Simular los próximos k valores de la serie de tiempo: $\{x_{T+1}, x_{T+2}, \dots, x_{T+k}\}$. Este conjunto sería un “futuro posible”.
- 2 Repetir el Paso 1 muchas veces, obteniendo por tanto muchos futuros posibles. Así, si lo repitiésemos B veces, tendríamos los B futuros posibles:

$$\left\{ x_{T+1}^{(1)}, x_{T+2}^{(1)}, \dots, x_{T+k}^{(1)} \right\}, \dots, \left\{ x_{T+1}^{(B)}, x_{T+2}^{(B)}, \dots, x_{T+k}^{(B)} \right\}.$$

- 3 Los extremos del correspondiente intervalo de predicción se obtendrían, para cada horizonte $1 \leq h \leq k$, a partir de los percentiles de orden $\alpha/2$ y $1 - \alpha/2$ de los valores $\left\{ x_{T+h}^{(j)} \right\}_{j=1}^B$.

¿Cómo simulamos cada futuro posible $\{x_{T+1}, x_{T+2}, \dots, x_{T+k}\}$?

A través del **bootstrap** sobre los residuos. Concretamente, si denotamos por e_t al error de predicción con origen en $t - 1$ y horizonte 1; esto es,

$$e_t := e_{t-1}(1) = X_t - \hat{X}_{t-1}(1),$$

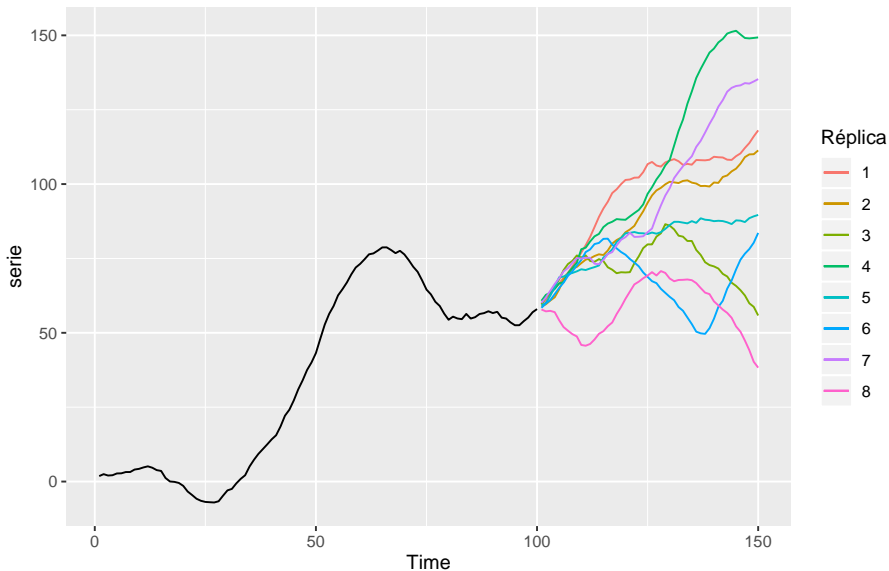
se tiene que

$$X_t = \hat{X}_{t-1}(1) + e_t.$$

A partir de aquí, los pasos a seguir son:

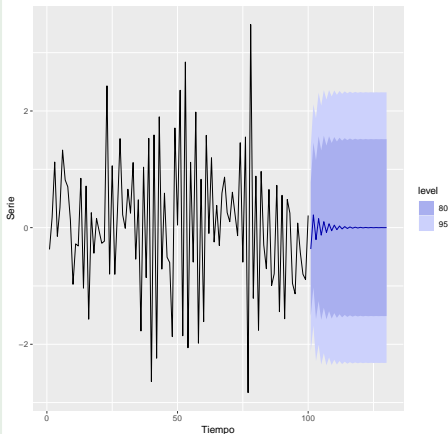
- ❶ Simular x_{T+1} : se tiene que $X_{T+1} = \hat{X}_T(1) + e_{T+1}$. Puesto que los errores futuros (en particular, e_{T+1}) son similares a los errores pasados (esto es, a los residuos $\hat{a}_1, \dots, \hat{a}_T$), se obtiene un valor para e_{T+1} muestreando los residuos.
- ❷ Simular x_{T+2} : El valor simulado en el Paso 1 (x_{T+1}) se añade al conjunto de valores observados, lo que nos permite construir el primer sumando de la igualdad $X_{T+2} = \hat{X}_{T+1}(1) + e_{T+2}$. Un valor para e_{T+2} se obtiene como en el Paso 1 (muestreando los residuos).
- ❸ Se repite el procedimiento anterior hasta completar el futuro posible $\{x_{T+1}, x_{T+2}, \dots, x_{T+k}\}$.

B=8 futuros posibles (k=50) obtenidos vía bootstrap sobre los residuos

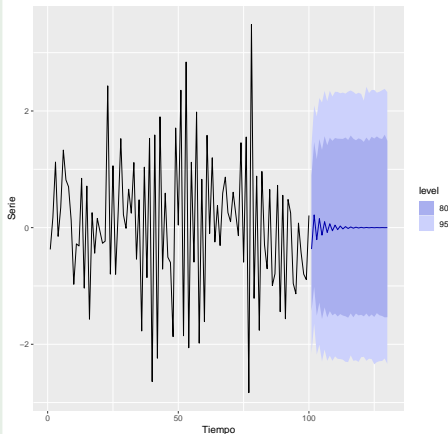


Finalizamos el estudio de la serie simulada: predic. punt. e interv. de predic.

(I) IP vía distribución asintótica



(II) IP vía bootstrap sobre los resids



```
forecast(...), forecast(..., bootstrap=TRUE); autoplot(...)
```

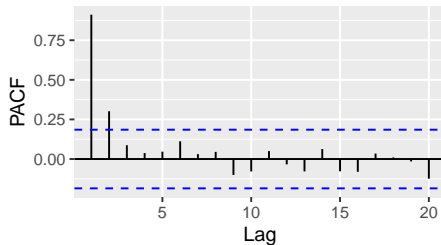
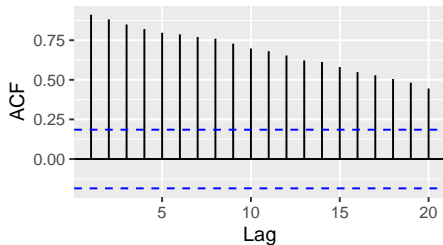
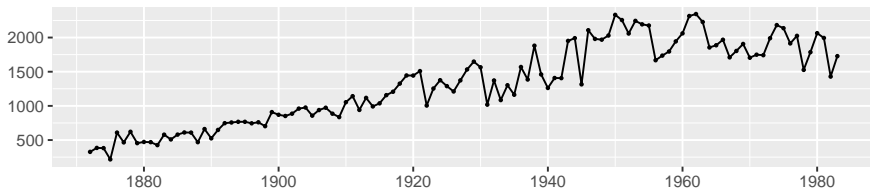
Presentamos a continuación un ejemplo con datos reales en el que se hace uso de gran parte de lo expuesto hasta este momento.

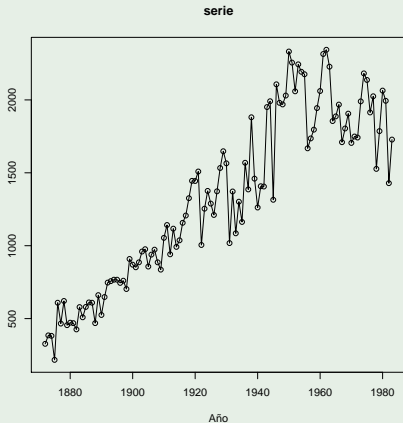
La serie de tiempo que analizaremos se corresponde con la **producción anual de tabaco** en EE.UU. (1872-1983).

El objetivo es proponer un modelo susceptible de haberla generado y, en base a él, predecir la producción de tabaco en los próximos 5 años (1984-1988).

1: Identificación del modelo (gráf. secuencial, fas y fap muestrales)

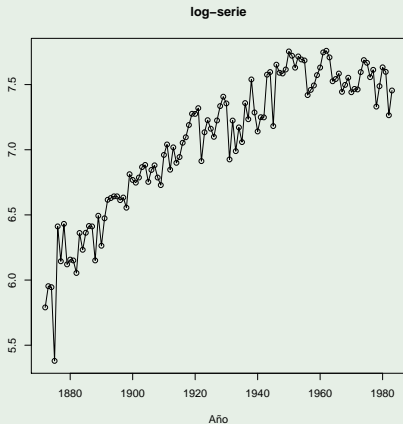
serie





El gráfico de la izquierda muestra presencia de **heterocedasticidad** y tendencia.

Comenzamos transformando la serie para tratar de estabilizar la variabilidad. Puesto que ésta aumenta con el nivel de la serie (quizás la desviación típica sea proporcional a la media), le aplicamos a la producción de tabaco la función logaritmo neperiano.

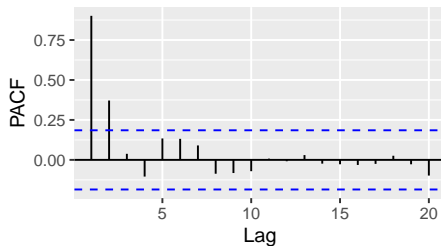
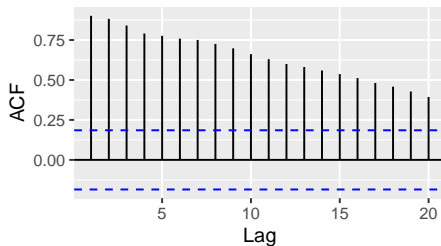
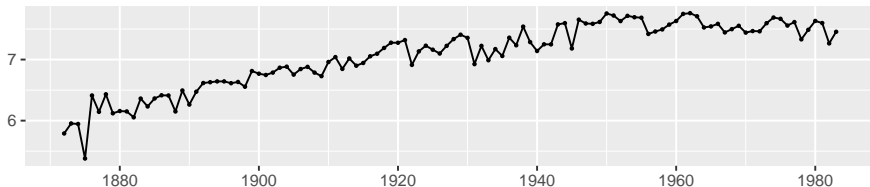


El gráfico de la izquierda muestra que:

- La varianza de la producción se ha estabilizado al aplicarle la función logaritmo neperiano.
- La serie transformada tiene **tendencia**.

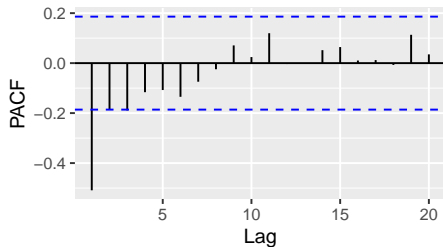
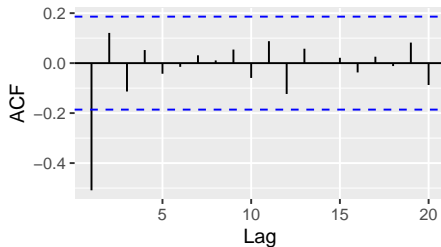
La diferenciación regular podría eliminar la tendencia.

log-serie



Sugerencia: Tanto el gráfico secuencial como la fas (ACF) indican la presencia de tendencia.

dif-log-serie



Sugerencia: MA(1) para dif-log-serie; esto es, ARIMA(0,1,1) para log-serie.

El modelo identificado (ARIMA(0,1,1)) se puede expresar como

$$(1 - B) Y_t = c + (1 + \theta_1 B) a_t,$$

siendo $Y_t = \log(X_t)$ y X_t la producción de tabaco del año t .

Una representación equivalente es

$$Y_t = c + Y_{t-1} + a_t + \theta_1 a_{t-1}.$$

Nota: Obsérvese que, puesto que el proceso diferenciado no tiene parte AR, la constante c coincide con su media μ (no la de Y_t , sino la de $Y_t - Y_{t-1}$).

2: Selección del modelo (criterio AICc)

El modelo ARIMA(0,1,1) con constante resultó ser el de menor AICc (AICc= -82.21). Nótese que este modelo coincide con el previamente identificado en base a las fas y fap muestrales.

Ninguno de los demás modelos evaluados alcanzó un AICc que distase del mínimo menos de 2 unidades.

3: Estimación del modelo identificado ARIMA(0,1,1)

Estimando los parámetros por máxima verosimilitud resulta:

$$\hat{\theta}_1 = -0.6899 (0.0698), \hat{c} = 0.0145 (0.0049),$$

$$\text{y } \hat{\sigma}_a^2 = 0.02672.$$

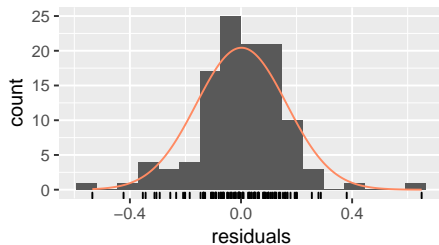
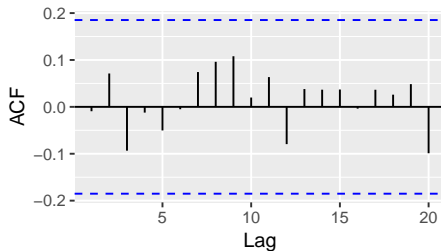
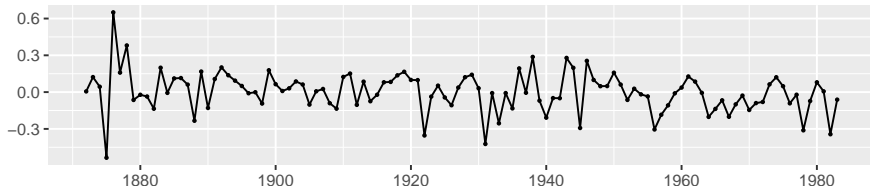
Nótese que tanto $\hat{\theta}_1$ como \hat{c} son significativamente distintos de cero (al 5 %), pues:

$$|-0.6899| \geq 1.96 \times 0.0698 \text{ y } |0.0145| \geq 1.96 \times 0.0049.$$

Puesto que el modelo no tiene parte AR, la constante c del modelo coincide con la **media** μ del **proceso diferenciado**. Por tanto, $\hat{\mu} = 0.0145$.

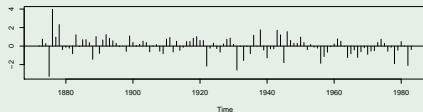
4: **Diagnosis** del modelo ARIMA(0,1,1)

Residuals from ARIMA(0,1,1) with drift



Sugerencia: Posible presencia de valores atípicos y ausencia de normalidad.

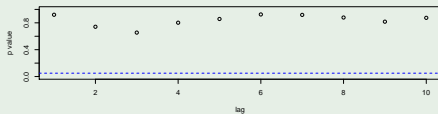
Standardized Residuals



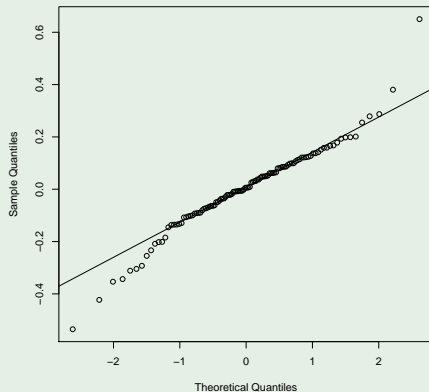
ACF of Residuals



p values for Ljung-Box statistic



Normal Q-Q Plot



Sugerencia: Independencia y posible ausencia de normalidad provocada por la presencia de valores atípicos.

Contrastes de independencia, media cero y normalidad

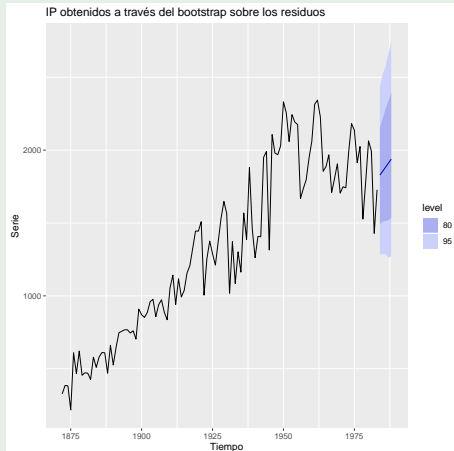
- Test de Ljung-Box con $H = 10$ (8 g.l.): p -valor=0.7314
- $\mu_a = 0$: p -valor=0.9118
- Normalidad:
 - Jarque-Bera:
 p -valor= $4.701e - 07$
 - Shapiro-Wilk: p -valor=
0.001877

Conclusión: Un modelo ARIMA(0,1,1) con constante e innovaciones no gaussianas resulta **adecuado** como generador de la serie de la producción de tabaco (transformada a través de la función logaritmo neperiano).

5: Predicción en base al modelo ARIMA(0,1,1)

El ARIMA(0,1,1) que hemos seleccionado, estimado y chequeado fue utilizado para realizar predicciones a horizontes de predicción $k = 1, \dots, 5$.

Tanto las predicciones puntuales como los intervalos de predicción bootstrap (lila), junto con la serie histórica (negro), se muestran en el gráfico de la derecha. Puesto que no tenemos gaussianidad, no presentamos los intervalos de predicción basados en la distribución asintótica de los errores de predicción.



La clase de procesos ARIMA(p,d,q) que acabamos de estudiar (ARIMAs no estacionales):

- Captura no estacionariedades provocadas por la presencia de tendencia (incluso no determinista).
- **No captura** no estacionariedades provocadas por la presencia de **componente estacional**.

A continuación, ampliaremos la clase de procesos ARIMA (no estacionales) estudiada, de modo que la nueva clase sea capaz de modelizar no estacionariedades provocadas tanto por la presencia de tendencia (determinista o estocástica) como por la presencia de componente estacional (determinista o estocástica).

En la construcción de los procesos ARIMA (no estacionales) ya estudiados jugaban un papel fundamental los procesos ARMA (no estacionales).

Recuérdese que:

$$\{X_t\}_t \text{ es ARIMA}(p,d,q) \Leftrightarrow (1 - B)^d X_t \text{ es ARMA}(p,q).$$

Del mismo modo, necesitaremos de los procesos **ARMA estacionales** para construir la nueva clase de procesos ARIMA estacionales.

Los procesos ARMA(p,q) que ya hemos estudiado (ARMA no estacionales) modelizan la **dependencia regular**: dependencia entre la variable X_t y observaciones/innovaciones consecutivas ocurridas en el pasado inmediato. Por ejemplo:

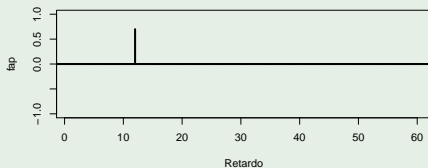
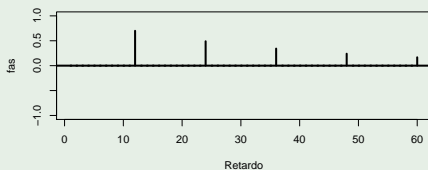
- AR(1): $X_t = c + \phi_1 X_{t-1} + a_t$.
- MA(2): $X_t = c + \theta_1 a_{t-1} + \theta_2 a_{t-2} + a_t$.
- ARMA(1,1): $X_t = c + \phi_1 X_{t-1} + \theta_1 a_{t-1} + a_t$.

Los procesos ARMA estacionales modelizan la **dependencia estacional**: dependencia entre la variable X_t y observaciones/innovaciones ocurridas en instantes separados por múltiplos del período estacional s . Así, si $s=12$, se tienen, por ejemplo:

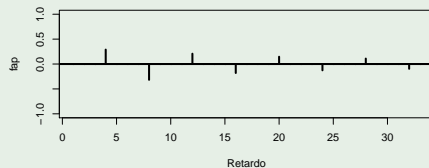
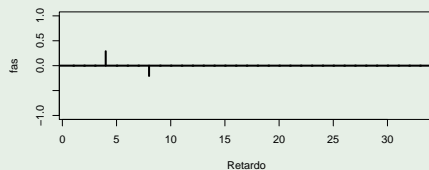
- AR(1)₁₂: $X_t = c + \Phi_1 X_{t-12} + a_t$.
- MA(2)₁₂: $X_t = c + \Theta_1 a_{t-12} + \Theta_2 a_{t-24} + a_t$.
- ARMA(1,1)₁₂: $X_t = c + \Phi_1 X_{t-12} + \Theta_1 a_{t-12} + a_t$.

Ejemplo de la fas y la fap de procesos $AR(1)_{12}$ y $MA(2)_4$

$AR(1)_{12}$

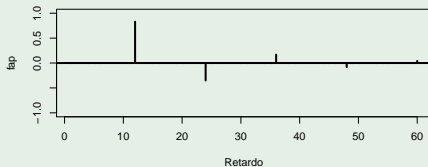
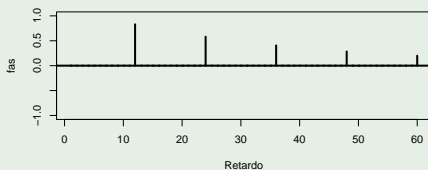


$MA(2)_4$

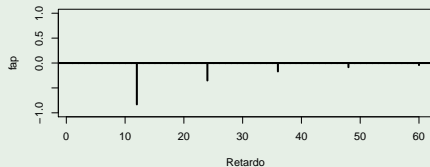
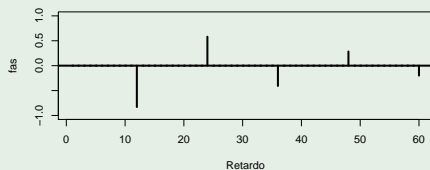


Ejemplo de la fas y la fap de procesos $\text{ARMA}(1,1)_{12}$

$\text{ARMA}(1,1)_{12}: \Phi_1 > 0, \Theta_1 > 0$



$\text{ARMA}(1,1)_{12}: \Phi_1 < 0, \Theta_1 < 0$



Procesos $\text{ARMA}(P,Q)_s$

Un proceso estacionario $\{X_t\}_t$ que admite la representación

$$X_t = c + \Phi_1 X_{t-s} + \Phi_2 X_{t-2s} + \cdots + \Phi_P X_{t-Ps} \\ + a_t + \Theta_1 a_{t-s} + \Theta_2 a_{t-2s} + \cdots + \Theta_Q a_{t-Qs},$$

donde $c, \Phi_1, \dots, \Phi_P, \Theta_1, \dots, \Theta_Q$ ($\Phi_P \neq 0 \neq \Theta_Q$) son constantes, se conoce como un proceso $\text{ARMA}(P,Q)_s$ (proceso ARMA estacional).

Se tiene que:

- Un $\text{ARMA}(P,Q)_s$ es, en particular, un $\text{ARMA}(sP,sQ)$ con muchos coeficientes nulos. Por tanto, las condiciones de estacionariedad, causalidad e invertibilidad se deducen de las de los $\text{ARMA}(p,q)$ (no estacionales).
- $\text{ARMA}(P,0)_s \Leftrightarrow \text{AR}(P)_s$.
- $\text{ARMA}(0,Q)_s \Leftrightarrow \text{MA}(Q)_s$.

La ecuación que define al proceso $\text{ARMA}(P,Q)_s$

$$X_t = c + \Phi_1 X_{t-s} + \Phi_2 X_{t-2s} + \cdots + \Phi_P X_{t-Ps} \\ + a_t + \Theta_1 a_{t-s} + \Theta_2 a_{t-2s} + \cdots + \Theta_Q a_{t-Qs},$$

se puede escribir en la forma compacta

$$\Phi(B^s) X_t = c + \Theta(B^s) a_t,$$

donde

$$\Phi(B^s) = (1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps}), \\ \Theta(B^s) = (1 + \Theta_1 B^s + \Theta_2 B^{2s} + \cdots + \Theta_Q B^{Qs})$$

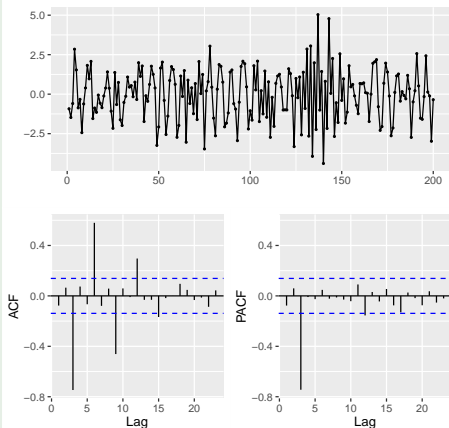
y B^s denota al **operador retardo estacional**, definido por $B^s X_t = X_{t-s}$.

La siguiente tabla resume la estructura general de las correlaciones simples y parciales de los modelos anteriormente presentados:

	f_{as}^*	f_{ap}^*
$AR(P)_s$	Retardos $s, 2s, \dots$: muchos coeficientes no nulos	El último retardo en el que no se anula es P_s
$MA(Q)_s$	El último retardo en el que no se anula es Q_s	Retardos $s, 2s, \dots$: muchos coeficientes no nulos
$ARMA(P,Q)_s$	Retardos $s, 2s, \dots$: muchos coeficientes no nulos	Retardos $s, 2s, \dots$: muchos coeficientes no nulos

* Los valores en los retardos no estacionales (distintos de ks) son nulos.

Serie, fas y fap

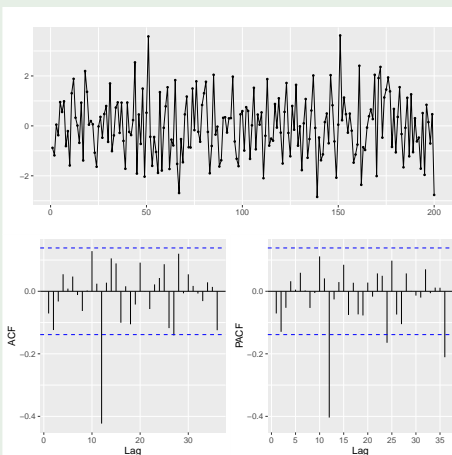


Conclusión

Los gráficos de la izquierda **sugieren** que la serie:

- 1 Es **estacionaria**.
- 2 Ha sido generada por un proceso **AR(1)₃**.

Serie, fas y fap



Conclusión

Los gráficos de la izquierda **sugieren** que la serie:

- 1 Es **estacionaria**.
- 2 Ha sido generada por un proceso **MA(1)₁₂**.

Se tiene que:

- ARMA(p,q) (no estacional): $\phi(B) X_t = c + \theta(B) a_t$. Modeliza la dependencia regular.
- ARMA(P,Q)_s (estacional): $\Phi(B^s) X_t = c + \Theta(B^s) a_t$. Modeliza la dependencia estacional.

Procesos ARMA(p,q) × (P,Q)_s

Combinando los modelos ARMA(p,q) y ARMA(P,Q)_s, podemos modelizar conjuntamente la dependencia regular y la estacional. Para ello, es suficiente considerar el modelo

$$\phi(B) \Phi(B^s) X_t = c + \theta(B) \Theta(B^s) a_t.$$

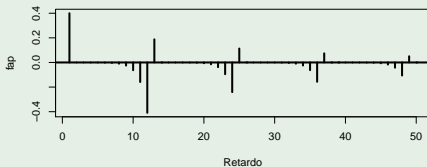
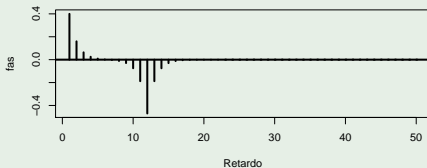
Este modelo se denota por ARMA(p,q) × (P,Q)_s (ARMA estacional multiplicativo) y es, en particular, un ARMA(p+sP,q+sQ) con muchos coeficientes nulos.

Procesos ARMA estacionales multiplicativos: Ejemplos

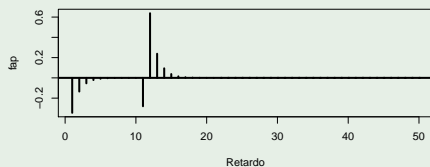
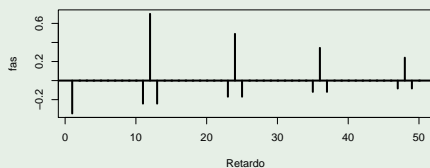
- El proceso $\text{ARMA}(1,0) \times (0,1)_{12}$ (o $\text{AR}(1) \times \text{MA}(1)_{12}$).
 - ① $(1 - \phi_1 B) X_t = c + (1 + \Theta_1 B^{12}) a_t.$
 - ② $X_t = c + \phi_1 X_{t-1} + a_t + \Theta_1 a_{t-12}.$
- El proceso $\text{ARMA}(0,1) \times (1,0)_{12}$ (o $\text{MA}(1) \times \text{AR}(1)_{12}$).
 - ① $(1 - \Phi_1 B^{12}) X_t = c + (1 + \theta_1 B) a_t.$
 - ② $X_t = c + \Phi_1 X_{t-12} + a_t + \theta_1 a_{t-1}.$
- El proceso $\text{ARMA}(1,0) \times (1,0)_{12}$ (o $\text{AR}(1) \times \text{AR}(1)_{12}$).
 - ① $(1 - \phi_1 B) (1 - \Phi_1 B^{12}) X_t = c + a_t.$
 - ② $X_t = c + \phi_1 X_{t-1} + \Phi_1 X_{t-12} - \phi_1 \Phi_1 X_{t-13} + a_t.$
- El proceso $\text{ARMA}(0,1) \times (0,1)_{12}$ (o $\text{MA}(1) \times \text{MA}(1)_{12}$).
 - ① $X_t = c + (1 + \theta_1 B) (1 + \Theta_1 B^{12}) a_t.$
 - ② $X_t = c + a_t + \theta_1 a_{t-1} + \Theta_1 a_{t-12} + \theta_1 \Theta_1 a_{t-13}.$

Ejemplos de la fas y la fap de ARMA estacionales multiplicativos

$AR(1) \times MA(1)_{12}$

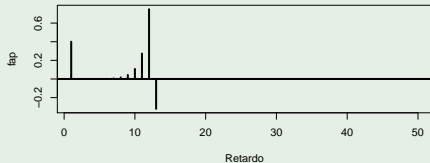
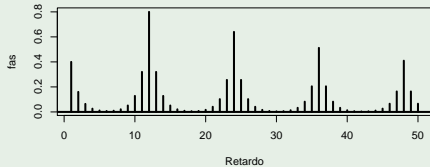


$MA(1) \times AR(1)_{12}$

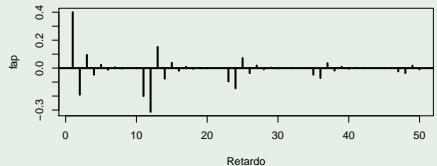
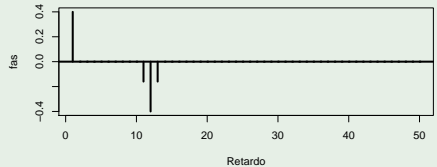


Ejemplos de la fas y la fap de ARMA estacionales multiplicativos

$AR(1) \times AR(1)_{12}$



$MA(1) \times MA(1)_{12}$



Estructura de la fas de un proceso ARMA estacional multiplicativo

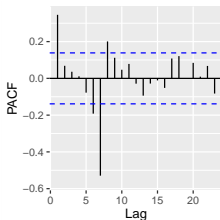
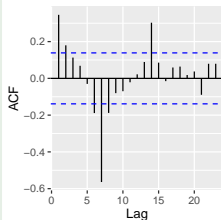
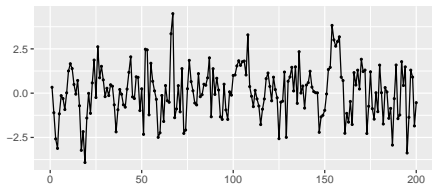
- En los retardos bajos ($1, 2, \dots, [s/2]$) se observará la fas de la parte regular.
- En los retardos estacionales ($s, 2s, 3s \dots$) se observará la fas de la parte estacional.
- A ambos lados de los retardos estacionales se repetirá la fas de la parte regular (invertida, si la fas en el retardo estacional es negativa).

Estructura de la fap de un proceso ARMA estacional multiplicativo

- En los retardos bajos $(1, 2, \dots, [s/2])$ se observará la fap de la parte regular.
- En los retardos estacionales $(s, 2s, 3s \dots)$ se observará la fap de la parte estacional.
- A la derecha de cada retardo estacional aparecerá la fap de la parte regular (invertida, si la fap en el retardo estacional es positiva).
- A la izquierda de cada retardo estacional aparecerá la fas de la parte regular (invertida, si la fap en el retardo estacional es negativa).

A continuación mostraremos dos ejemplos en los que, utilizando la información contenida en las dos transparencias anteriores y también la distribución muestral de $\hat{\rho}_k$ ó $\hat{\alpha}_k$ bajo procesos MA ó AR, respectivamente, se identifican dos procesos ARMA estacionales multiplicativos.

Serie, fas y fap

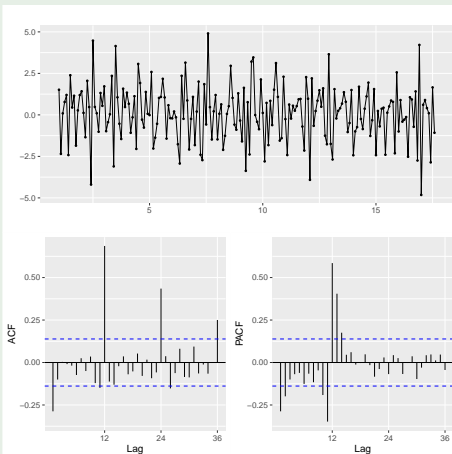


Conclusión

Los gráficos de la izquierda **sugieren** que la serie:

- 1 Es **estacionaria**.
- 2 Ha sido generada por un proceso **$AR(1) \times AR(1)_7$** .

Serie, fas y fap



Conclusión

Los gráficos de la izquierda **sugieren** que la serie:

- 1 Es **estacionaria**.
- 2 Ha sido generada por un proceso **$MA(1) \times AR(1)_{12}$** .

Motivación:

Sea $X_t = S_t + V_t$, donde $\{S_t\}_t$ no es estacionario, $\{V_t\}_t$ sí y

$$\textcircled{1} S_t = S_{t-s} \text{ (componente estacional determinista),}$$

ó

$$\textcircled{2} S_t = S_{t-s} + W_t \text{ donde } \{W_t\}_t \text{ es estacionario con media 0}$$

(componente estacional estocástica).

$\{X_t\}_t$ no es estacionario (suma de un proceso estacionario y de otro no estacionario). Sin embargo, sí lo es el proceso **diferenciado estacionalmente**

$$\textcircled{1} X_t - X_{t-s} = V_t - V_{t-s}$$

ó

$$\textcircled{2} X_t - X_{t-s} = W_t + V_t - V_{t-s}.$$

Conclusión: A veces, la diferenciación estacional consigue eliminar la componente estacional.

Basándonos en los ejemplos anteriores, ante una serie con tendencia y/o componente estacional, sugerimos:

- **Eliminar la tendencia** aplicando d diferencias regulares $((1 - B)^d)$. En general, es suficiente $d \leq 3$.
- **Eliminar la componente estacional** aplicando D diferencias estacionales $((1 - B^s)^D)$. En general, es suficiente $D = 1$.
- Una vez que la serie diferenciada es estacionaria, modelizarla a través de un ARMA:
 - Sólo dependencia regular: $\text{ARMA}(p, q)$.
 - Sólo dependencia estacional: $\text{ARMA}(P, Q)_s$.
 - Ambos tipos de dependencia: $\text{ARMA}(p, q) \times (P, Q)_s$.

Procesos $\text{ARIMA}(p,d,q) \times (P,D,Q)_s$

Un proceso $\text{ARIMA}(p,d,q) \times (P,D,Q)_s$ (o ARIMA estacional multiplicativo) es aquél que, después de aplicarle d diferencias regulares y D diferencias estacionales de periodo s , se convierte en un proceso $\text{ARMA}(p,q) \times (P,Q)_s$.

Equivalentemente:

$\{X_t\}_t$ es un proceso $\text{ARIMA}(p,d,q) \times (P,D,Q)_s$ (o ARIMA estacional multiplicativo) si admite una representación del tipo:

$$\phi(B) \Phi(B^s) (1-B)^d (1-B^s)^D X_t = c + \theta(B) \Theta(B^s) a_t,$$

donde el polinomio $\phi(z) \Phi(z^s)$ no tiene raíces de módulo 1.

Ejemplo: Proceso ARIMA(1,1,1)×(1,1,1)₁₂

La expresión del ARIMA(1,1,1)×(1,1,1)₁₂ es

$$\begin{array}{cccc}
 \text{AR} & \text{AR} & \text{Dif.} & \text{Dif.} \\
 \text{reg.} & \text{est.} & \text{reg.} & \text{est.} \\
 \downarrow & \downarrow & \downarrow & \downarrow \\
 (1 - \phi_1 B) & (1 - \Phi_1 B^{12}) & (1 - B) & (1 - B^{12}) X_t = \\
 \\
 c + (1 + \theta_1 B) & (1 + \Theta_1 B^{12}) a_t \\
 \uparrow & \uparrow \\
 \text{MA} & \text{MA} \\
 \text{reg.} & \text{est.}
 \end{array}$$

Operando en la expresión del ARIMA(1,1,1)×(1,1,1)₁₂

$$(1 - \phi_1 B) (1 - \Phi_1 B^{12}) (1 - B) (1 - B^{12}) X_t = c + (1 + \theta_1 B) (1 + \Theta_1 B^{12}) a_t$$

se obtiene la representación:

$$\begin{aligned} X_t = & c + (1 + \phi_1) X_{t-1} - \phi_1 X_{t-2} + (1 + \Phi_1) X_{t-12} \\ & - (1 + \phi_1 + \Phi_1 + \phi_1 \Phi_1) X_{t-13} \\ & + (\phi_1 + \phi_1 \Phi_1) X_{t-14} - \Phi_1 X_{t-24} \\ & + (\Phi_1 + \phi_1 \Phi_1) X_{t-25} - \phi_1 \Phi_1 X_{t-26} \\ & + a_t + \theta_1 a_{t-1} + \Theta_1 a_{t-12} + \theta_1 \Theta_1 a_{t-13} \end{aligned}$$

El proceso $ARIMA(p,d,q) \times (P,D,Q)_s$:

- Es **estacionario** cuando $d = D = 0$ (se convierte en un proceso $ARMA(p,q) \times (P,Q)_s$).
- Modeliza la **dependencia regular** (p o $q \neq 0$).
- Modeliza la **dependencia estacional** ($s > 1$, y P o $Q \neq 0$).
- **Captura** no estacionariedades provocadas por la presencia de **tendencia** ($d > 0$).
- **Captura** no estacionariedades provocadas por la presencia de **componente estacional** ($s > 1$ y $D > 0$).
- **Generaliza** a todos los procesos que hemos estudiado.
- Es, posiblemente, el proceso **más utilizado** en la modelización de series de tiempo univariantes.

En la práctica, ante una [serie real...](#)

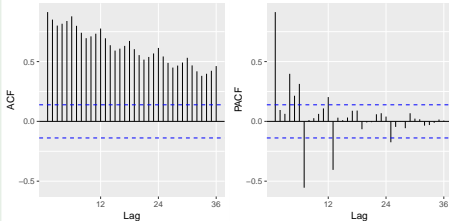
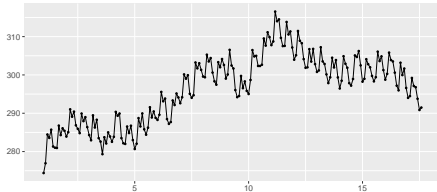
¿cuándo propondremos un [ARIMA](#) estacional como su generador?

Cuando, siendo la serie homocedástica, detectemos la presencia de componente estacional.

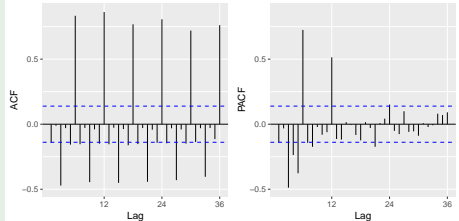
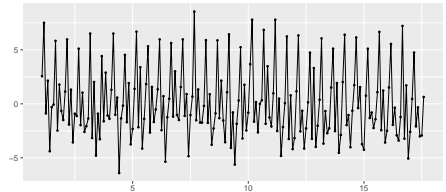
La presencia de componente estacional en una serie (y, por tanto, la necesidad de diferenciarla estacionalmente para eliminarla) suele ser delatada por:

- El gráfico secuencial de la serie.
- La fas muestral:
 - Presenta [fuerte correlación positiva](#) en el [retardo estacional](#) (y, posiblemente, en sus múltiplos),
 - Converge [lentamente a cero](#) a medida que el retardo crece.
 - Presenta [periodicidad](#) del mismo periodo que la serie.

serie

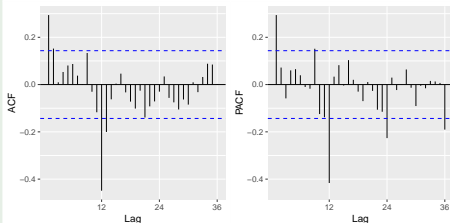
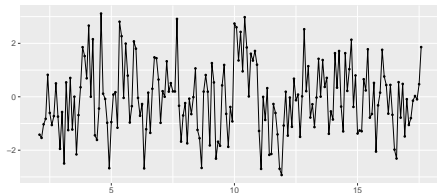


dif-serie



Sugerencia: (Izq) La serie original tiene tendencia y componente estacional. (Der) La serie diferenciada regularmente tiene componente estacional con periodo $s = 12$.

dif12-dif-serie



Conclusión

Los gráficos estudiados **sugieren** que la serie original:

- ❶ No es **estacionaria**.
- ❷ Se convierte en estacionaria a través de una diferenciación regular ($d = 1$) y otra estacional con periodo $s = 12$ ($D = 1$).
- ❸ Ha sido generada por un proceso **ARIMA(1,1,0) × (0,1,1)₁₂**, o quizás por un **ARIMA(0,1,1) × (0,1,1)₁₂**.

Estimación y diagnosis en procesos ARIMA estacionales

Como consecuencia de la estrecha relación existente entre los procesos $\text{ARIMA}(p,d,q) \times (P,D,Q)_s$ y los procesos $\text{ARMA}(p,q) \times (P,Q)_s$:

$$\{X_t\}_t \text{ es } \text{ARIMA}(p,d,q) \times (P,D,Q)_s \Leftrightarrow (1-B)^d (1-B^s)^D X_t \text{ es } \text{ARMA}(p,q) \times (P,Q)_s,$$

se tiene que para estimar y realizar la diagnosis de un modelo $\text{ARIMA}(p,d,q) \times (P,D,Q)_s$ es suficiente:

- Identificar los órdenes de diferenciación regular (d) y estacional (D), y el período estacional (s).
- Saber estimar y realizar la diagnosis de un modelo $\text{ARMA}(p,q) \times (P,Q)_s$.

Selección del modelo en procesos ARIMA estacionales

Como consecuencia de la estrecha relación existente entre los procesos $\text{ARIMA}(p,d,q) \times (P,D,Q)_s$ y los procesos $\text{ARMA}(p,q) \times (P,Q)_s$:

$$\{X_t\}_t \text{ es } \text{ARIMA}(p,d,q) \times (P,D,Q)_s \Leftrightarrow (1-B)^d (1-B^s)^D X_t \text{ es } \text{ARMA}(p,q) \times (P,Q)_s,$$

se tiene que para seleccionar un modelo $\text{ARIMA}(p,d,q) \times (P,D,Q)_s$ es suficiente:

- Identificar los órdenes de diferenciación regular (d) y estacional (D), y el período estacional (s).
- Saber seleccionar un modelo $\text{ARMA}(p,q) \times (P,Q)_s$.

Predicción en procesos ARIMA estacionales

Las técnicas mostradas anteriormente para obtener predicciones puntuales e intervalos de predicción en procesos ARIMA no estacionales ($\text{ARIMA}(p,d,q)$) siguen siendo aplicables en el caso de procesos ARIMA estacionales ($\text{ARIMA}(p,d,q) \times (P,D,Q)_s$).

Una vez que se ha realizado una predicción con origen en T y horizonte h ($\hat{x}_T(h)$) y posteriormente se ha observado el valor real que se pretendía predecir (x_{T+h}), podemos construir el error de predicción (observado):

$$e_T(h) = x_{T+h} - \hat{x}_T(h).$$

(Nótese que hemos utilizado la misma notación, $e_T(h)$, tanto para el error de predicción aleatorio como para el observado)

A continuación se presentan distintas medidas, basadas en $e_T(h)$, de la adecuación de las predicciones obtenidas a diferentes horizontes $h = 1, \dots, H$.

Su interés radica en que a través de ellas podremos comparar la potencia predictiva de diferentes modelos/procedimientos de predicción.

Media de los errores absolutos (MAE, del acrónimo en inglés)

$$MAE = \frac{1}{H} \sum_{h=1}^H |e_T(h)|.$$

Algunas de sus propiedades de interés son:

- Cálculo sencillo y fácil de interpretar.
- El predictor que minimiza al MAE es la mediana.
- Útil para comparar métodos de predicción aplicados a una serie.
- Depende de las unidades; por tanto, no es válido para comparar la adecuación de métodos de predicción aplicados a series medidas en distintas unidades (o de distintas magnitudes).

Raíz cuadrada de la media de los errores al cuadrado (RMSE)

$$RMSE = \sqrt{\frac{1}{H} \sum_{h=1}^H e_T^2(h)}.$$

Algunas de sus propiedades de interés son:

- Cálculo sencillo, pero más difícil de interpretar que el MAE.
- El predictor que minimiza al RMSE es la media.
- Útil para comparar métodos de predicción aplicados a una serie.
- Depende de las unidades; por tanto, no es válido para comparar la adecuación de métodos de predicción aplicados a series medidas en distintas unidades (o de distintas magnitudes).

Media de los errores porcentuales absolutos (MAPE)

$$MAPE = \frac{1}{H} \sum_{h=1}^H |p_T(h)|,$$

donde

$$p_T(h) = \frac{100e_T(h)}{x_{T+h}}$$

es el error porcentual con origen en T y horizonte h .

Algunas de sus propiedades de interés son:

- Cálculo sencillo y fácil de interpretar.
- No depende de las unidades; por tanto, es válido para comparar la adecuación de métodos de predicción aplicados a series medidas en distintas unidades (o de distintas magnitudes).
- No es apropiado para utilizar sobre series que alcanzan algún valor próximo (o igual) a 0 en el período de interés (es inestable).

Media de los errores escalados absolutos (MASE)

$$MASE = \frac{1}{H} \sum_{h=1}^H |q_T(h)|,$$

donde

$$q_T(h) = \frac{e_T(h)}{\frac{1}{T-1} \sum_{t=2}^T |x_t - x_{t-1}|}.$$

Nótese que $q_T(h)$ (error escalado) se construye escalando $e_T(h)$ a través del MAE obtenido a partir del predictor *naïve* aplicado a la muestra histórica (no a futuros valores) y siempre a horizonte 1 (la predicción *naïve* consiste en predecir el valor de la serie en el instante t (x_t) a través del valor observado en el instante anterior (x_{t-1})).

Nota: Si la serie tuviese componente estacional de periodo s , la predicción *naïve* consistiría en predecir el valor x_t a través del valor observado s instantes antes (x_{t-s}); por tanto, el denominador de $q_T(h)$ debiera cambiarse por $\frac{1}{T-s} \sum_{t=s+1}^T |x_t - x_{t-s}|$.

Es interesante hacer notar que:

- **Interpretación:** un error escalado absoluto a horizonte h ($|q_T(h)|$) menor que 1 indica que el predictor utilizado se comporta mejor, a horizonte h , que la media (sobre la serie histórica) de las predicciones *naïve* a horizonte 1. Análogamente, si es mayor que 1, el predictor utilizado se comporta peor, a horizonte h , que la media (sobre la serie histórica) de las predicciones *naïve* a horizonte 1. Por tanto, el valor del MASE, promedio de los H errores escalados absolutos, compara el comportamiento de un predictor a distintos horizontes con el del *naïve* a horizonte 1 (este último, sobre la serie histórica).
- El MASE no depende de las unidades; por tanto, es válido para comparar la adecuación de métodos de predicción aplicados a **series medidas en distintas unidades** (o de distintas magnitudes): será mejor aquel método que dé lugar a un menor valor del MASE.

Heterocedasticidad

En los modelos teóricos que hemos presentado, la falta de estacionariedad venía provocada por la presencia de tendencia y/o componente estacional.

- Aplicando diferencias (regulares y/o estacionales, respectivamente) conseguíamos eliminar este tipo de no estacionariedad.

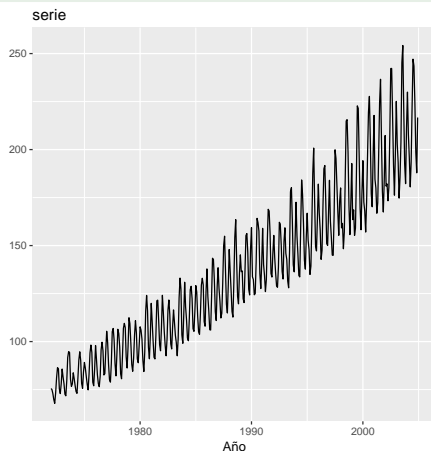
Otra fuente que provoca falta de estacionariedad es la **heterocedasticidad** (la varianza no es constante/estable).

A continuación veremos cómo eliminar la heterocedasticidad.

En el gráfico de la derecha, se intuye que la **variabilidad** de la serie (consumo de electricidad...) **no** es **constante**.

Concretamente, parece que la variabilidad aumenta al hacerlo el nivel de la serie.

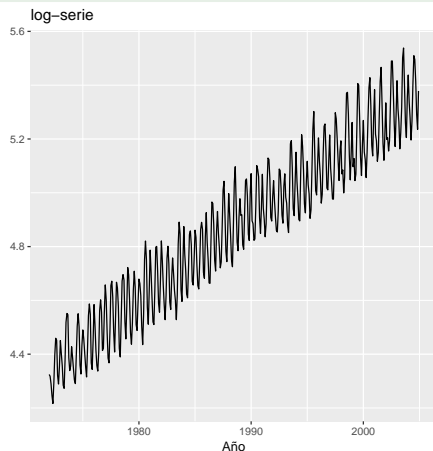
Heterocedasticidad



En el gráfico de la derecha, se muestra la **serie transformada** a través de la función **logaritmo neperiano** ($\log(\cdot)$).

Se observa que la aplicación de dicha función ha conseguido **estabilizar la varianza**.

Homocedasticidad



Transformaciones para estabilizar la varianza

Box-Cox

La familia de transformaciones de **Box-Cox** se define como aquella que transforma a x_t en:

$$\begin{cases} \frac{x_t^\lambda - 1}{\lambda}, & \text{si } \lambda \neq 0 \\ \log(x_t), & \text{si } \lambda = 0 \end{cases}$$

Estas transformaciones se aplican únicamente a series, $\{x_t\}$, cuyos valores son todos positivos.

- Si la serie contiene algún valor nulo o negativo, la transformación se aplicará a la serie $\{y_t\}$, donde $y_t = x_t + c$ siendo c una constante que garantice que $y_t > 0 \forall t$.
- Si la **desviación típica** es una **función** potencial de la **media** ($\sigma_t = k\mu_t^{1-\lambda}$), entonces la transformación de Box-Cox con parámetro λ consigue estabilizar la varianza.
- Si $\sigma_t = k\mu_t$ se tiene que $\lambda = 0$ y la aplicación del logaritmo neperiano estabiliza la varianza.

En la práctica, ante una **serie real**,...

¿de qué métodos disponemos para proponer un **valor apropiado para λ** ?

Un método muy intuitivo consiste en aplicar la función $\log(\cdot)$ (logaritmo neperiano) a la igualdad $\sigma_t = k\mu_t^{1-\lambda}$, obteniendo

$$\log(\sigma_t) = a + b \log(\mu_t),$$

donde $a = \log(k)$ y $b = 1 - \lambda$. Dividiendo la serie en I subseries de observaciones consecutivas*, se obtienen, a partir de cada una de dichas subseries, pares de estimaciones $(\hat{\mu}_i, \hat{\sigma}_i)$ (media y desviación típica). Finalmente, se ajusta un modelo lineal a $\{(\log(\hat{\mu}_i), \log(\hat{\sigma}_i))\}$ y, a partir de la estimación de b , se obtiene una estimación para el valor de λ : $\hat{\lambda} = 1 - \hat{b}$.

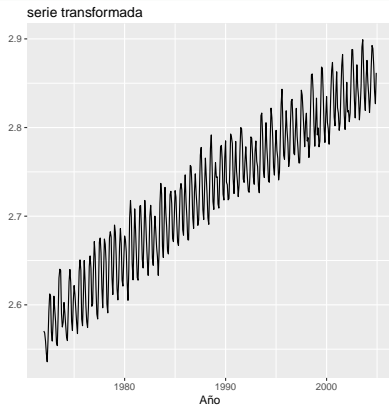
Nota*: Todos las subseries deben tener el mismo tamaño: por lo general, entre 5 y 10 para series no estacionales, y s (período estacional) para series estacionales.

Nota: Es necesario que la igualdad de partida sea cierta (lo cual se puede chequear a través del coeficiente de determinación del ajuste lineal).

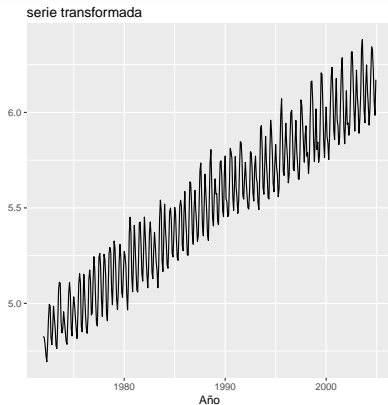
Por otro lado, la función `BoxCox.lambda(...)` del paquete `forecast` (que se carga automáticamente al cargar el paquete `fpp2`) permite obtener un valor para λ a través de dos métodos:

- **Método de Guerrero (1993):** Calcula el valor de λ que minimiza el coeficiente de variación de $\left\{ \hat{\sigma}_i / \hat{\mu}_i^{1-\lambda} \right\}_{i=1}^I$, y se basa en la misma igualdad ($\sigma_t = k\mu_t^{1-\lambda}$) que el método expuesto en la transparencia anterior. La función `BoxCox.lambda(...)` utiliza este método por defecto.
- **Máxima verosimilitud:** Calcula el valor de λ que da lugar al modelo lineal más verosímil ajustado a la serie. Si la serie no tiene componente estacional, el modelo incluye una tendencia lineal temporal; si tiene componente estacional, el modelo incluye, además, variables “dummy” (variables artificiales o ficticias) para manejar dicha componente.

Método de Guerrero: $\hat{\lambda} = -0.266$



Máxima verosimilitud: $\hat{\lambda} = 0.05$

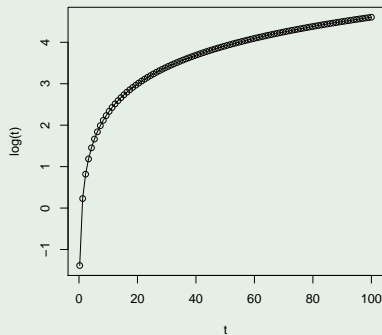


`autoplot(BoxCox(...), ...)`

La transformación logarítmica (I)

La transformación logarítmica ($\lambda = 0$) es, posiblemente, la más utilizada en la práctica para estabilizar la variabilidad en situaciones en la que ésta aumenta al aumentar el nivel.

El gráfico de la derecha puede ayudar a comprender el por qué de que esta transformación funcione en algunas de dichas situaciones: representa los pares $(t, \log(t))$, donde los valores de t están igualmente espaciados sobre el intervalo $[0,1]$.



Valores de t : igualmente espaciados. Valores de $\log(t)$: la distancia entre consecutivos disminuye al aumentar t .

La transformación logarítmica (II)

La transformación $\log(\cdot)$ posee la siguiente propiedad, que contribuye a hacerla popular entre las transformaciones de Box-Cox (en especial en series financieras):

Si denotamos

$$R_t = \frac{x_t - x_{t-1}}{x_{t-1}},$$

se tiene que R_t es la proporción que sobre x_{t-1} representa el cambio de x_{t-1} a x_t (así, si la serie mide el precio o valor de un activo, R_t se denomina el “rendimiento” de la serie).

Si R_t es pequeño, se puede utilizar la aproximación

$$R_t \approx \log(x_t) - \log(x_{t-1}).$$

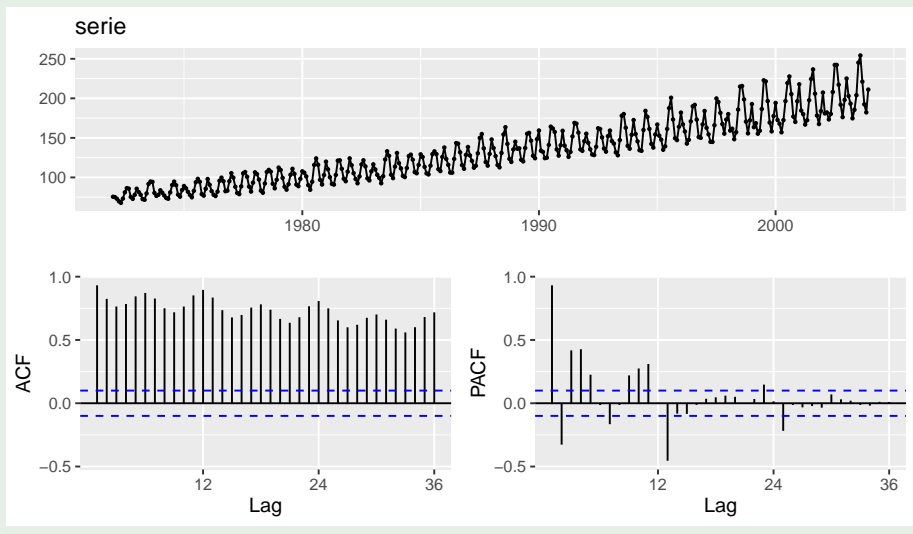
Por tanto, se tiene que la serie en logaritmos diferenciada representa el rendimiento de la serie original.

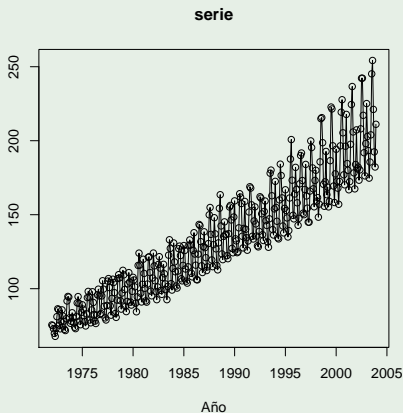
Presentamos a continuación un ejemplo con datos reales en el que se hace uso de gran parte de lo expuesto en este capítulo.

Para ello:

- Dividiremos en 2 trozos la serie del **consumo mensual de electricidad** en EE.UU. (1972-2004) (introducida en el primer capítulo):
 - ➊ Consumo entre los meses de enero 1972 y diciembre 2003 ($T = 384$).
 - ➋ Consumo durante los próximos 12 meses.
- El primer trozo será utilizado para seleccionar y ajustar un modelo.
- En base a dicho modelo, realizaremos predicciones para el consumo correspondiente a los próximos 12 meses, que serán comparadas con los consumos reales (segundo trozo).

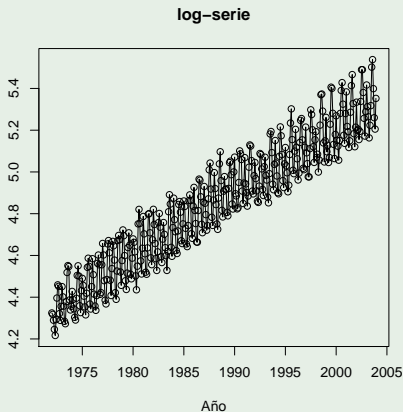
1: Identificación del modelo (gráf. secuencial, fas y fap muestrales)





El gráfico de la izquierda muestra presencia de **heterocedasticidad** (también tendencia y componente estacional).

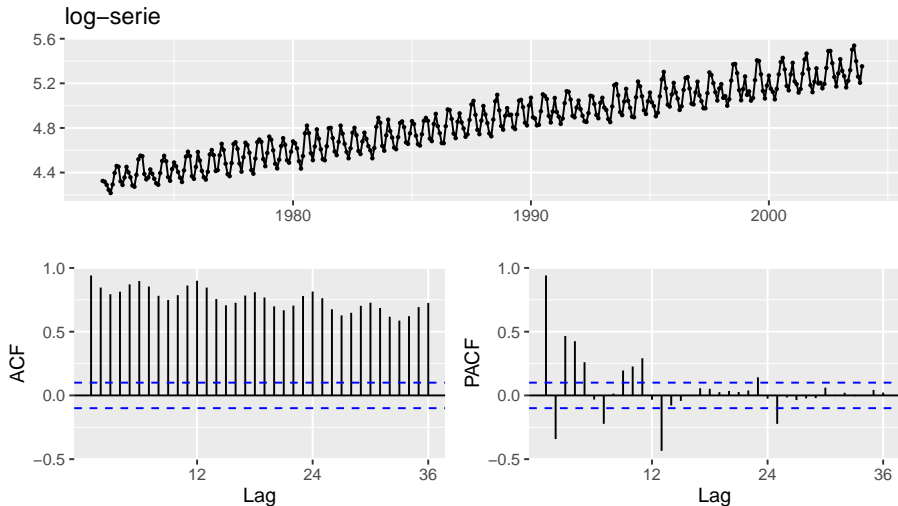
Comenzamos transformando la serie para tratar de estabilizar la variabilidad. Puesto que ésta aumenta con el nivel de la serie (quizás la desviación típica sea proporcional a la media), le aplicamos a la serie de consumo de electricidad la función logaritmo neperiano.



El gráfico de la izquierda muestra que:

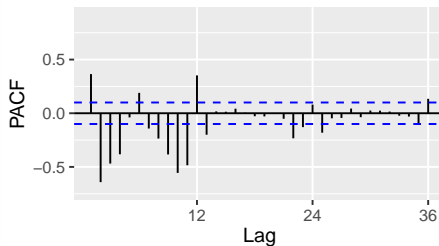
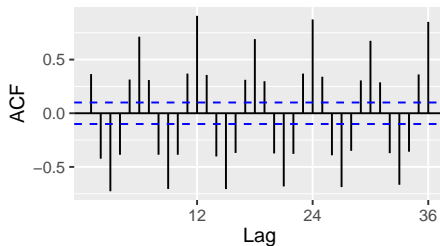
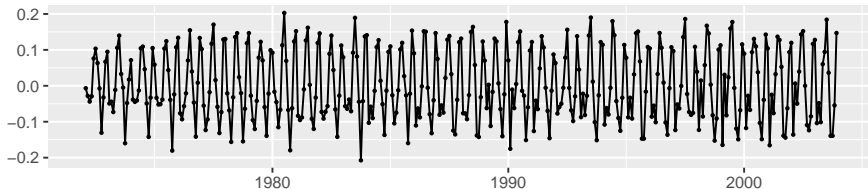
- La varianza del consumo se ha estabilizado al aplicarle la función logaritmo neperiano (log-serie es homocedástica).
- La serie transformada (log-serie) tiene **tendencia** y componente estacional.

La diferenciación regular podría eliminar la tendencia.



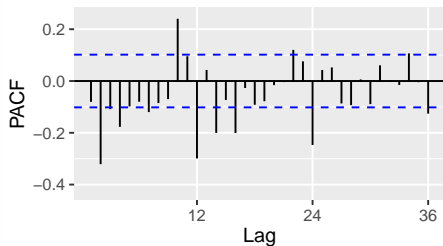
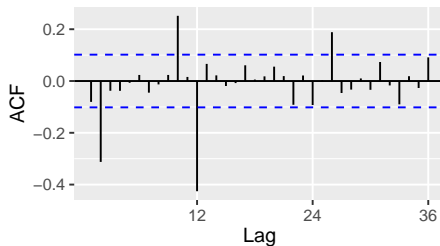
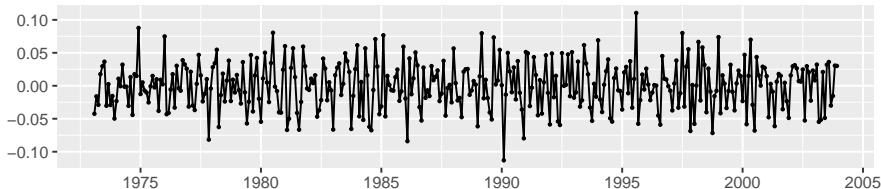
Sugerencia: log-serie es homocedástica, tiene **tendencia** y también componente estacional. Necesidad de diferenciación regular.

dif-log-serie



Sugerencia: dif-log-serie tiene **componente estacional** de periodo 12 ($s = 12$). Necesidad de diferenciación estacional.

dif-dif-log-serie



Sugerencia: dif-dif-log-serie es estacionaria y compatible con un proceso $\text{ARMA}(0,2) \times (0,1)_{12}$. Equivalent., $\text{ARIMA}(0,1,2) \times (0,1,1)_{12}$ para log-serie.

Expresión del modelo identificado (ARIMA(0,1,2) \times (0,1,1)₁₂):

$$(1 - B)(1 - B^{12}) Y_t = c + (1 + \theta_1 B + \theta_2 B^2)(1 + \Theta_1 B^{12}) a_t,$$

siendo $Y_t = \log(X_t)$ y X_t el consumo eléctrico del mes t . Equivalentemente:

$$\begin{aligned} Y_t &= Y_{t-1} + Y_{t-12} - Y_{t-13} \\ &\quad + c + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} \\ &\quad + \Theta_1 a_{t-12} + \theta_1 \Theta_1 a_{t-13} + \theta_2 \Theta_1 a_{t-14}. \end{aligned}$$

Nota: La inclusión de la constante c en el modelo es equivalente a considerar una tendencia polinómica de grado $d+D=2$. En general, a efectos de predicción, puede resultar peligroso incluir tendencias polinómicas de grado superior a 1. Por este motivo, **no incluiremos la constante** (consideramos $c = 0$).

2: Selección del modelo (criterio AICc)

El modelo $\text{ARIMA}(1,1,2) \times (0,1,1)_{12}$ resultó ser el de menor AICc ($\text{AICc} = -1645.32$). Nótese que este modelo es ligeramente distinto del previamente identificado en base a las fas y fap muestrales ($\text{ARIMA}(0,1,2) \times (0,1,1)_{12}$; $\text{AICc} = -1636.292$).

Algunos de los demás modelos evaluados alcanzaron un AICc que distó del mínimo menos de 2 unidades, pero ninguno fue más simple.

Expresión del modelo seleccionado $(ARIMA(1,1,2) \times (0,1,1)_{12})$:

$$(1 - \phi_1 B)(1 - B)(1 - B^{12}) Y_t = (1 + \theta_1 B + \theta_2 B^2)(1 + \Theta_1 B^{12}) a_t,$$

siendo $Y_t = \log(X_t)$ y X_t el consumo eléctrico del mes t .

Equivalentemente:

$$\begin{aligned} Y_t = & (1 + \phi_1) Y_{t-1} - \phi_1 Y_{t-2} + Y_{t-12} - (1 + \phi_1) Y_{t-13} \\ & + c + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} \\ & + \Theta_1 a_{t-12} + \theta_1 \Theta_1 a_{t-13} + \theta_2 \Theta_1 a_{t-14}. \end{aligned}$$

3: Estimación del modelo seleccionado $ARIMA(1,1,2) \times (0,1,1)_{12}$

Estimando los parámetros por máxima verosimilitud resulta:

$$\hat{\phi}_1 = 0.3214 (0.0901), \hat{\theta}_1 = -0.5804 (0.0898), \hat{\theta}_2 = -0.3762 (0.0784),$$

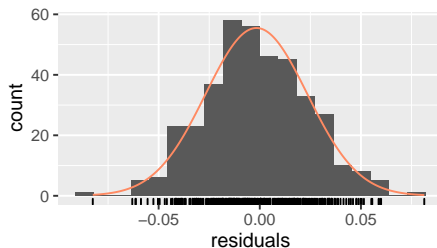
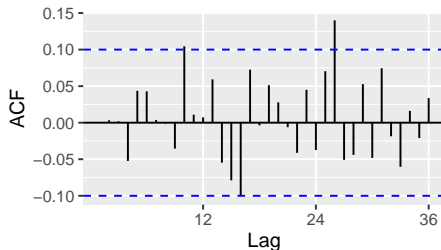
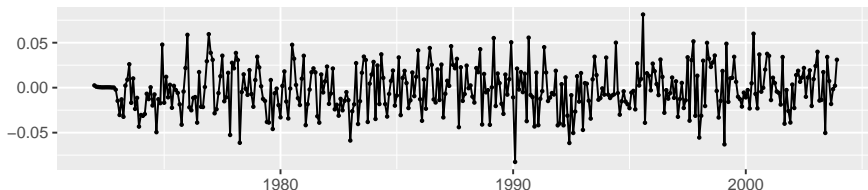
$$\hat{\Theta}_1 = -0.7135 (0.0391) \text{ y } \hat{\sigma}_a^2 = 0.0006621.$$

Nótese que las estimaciones $\hat{\phi}_1$, $\hat{\theta}_1$, $\hat{\theta}_2$ y $\hat{\Theta}_1$ son significativamente distintas de cero (al 5 %) (recuérdese que esta afirmación está basada en el hecho de que se verifica que

$$|\hat{\gamma}| \geq 1.96 \times \hat{\sigma}_{\hat{\gamma}},$$

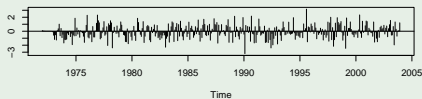
donde hemos denotado por $\hat{\gamma}$ a cualquiera de las anteriores estimaciones (salvo $\hat{\sigma}_a^2$) y por $\hat{\sigma}_{\hat{\gamma}}$ a la estimación de la desviación típica del estimador correspondiente; véase la transparencia dedicada a las propiedades de los estimadores de máxima verosimilitud).

4: Diagnosis del modelo $ARIMA(1,1,2) \times (0,1,1)_{12}$

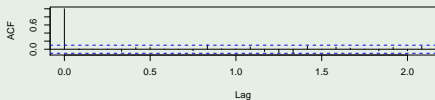
Residuals from $ARIMA(1,1,2)(0,1,1)_{12}$ 

Sólo $\hat{\rho}_{26}$ (y $\hat{\rho}_{10}$ por poco) se sale de la zona de no rechazo: no problem!

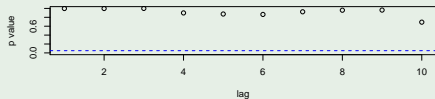
Standardized Residuals



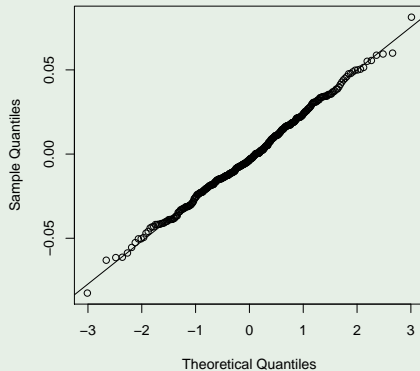
ACF of Residuals



p values for Ljung-Box statistic



Normal Q-Q Plot



Sugerencia: Independencia y posible normalidad.

Contrastes de independencia, media cero y normalidad

- Test de Ljung-Box con $H = 24$ (20 g.l.): p -valor=0.3339
- $\mu_a = 0$: p -valor=0.2012
- Normalidad:
 - Jarque-Bera: p -valor=0.7272
 - Shapiro-Wilk: p -valor= 0.7197

Conclusión: Un modelo $ARIMA(1,1,2) \times (0,1,1)_{12}$ sin constante y con innovaciones gaussianas resulta **adecuado** como generador de la serie del consumo de electricidad (transformada a través de la función logaritmo neperiano).

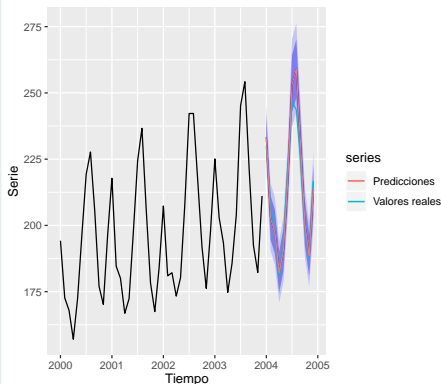
5: Predicción en base al modelo $\text{ARIMA}(1,1,2) \times (0,1,1)_{12}$

Para finalizar el estudio, el $\text{ARIMA}(1,1,2) \times (0,1,1)_{12}$ que hemos seleccionado, estimado y chequeado fue utilizado para realizar **predicciones puntuales** con origen en $T = 384$ y horizontes de predicción $k = 1, \dots, 12$.

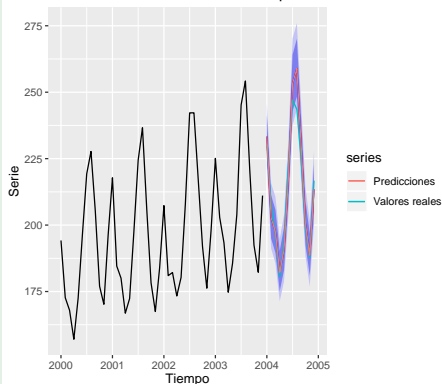
También se obtuvieron los correspondientes **intervalos de predicción** (al 80 % y al 95 %) contruidos en base a la distribución asintótica del error de predicción y también los intervalos obtenidos vía bootstrap.

Todo ello se pueden observar en los gráficos de la siguiente transparencia, en los que además se han añadido los **valores reales a predecir** (lo que permite ver la adecuación tanto de las predicciones puntuales como de los intervalos de predicción).

IP obtenidos a través de la distribución asintótica



IP obtenidos a través del bootstrap sobre los residuos



Observación: En este ejemplo, los intervalos de predicción parecen depender poco de la metodología utilizada.

```
forecast(...); autoplot(...); autolayer(...)
```

Valores numéricos

Mes	Consumo	Predicción	Intervalo de Predicción (95 %)	
			Distr. asintótica	Bootstrap
enero	229.922	233.299	(221.825 , 245.366)	(222.589 , 245.663)
febr.	207.913	202.758	(190.422 , 215.893)	(191.216 , 215.761)
mar.	195.917	197.380	(185.078 , 210.500)	(185.662 , 210.595)
abril	180.561	182.247	(170.828 , 194.430)	(171.335 , 194.754)
mayo	193.574	190.710	(178.734 , 203.488)	(178.825 , 203.767)
junio	222.073	216.954	(203.310 , 231.513)	(203.196 , 231.774)
julio	247.093	253.254	(237.306 , 270.272)	(237.630 , 270.252)
agos.	243.509	258.902	(242.580 , 276.323)	(243.499 , 275.983)
sept.	224.615	229.924	(215.411 , 245.414)	(216.094 , 245.594)
oct.	198.691	200.532	(187.859 , 214.059)	(188.360 , 214.355)
nov.	187.896	188.694	(176.755 , 201.439)	(176.882 , 201.557)
dic.	216.703	213.443	(199.923 , 227.878)	(200.257 , 227.893)

A continuación se muestran varias medidas de la adecuación de las predicciones obtenidas:

MAE	RMSE	MAPE	MASE
4.3688	5.7381	1.9539	0.7943

Como era de esperar, el MAPE y el MASE son las que resultan más informativas (pues son medidas relativas, no absolutas): ambas dan lugar a valores bajos*, por lo que podemos concluir que el modelo $\text{ARIMA}(1,1,2) \times (0,1,1)_{12}$ ajustado predice bien.

*Recuérdese que el MAPE se puede asimilar a un porcentaje, y el MASE a una proporción del error cometido por un método, el *naïve*, que predice sobre la serie histórica.

`accuracy(...)`

A lo largo de este tema:

- Se ha **construido** la clase de modelos Box-Jenkins.
- Se han propuesto métodos para **identificar** sus órdenes:
 - Basados en el estudio de sus fas y fap muestrales.
 - Basados en el estudio de las funciones AIC, AICc y BIC.
- Se han propuesto **estimadores** de sus parámetros y se han mostrado algunas de sus propiedades asintóticas.
- Se han propuesto técnicas para **chequear** el modelo ajustado.
- Se han propuesto métodos para **predecir** sus valores futuros, y se han construido intervalos de predicción.
- Se han propuesto **medidas de adecuación** de las predicciones.
- Se ha propuesto una familia de transformaciones para **estabilizar la varianza**.

Resumen (1/3): Construcción del modelo generador de la serie

- **Etapa 1:** Representar gráficamente la serie frente al tiempo, y su fas muestral frente al retardo.
 - 1 Si el gráfico de la serie sugiere presencia de variabilidad no constante, transformar (Box-Cox) la serie para estabilizar la varianza.
 - 2 Si el gráfico de la serie (quizás transformada en el paso anterior) y/o el gráfico de su fas muestral sugiere/n presencia de tendencia, aplicar diferencias regulares (**d**) hasta eliminarla.
 - 3 Si el gráfico de la serie (posiblemente transformada en alguno de los 2 pasos anteriores) y/o el gráfico de su fas muestral sugiere/n presencia de componente estacional (**s**), aplicar diferencias estacionales (**D**) hasta eliminarla.

Resumen (2/3): Construcción del modelo generador de la serie

- **Etapla 2:** Representar gráficamente la serie (posiblemente transformada en la etapa 1) frente al tiempo, y sus fas y fap muestrales frente al retardo.

Dichos gráficos debieran sugerir la procedencia de la serie (posiblemente transformada en la etapa 1) de un proceso estacionario (ARMA, quizás multiplicativo), pues en caso contrario no deberíamos haber pasado a esta etapa 2.

- 1 Tratar de identificar sus órdenes p , q , P y Q a través del estudio de su fas y fap muestrales.
- 2 Identificar sus órdenes p , q , P y Q a través del estudio de las funciones AIC, AICc y/o BIC.

Resumen (3/3): Construcción del modelo generador de la serie

Una vez que uno o varios modelos han sido identificados, la siguiente etapa es su **estimación**.

A continuación, el/los modelo/s estimado/s debe/n ser **chequeado/s** (es necesario comprobar que verifica/n las hipótesis básicas que se han supuesto en su construcción). Principalmente, comprobaremos la hipótesis de que las innovaciones son **ruido blanco** (preferiblemente **gaussiano**).

Si disponemos de varios modelos que han superado el análisis de residuos (diagnosis), seleccionaremos aquél que, teniendo un AIC, AICc y/o BIC **pequeño** (diferencias de hasta 2 unidades no se consideran relevantes), resulte más **simple**. En base a dicho modelo, realizaremos las **predicciones**.

Parte III

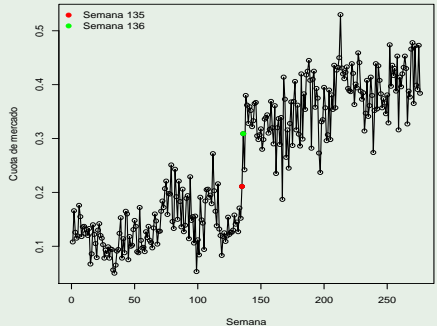
Tópicos adicionales

En los modelos estudiados hasta ahora, toda la información procedía de la propia serie de tiempo. En este capítulo incorporaremos en la modelización **información externa**.

En agosto de 1960 la Asociación Dental Americana publicó un estudio según el cual el flúor introducido en la pasta dentífrica Crest produce un efecto positivo en la reducción de la caries.

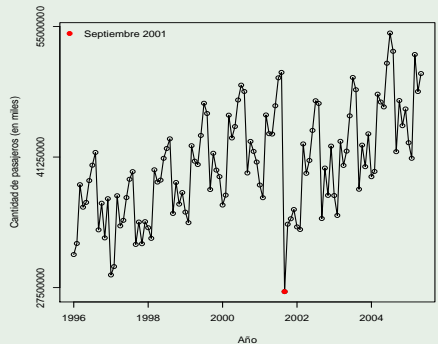
La figura de la derecha muestra la serie semanal de cuota de mercado de Crest (Enero-1958 hasta Abril-1963).

La publicación del estudio se realizó en la semana 135.



El día 11 de septiembre de 2001 EE.UU. fue víctima de un ataque terrorista: cuatro aviones comerciales fueron secuestrados y estrellados.

La figura de la derecha muestra la serie de la cantidad mensual de pasajeros en líneas aéreas en EE.UU. desde Enero-1996 hasta Mayo-2005.



El **análisis de intervención** estudia modelos de series de tiempo en los que se incluyen variables ficticias para representar sucesos (**intervenciones**) que producen efectos deterministas (por ejemplo, cambios en el nivel o tendencia).

En general, hay dos tipos de efectos provocados por la intervención:

- **Efectos permanentes:** provocan un cambio en el nivel de la serie a partir de un instante conocido.
- **Efectos transitorios:** provocan un cambio en algunos valores de la serie, pero a la larga el nivel no se ve afectado.

A lo largo de esta sección supondremos que:

- La intervención tuvo lugar en el instante $t = h$ (conocido).
- y_1, \dots, y_T denota la serie intervenida (es la que observamos), mientras que $\{Y_t\}$ es el proceso que la genera.
- x_1, \dots, x_T denota la serie si no hubiese sufrido intervención (sólo observable hasta $t = h - 1$). $\{X_t\}$, el proceso que la genera, es un ARIMA.

Para modelizar los efectos permanentes y los efectos transitorios provocados por la intervención, son de utilidad las funciones (variables ficticias) **escalón** ($S_t^{(h)}$) e **impulso** ($I_t^{(h)}$), respectivamente, que definimos a continuación:

$$S_t^{(h)} = \begin{cases} 0, & \text{si } t < h \\ 1, & \text{si } t \geq h \end{cases} \quad I_t^{(h)} = \begin{cases} 0, & \text{si } t \neq h \\ 1, & \text{si } t = h \end{cases}$$

Nota: Se verifica que $S_t^{(h)} - S_{t-1}^{(h)} = I_t^{(h)}$; también, $S_t^{(h+j)} = S_{t-j}^{(h)}$ ($j \geq 0$).

Modelos para series que sufren efectos permanentes (I)

El efecto de la intervención consiste en variar, desde el instante $t = h$, el nivel de la serie en w_0 unidades.

El modelo sería:

$$Y_t = \begin{cases} X_t, & \text{si } t < h, \\ w_0 + X_t, & \text{si } t \geq h. \end{cases}$$

Equivalentemente:

$$Y_t = w_0 S_t^{(h)} + X_t.$$

Modelos para series que sufren efectos permanentes (II)

El efecto de la intervención consiste en variar gradualmente el nivel de la serie desde el instante $t = h$ hasta el instante $t = h + m$, manteniéndose a partir de ahí las variaciones acumuladas.

El modelo sería:

$$Y_t = \begin{cases} X_t, & \text{si } t < h, \\ w_0 + X_t, & \text{si } t = h, \\ w_0 + w_1 + X_t, & \text{si } t = h + 1, \\ \vdots & \vdots \\ w_0 + \cdots + w_{m-1} + X_t, & \text{si } t = h + m - 1, \\ w_0 + \cdots + w_{m-1} + w_m + X_t, & \text{si } t \geq h + m. \end{cases}$$

Equivalentemente:

$$Y_t = w_0 S_t^{(h)} + w_1 S_t^{(h+1)} + \cdots + w_m S_t^{(h+m)} + X_t.$$

Modelos para series que sufren efectos permanentes (III)

El efecto de la intervención consiste en variar gradualmente el nivel de la serie desde el instante $t = h$ hasta el final, estando las variaciones correspondientes generadas a partir de dos parámetros, w_0 y δ ($0 < \delta < 1$).

El modelo sería:

$$Y_t = \begin{cases} X_t, & \text{si } t < h, \\ w_0 + X_t, & \text{si } t = h, \\ w_0 + w_0\delta + X_t, & \text{si } t = h + 1, \\ \vdots & \vdots \\ w_0 + w_0\delta + w_0\delta^2 + \cdots + w_0\delta^j + X_t, & \text{si } t = h + j, \\ \vdots & \vdots \end{cases}$$

Equivalentemente:

$$Y_t = w_0 \sum_{j=0}^{\infty} \delta^j S_t^{(h+j)} + X_t.$$

Modelos para series que sufren efectos transitorios (I)

El efecto de la intervención consiste en variar, exclusivamente en el instante $t = h$, el valor de la serie en w_0 unidades.

El modelo sería:

$$Y_t = \begin{cases} X_t, & \text{si } t \neq h, \\ w_0 + X_t, & \text{si } t = h. \end{cases}$$

Equivalentemente:

$$Y_t = w_0 I_t^{(h)} + X_t.$$

Modelos para series que sufren efectos transitorios (II)

El efecto de la intervención consiste en variar los valores de la serie desde el instante $t = h$ hasta el instante $t = h + m$.

El modelo sería:

$$Y_t = \begin{cases} X_t, & \text{si } t < h, \\ w_0 + X_t, & \text{si } t = h, \\ w_1 + X_t, & \text{si } t = h + 1, \\ \vdots & \vdots \\ w_m + X_t, & \text{si } t = h + m, \\ X_t, & \text{si } t > h + m. \end{cases}$$

Equivalentemente:

$$Y_t = w_0 I_t^{(h)} + w_1 I_t^{(h+1)} + \dots + w_m I_t^{(h+m)} + X_t.$$

Modelos para series que sufren efectos transitorios (III)

El efecto de la intervención consiste en variar los valores de la serie desde el instante $t = h$ hasta el final, estando las variaciones correspondientes generadas a partir de dos parámetros, w_0 y δ ($0 < \delta < 1$).

El modelo sería:

$$Y_t = \begin{cases} X_t, & \text{si } t < h, \\ w_0 + X_t, & \text{si } t = h, \\ w_0\delta + X_t, & \text{si } t = h + 1, \\ w_0\delta^2 + X_t, & \text{si } t = h + 2, \\ \vdots & \vdots \\ w_0\delta^j + X_t, & \text{si } t = h + j, \\ \vdots & \vdots \end{cases}$$

Equivalentemente:

$$Y_t = w_0 \sum_{j=0}^{\infty} \delta^j I_t^{(h+j)} + X_t.$$

Tanto los modelos para efectos permanentes como para efectos transitorios que acabamos de presentar pueden escribirse como

$$Y_t = w(B)S_t^{(h)} + X_t \quad \text{e} \quad Y_t = w(B)I_t^{(h)} + X_t,$$

respectivamente, donde:

- (I) $w(B) = w_0$.
- (II) $w(B) = w_0 + w_1B + \dots + w_mB^m$.
- (III) $w(B) = w_0(1 + \delta B + \delta^2 B^2 + \dots + \delta^j B^j + \dots)$. Si además $0 < \delta < 1$, se verifica que $w(B) = w_0/(1 - \delta B)$.

La función $w(B)$ se denomina **función de transferencia**, y describe el efecto que sobre la variable respuesta, Y_t , ejerce la función escalón $S_t^{(h)}$ (o la función impulso $I_t^{(h)}$). Por lo tanto, la componente $w(B)S_t^{(h)}$ (o $w(B)I_t^{(h)}$) describe el impacto de la intervención.

A partir de los 6 modelos de intervención presentados se pueden construir e interpretar modelos más sofisticados.

Así, el modelo

$$Y_t = (u_0 + \frac{v_0}{1 - \delta B})I_t^{(h)} + X_t$$

modelizaría una serie que sufre un efecto transitorio cuyo resultado es:

- En el instante de la intervención: Aumentar el valor de la serie en $u_0 + v_0$ unidades.
- j instantes después: Aumentar el valor de la serie en $v_0\delta^j$ unidades.

También, sin más que cambiar h por $h + r$ en los modelos

$$Y_t = w(B)S_t^{(h)} + X_t \text{ e } Y_t = w(B)I_t^{(h)} + X_t,$$

se consigue **retardar r instantes de tiempo el efecto de la intervención**. Así, el modelo $Y_t = (w_0 + w_1 B)S_t^{(h+r)} + X_t$ o, equivalentemente,

$$Y_t = (w_0 B^r + w_1 B^{r+1})S_t^{(h)} + X_t,$$

modelizaría un efecto permanente cuyo resultado es:

- r instantes después de la intervención: Aumentar el nivel de la serie en w_0 unidades.
- $r + 1$ instantes después de la intervención: Un aumento adicional con respecto al del instante r de w_1 unidades.

Para **construir** un modelo de intervención es necesario:

- Proponer el tipo de efecto (permanente o transitorio).
- Proponer una función de transferencia ($w(B)$).
- Proponer un modelo ARIMA para $\{X_t\}$.
- Estimar los parámetros del ARIMA y de $w(B)$.

Naturalmente, todo ello debe hacerse en base a las observaciones disponibles; esto es, a partir de la serie intervenida: y_1, \dots, y_T . **¿Cómo?**

- **Tipo de efecto y función de transferencia.** Se proponen en base a la exploración de la serie completa: y_1, \dots, y_T .
- **Modelo ARIMA** para $\{X_t\}$. Podemos utilizar 2 opciones:
 - (a) Se propone en base al estudio de las observaciones anteriores a la intervención: y_1, \dots, y_{h-1} (nótese que, $\forall t < h, y_t = x_t$).
 - (b) Se propone el ARIMA que dé lugar al menor valor de un criterio de información (AIC, AICc, BIC) en la estimación del modelo para $\{Y_t\}$ (en este caso, se utiliza la serie completa: y_1, \dots, y_T).
- **Estimación** del ARIMA y de $w(B)$. Máxima verosimilitud.

La estimación (máxima verosimilitud) de los parámetros del ARIMA y de $w(B)$ podría llevarse a cabo a través del siguiente algoritmo iterativo.

Para ilustrarlo, consideramos que el efecto de la intervención fue transitorio, representamos $X_t = \psi(B)a_t$ y utilizaremos que:

$$Y_t = w(B)I_t^{(h)} + \psi(B)a_t$$

y

$$\psi(B)^{-1}Y_t = w(B)\psi(B)^{-1}I_t^{(h)} + a_t.$$

El algoritmo consiste en:

- ➊ Obtener una estimación inicial, $\hat{w}(B)$, de $w(B)$.
- ➋ Ajustar un ARIMA a $Y_t - \hat{w}(B)I_t^{(h)}$ ($\Rightarrow \hat{\psi}(B)$).
- ➌ Estimar de nuevo $w(B)$ en base al modelo

$$\hat{\psi}(B)^{-1}Y_t = w(B)\hat{\psi}(B)^{-1}I_t^{(h)} + a_t$$

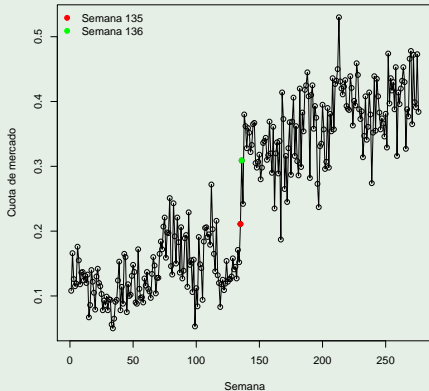
- ➍ Iterar los pasos 2-3 hasta obtener convergencia.

Aplicación a datos reales

Presentamos a continuación un ejemplo con datos reales en el que se hace uso de gran parte de lo expuesto hasta este momento.

La serie de tiempo que analizaremos se corresponde con la [cuota semanal de mercado de la pasta dentífrica Crest](#) (enero 1958 - abril 1963).

En la semana número 135 se publicó un estudio favorable a dicho dentífrico.



El gráfico de la izquierda sugiere que la intervención ha ejercido un **efecto permanente** sobre la serie.

Tal efecto podría haber dado lugar a dos saltos: uno en el mismo instante de la intervención (semana 135) y otro, acumulativo, en el instante siguiente (semana 136). A partir de ahí, el nivel de la serie ha cambiado.

Construimos un modelo que contemple tal situación.

Asumimos que el efecto de la intervención es **permanente** y consiste en aumentar gradualmente el nivel de la serie desde el instante de la intervención hasta el instante siguiente (manteniéndose el efecto acumulativo en los instantes posteriores). Concretamente:

$$Y_t = \begin{cases} X_t, & \text{si } t < 135, \\ w_0 + X_t, & \text{si } t = 135, \\ w_0 + w_1 + X_t, & \text{si } t \geq 136, \end{cases}$$

o, equivalentemente:

$$Y_t = w_0 S_t^{(135)} + w_1 S_t^{(136)} + X_t.$$

Escrito de forma compacta:

$$Y_t = w(B) S_t^{(135)} + X_t,$$

donde $w(B) = w_0 + w_1 B$.

Se propuso, para $\{X_t\}$, el modelo ARIMA que da lugar al menor valor del AICc en la estimación del modelo para la serie intervenida (y_1, \dots, y_T) . Tal propuesta fue un ARIMA(0,1,1) sin constante para la serie x_1, \dots, x_T (nótese que de esta serie sólo se observan los valores x_1, \dots, x_{h-1} , pues los demás están afectados por la intervención).

Por lo tanto, el modelo completo a estimar es:

$$Y_t = w_0 S_t^{(135)} + w_1 S_t^{(136)} + X_{t-1} + \theta_1 a_{t-1} + a_t.$$

La estimación de w_0 resultó no significativamente distinta de cero ($\hat{w}_0 = 0.0654$ (0.0434)). Por tanto, w_0 fue eliminado, y se construyó y ajustó un nuevo modelo.

```
auto.arima(..., xreg=..., ...)
```

El nuevo modelo es:

$$Y_t = w_1 S_t^{(136)} + X_{t-1} + \theta_1 a_{t-1} + a_t.$$

Nótese que, según este modelo, el efecto de la intervención se retrasa una semana.

Las estimaciones obtenidas fueron:

$$\hat{w}_1 = 0.1624 \text{ (0.0283)}, \hat{\theta}_1 = -0.7758 \text{ (0.0442)} \text{ y } \hat{\sigma}_a^2 = 0.0019.$$

Todos sus coeficientes son significativamente distintos de cero, y los residuos superaron los distintos contrastes (incorrelación, media nula y normalidad).

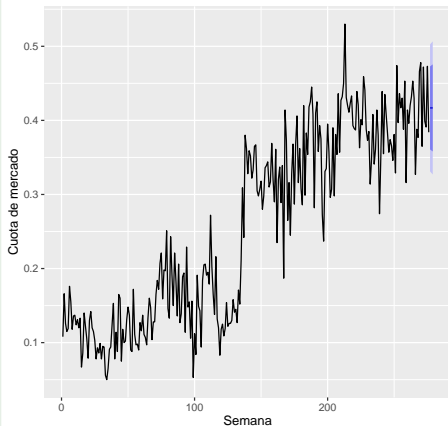
Interpretación: El efecto de la intervención consistió en aumentar la cuota de mercado de la pasta dentífrica Crest, a partir de la semana siguiente al anuncio del estudio, en 0.1624 unidades.

El modelo que hemos seleccionado, estimado y chequeado fue utilizado para realizar predicciones a horizontes de predicción $h = 1, 2, 3$.

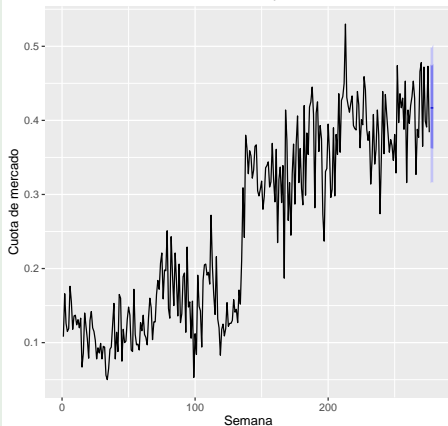
También se obtuvieron los correspondientes intervalos de predicción (al 80 % y al 95 %), tanto a través de la distribución asintótica del error de predicción (pues las innovaciones son gaussianas) como a través de bootstrap sobre los residuos.

Todo ello, junto con la serie histórica, se muestra en los gráficos de las dos páginas siguientes (en la segunda página, se ha aplicado un “zoom” para ver mejor los intervalos de predicción).

IP obtenidos a través de la distribución asintótica

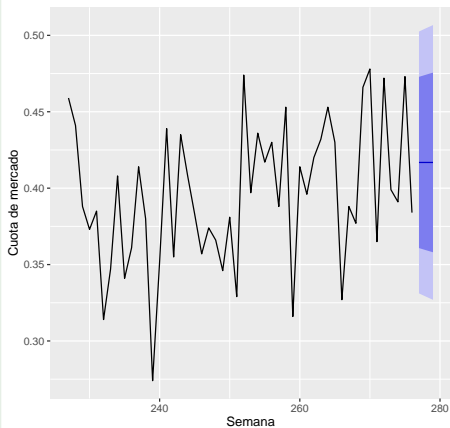


IP obtenidos a través del bootstrap sobre los residuos

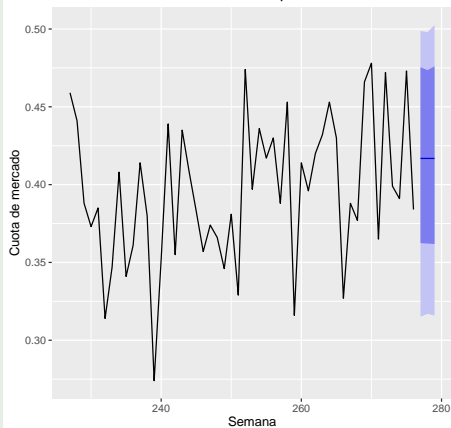


```
forecast(..., xreg=..., ...)
```

IP obtenidos a través de la distribución asintótica

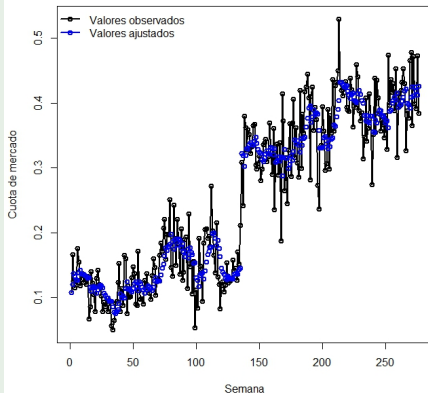


IP obtenidos a través del bootstrap sobre los residuos



El gráfico de la derecha muestra la serie original (negro) y los valores ajustados a través del modelo propuesto (azul).

Como se puede observar, el modelo ajustado capta bien el cambio provocado por la intervención.



Un **valor atípico** (“outlier”) es un valor que tiene “poca probabilidad de ocurrir” dada la estructura habitual de evolución de la serie.

Consideraremos dos tipos de valores atípicos:

- Atípicos aditivos (AO). El valor **de la serie** en el instante $t = h$ ha sido generado de manera distinta al resto.
- Atípicos innovativos (IO). El valor **de la innovación** en el instante $t = h$ ha sido generado de manera distinta al resto.

En la práctica, a efectos de mejorar tanto la precisión de las estimaciones de los parámetros del modelo como las predicciones obtenidas a partir de él, resulta muy importante detectar la presencia de valores atípicos y separarlos de la dinámica habitual de la serie.

A efectos del desarrollo de este tópico adicional, supondremos que hay un único valor atípico y que:

- El valor atípico tuvo lugar en el instante $t = h$ (desconocido).
- y_1, \dots, y_T denota la serie que contiene al valor atípico.
- Si no hubiese valor atípico, la serie podría ser modelizada a través del proceso ARIMA $\{X_t\}$.

Por lo tanto, las siguientes expresiones modelizan la presencia de un atípico aditivo y de un atípico innovativo, respectivamente:

$$\text{AO: } Y_t = w_A I_t^{(h)} + \psi(B)a_t; \quad \text{IO: } Y_t = \psi(B)(w_I I_t^{(h)} + a_t)$$

donde hemos considerado que $X_t = \psi(B)a_t$.

Nota: La serie observada es y_1, \dots, y_T (la afectada por el atípico).

En la práctica, los valores atípicos serán detectados en base al estudio de los residuos. Por lo tanto, resulta fundamental conocer qué efecto ejercen sobre ellos.

Para ello, relacionaremos las verdaderas innovaciones del ARIMA:

$$a_t = \psi(B)^{-1}X_t = X_t - \pi_1 X_{t-1} - \pi_2 X_{t-2} - \dots$$

con las que contienen el efecto del atípico:

$$e_t = \psi(B)^{-1}Y_t = Y_t - \pi_1 Y_{t-1} - \pi_2 Y_{t-2} - \dots$$

Efecto de los atípicos aditivos sobre los residuos

Supongamos que estamos ante un atípico aditivo. Entonces

$$Y_t = w_A l_t^{(h)} + \psi(B)a_t.$$

Puede demostrarse que

$$e_t = -w_A \pi_{t-h} + a_t,$$

donde $\pi_0 = -1$ y $\pi_j = 0 \ \forall j < 0$. Esto es,

$$e_t = \begin{cases} a_t, & \text{si } t < h, \\ w_A + a_t, & \text{si } t = h, \\ -w_A \pi_i + a_{h+i}, & \text{si } t = h + i \text{ con } i > 0. \end{cases}$$

Efecto de los atípicos innovativos sobre los residuos

Supongamos que estamos ante un atípico innovativo. Entonces

$$Y_t = \psi(B)(w_I l_t^{(h)} + a_t).$$

Puede demostrarse que

$$e_t = w_I l_t^{(h)} + a_t.$$

Esto es,

$$e_t = \begin{cases} a_t, & \text{si } t \neq h, \\ w_I + a_t, & \text{si } t = h. \end{cases}$$

En ambos casos (AO e IO) se tiene que

$$e_t = w z_t + a_t,$$

donde w denota el tamaño del atípico, y $\{z_t\}$ son valores que dependen del tipo de atípico y del modelo que sigue la serie sin contaminar.

La **detección de valores atípicos** consiste en estimar w en el modelo

$$\hat{e}_t = w \hat{z}_t + a_t$$

y estudiar si resulta o no significativamente distinta de cero.

Nota: $\{\hat{e}_t\}$ y $\{\hat{z}_t\}$ se construyen utilizando los valores $\{\hat{\pi}_i\}$ correspondientes al ARIMA ajustado a la serie y_1, \dots, y_T .

En la práctica, acerca de los valores atípicos se desconoce:

- Cuántos hay.
- Los instantes (h) en que surgen.
- El tipo de valor atípico (AO o IO).

Por ello, el test de valor atípico debe ser aplicado a todas las observaciones.

Detección de valores atípicos

A continuación mostramos un procedimiento conservador, basado en el método de Bonferroni, para detectar valores atípicos:

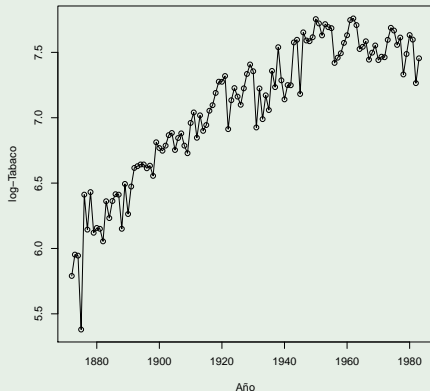
- 1 Candidato a AO: entre todos aquéllos cuyo p-valor en el contraste de AO sea inferior a $0.05/T$, aquél cuyo p-valor sea menor.
- 2 Candidato a IO: entre todos aquéllos cuyo p-valor en el contraste de IO sea inferior a $0.05/T$, aquél cuyo p-valor sea menor.
- 3 De los candidatos anteriores consideraremos como valor atípico aquél cuyo p-valor sea menor.
- 4 El efecto del atípico detectado en 1-3 se incorpora en el modelo, y se repiten los pasos 1-4 (aplicados ahora a los residuos del nuevo modelo) hasta que no haya atípicos.

Aplicación a datos reales

Modelizaremos la **producción anual de tabaco** en EE.UU. (1872–1983) y realizaremos predicciones de futuros valores.

Utilizaremos dos modelos: uno que no contemple atípicos y otro que sí (el primero ya fue analizado en el Tema 2).

La serie, en unidades logarítmicas, puede verse en el gráfico de la derecha.



Como se puede comprobar en el Tema 2, tanto las fas y fap muestrales como el criterio AICc propusieron un modelo ARIMA(0,1,1); esto es:

$$X_t = c + X_{t-1} + \theta_1 a_{t-1} + a_t,$$

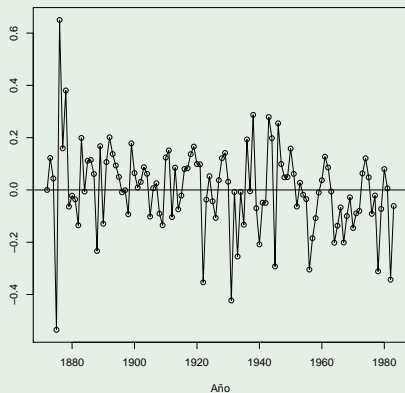
donde X_t denota al logaritmo neperiano (\log) de la producción de tabaco.

Las estimaciones obtenidas fueron:

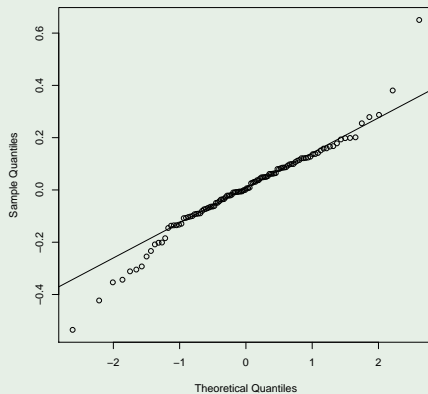
$$\hat{c} = 0.0145 \text{ (0.0049)}, \hat{\theta}_1 = -0.6899 \text{ (0.0698)} \text{ y } \hat{\sigma}_a^2 = 0.02672.$$

Los residuos del modelo ajustado superaron los contrastes de incorrelación y media nula, aunque NO los de normalidad (ni al 5 % ni al 1 %).

Residuos del modelo ARIMA



Residuos del modelo ARIMA



Ambos gráficos sugieren la presencia de valores atípicos.

A través de un análisis del ajuste anterior en busca de valores atípicos se localizó la presencia de un AO (posición 5) y de un IO (posición 4). Sin embargo, en ambos casos modelizamos sus efectos (paso a paso) como si los dos fuesen AOs debido a las carencias de la función Arima. El modelo resultante es:

$$Y_t = w_4 I_t^{(4)} + w_5 I_t^{(5)} + c + X_{t-1} + \theta_1 a_{t-1} + a_t,$$

donde el log de la producción de tabaco es ahora denotado por Y_t . Las estimaciones obtenidas fueron:

$$\hat{w}_4 = -0.6572 \text{ (0.1293)}, \hat{w}_5 = 0.3247 \text{ (0.1289)},$$

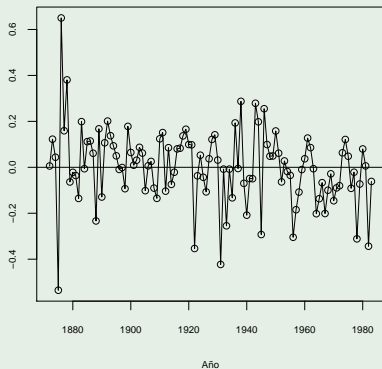
$$\hat{c} = 0.0141 \text{ (0.0052)}, \hat{\theta}_1 = -0.6159 \text{ (0.0866)} \text{ y } \hat{\sigma}_a^2 = 0.02036.$$

Los residuos del modelo ajustado superaron los contrastes de incorrelación y media nula, aunque NO los de normalidad (al 5 %, pero SÍ al 1 %).

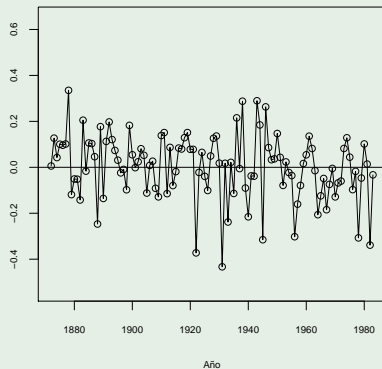
```
TSA::detectAO(...), TSA::detectIO(...), Arima(..., xreg=..., ...)
```

Comparación de los residuos de cada uno de los dos ajustes

Residuos del modelo ARIMA

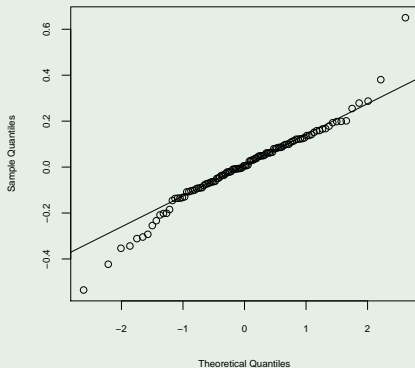


Residuos del modelo ARIMA+ATÍPICOS

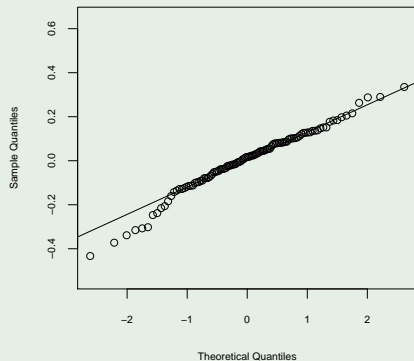


Se observa que al modelizar los atípicos desaparece su efecto sobre los residuos.

Residuos del modelo ARIMA

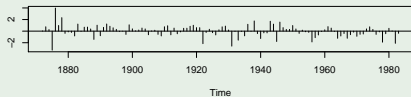


Residuos del modelo ARIMA+ATÍPICOS

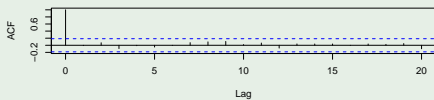


Residuos del modelo ARIMA

Standardized Residuals



ACF of Residuals

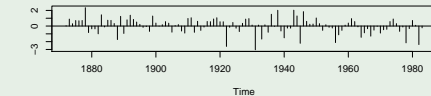


p values for Ljung-Box statistic

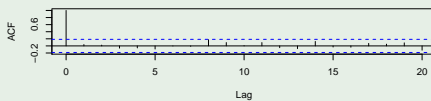


Residuos del modelo ARIMA+ATÍPICOS

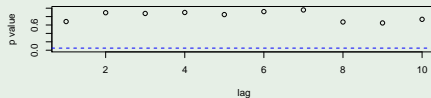
Standardized Residuals



ACF of Residuals



p values for Ljung-Box statistic



Residuos del modelo ARIMA

- Ljung-Box ($H = 10$, g.l.=8):
 p -valor=0.7314
- $\mu_a = 0$: p -valor=0.9118
- Normalidad:
 - Jarque-Bera: p -valor=0.0000
 - Shapiro-Wilk: p -valor= 0.0019

Residuos del modelo ARIMA+ATÍPICOS

- Ljung-Box ($H = 10$, g.l.=6):
 p -valor=0.3304
- $\mu_a = 0$: p -valor=0.8948
- Normalidad:
 - Jarque-Bera: p -valor=0.0216
 - Shapiro-Wilk: p -valor= 0.0379

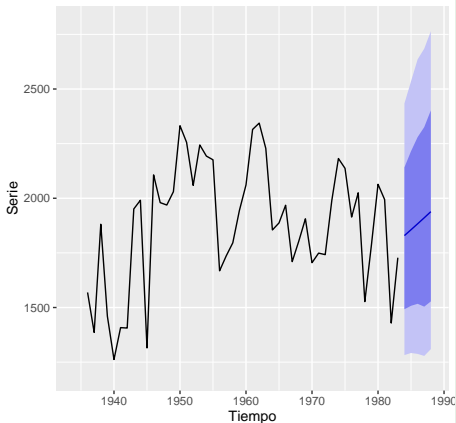
Modelo ARIMA

- AIC: -82.43
- AICc: -82.21
- BIC: -74.31

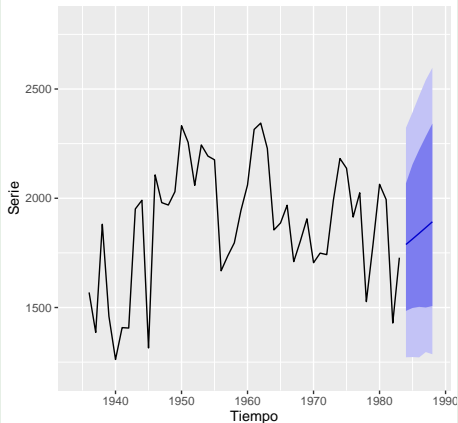
Modelo ARIMA+ATÍPICOS

- AIC: -110.86
- AICc: -110.29
- BIC: -97.31

Predicciones a partir del modelo ARIMA



Predicciones a partir del modelo ARIMA+ATÍPICOS



```
forecast(...)
```

```
forecast(..., xreg=..., ...)
```

Los modelos ARIMA explican el comportamiento de una serie de tiempo a partir de su propio pasado. Así, en el AR(1):

$$X_t = c + \phi_1 X_{t-1} + a_t.$$

Los modelos de intervención incluyen en la relación información externa relativa a la ocurrencia de un suceso (intervención) que puede afectar al comportamiento de la serie. Por ejemplo:

$$Y_t = wS_t^{(h)} + X_t,$$

donde

$$S_t^{(h)} = \begin{cases} 0, & \text{si } t < h, \\ 1, & \text{si } t \geq h. \end{cases}$$

y $\{X_t\}$ sigue un proceso ARIMA.

Hay situaciones en las que una serie de tiempo (generada por el proceso $\{Y_t\}$) está relacionada con (o es guiada por) otra serie de tiempo (generada por el proceso $\{X_t\}$), e interesa explicar el comportamiento de una a partir del de la otra. Por ejemplo:

$$Y_t = \beta_0 + \beta_1 X_{t-r} + e_t.$$

El modelo anterior es un modelo de **regresión lineal entre series de tiempo** (y, en particular, un modelo de regresión dinámica).

Su construcción y estudio serán los objetivos de esta parte final del curso.

Ante cualquier modelo de regresión, se hace necesario disponer de técnicas que nos asesoren acerca de la existencia, o no, de la relación indicada por el modelo.

Además, en el caso de los modelos de regresión con series de tiempo (datos dinámicos),

$$Y_t = \beta_0 + \beta_1 X_{t-r} + e_t,$$

hay tres particularidades que, de manera natural, están ausentes en la regresión clásica (datos estáticos):

- ❶ Desconocimiento del retardo, r .
- ❷ Posibilidad de que los errores $\{e_t\}$ estén correlados.
- ❸ Posibilidad de que los procesos $\{X_t\}$ e $\{Y_t\}$ no sean estacionarios.

A continuación, abordaremos todo esto.

Algunas definiciones

Comenzamos con varias definiciones que resultan de utilidad para asesorarnos tanto acerca de la existencia, o no, de relación lineal entre dos series de tiempo como sobre el valor de r .

- Función de covarianzas cruzadas:

$$\gamma_{s,t}(X, Y) = \text{Cov}(X_s, Y_t).$$

- Función de correlaciones cruzadas:

$$\rho_{s,t}(X, Y) = \frac{\gamma_{s,t}(X, Y)}{\sigma_{X_s} \sigma_{Y_t}}.$$

Ambas funciones miden el grado de dependencia lineal existente entre X_s e Y_t . La segunda, $\rho_{s,t}$, carece de unidades y toma valores en el intervalo $[-1, 1]$.

Los procesos $\{X_t\}$ e $\{Y_t\}$ se dice que son **conjuntamente estacionarios** si:

- 1 Ambos son estacionarios.
- 2 Las covarianzas cruzadas sólo dependen del retardo entre las variables, y no de los instantes iniciales considerados; esto es,

$$\gamma_{t,t-k}(X, Y) = \gamma_{s,s-k}(X, Y), \forall t, s, k.$$

En tal caso, utilizaremos la notación $\rho_k(X, Y) = \rho_{t,t-k}(X, Y)$.

Un caso particular

Supongamos que los procesos $\{X_t\}$ y $\{e_t\}$ son ruido blanco e independientes entre sí, y consideremos el proceso $\{Y_t\}$:

$$Y_t = \beta_0 + \beta_1 X_{t-r} + e_t.$$

Se tiene entonces que el único valor no nulo de $\rho_k(X, Y)$ ($k = \dots, -1, 0, 1, \dots$) se alcanza en $k = -r$.

Por lo tanto, en el caso particular que nos ocupa, $\rho_k(X, Y)$ nos asesora acerca del valor de r .

Sin embargo, puesto que $\rho_k(X, Y)$ es desconocida, debemos basarnos en su versión muestral, $\hat{\rho}_k(X, Y)$.

La versión muestral de $\rho_k(X, Y)$ es:

$$\hat{\rho}_k(X, Y) = \frac{\sum (x_t - \bar{x})(y_{t-k} - \bar{y})}{\sqrt{\sum (x_t - \bar{x})^2} \sqrt{\sum (y_t - \bar{y})^2}}.$$

Si $\{X_t\}$ e $\{Y_t\}$ son ruido blanco (salvo constantes) e independientes entre sí (esto es, $\beta_1 = 0$), se tiene que

$$\hat{\rho}_k(X, Y) \approx N\left(0, 1/\sqrt{T}\right).$$

Por lo tanto, se pueden utilizar los intervalos

$$\left(\frac{-1.96}{\sqrt{T}}, \frac{1.96}{\sqrt{T}}\right)$$

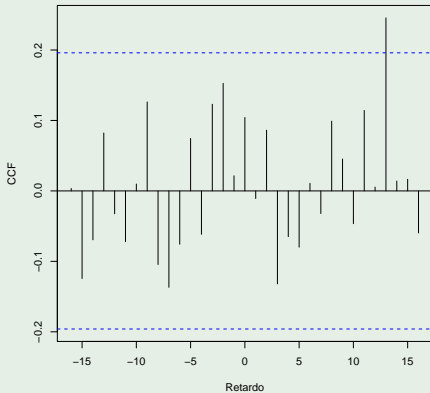
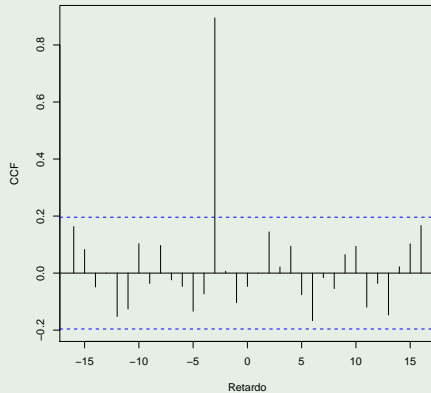
para detectar relaciones lineales entre $\{Y_t\}$ y $\{X_t\}$ y también para sugerir valores para r : habrá relación lineal si, para algún k , se verifica

$$|\hat{\rho}_{-k}(X, Y)| \geq 1.96/\sqrt{T}.$$

En dicho caso, sugeriríamos esos k como los candidatos a ser el valor del retardo r en el modelo de regresión $Y_t = \beta_0 + \beta_1 X_{t-r} + e_t$.

Ejemplos

Ruidos blancos independientes

Ruidos blancos relacionados: $r=3$ 

`ccf(· , ·)`

Aunque en lo expuesto en la antepenúltima transparencia se asume que $\{X_t\}$ e $\{Y_t\}$ son ruido blanco, se puede demostrar que los intervalos allí mostrados,

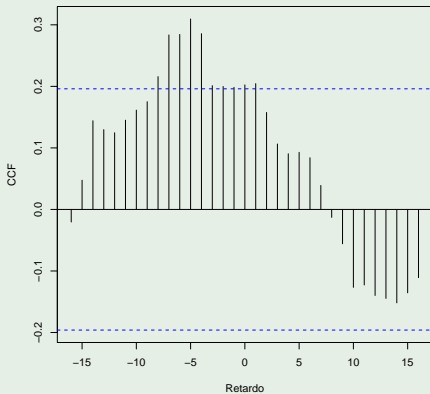
$$\left(\frac{-1.96}{\sqrt{T}}, \frac{1.96}{\sqrt{T}} \right),$$

siguen teniendo validez si únicamente uno de los dos procesos, $\{X_t\}$ o $\{Y_t\}$, es ruido blanco y el otro es estacionario.

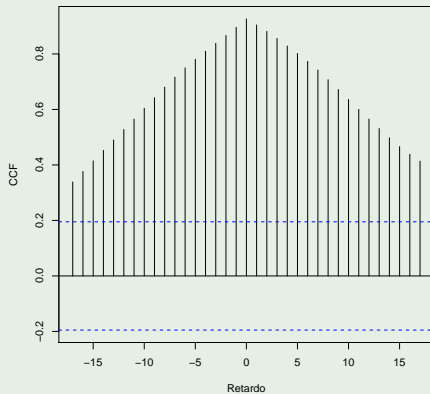
Sin embargo, la utilización de tales intervalos en situaciones en que ninguno de los dos procesos es ruido blanco (o, siendo ruido blanco uno de ellos, el otro no es estacionario) puede provocar la propuesta de falsas relaciones lineales (correlación espuria).

Ejemplos

AR(1)'s independientes



ARI(1,1)'s independientes



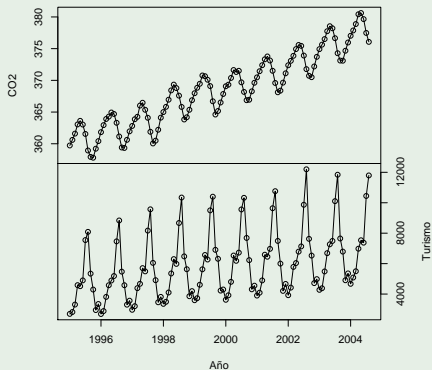
Las bandas dibujadas en los dos gráficos mostrados en la transparencia anterior están situadas a alturas $\pm 1.96/\sqrt{T}$.

Si utilizamos tales bandas para juzgar si las correlaciones cruzadas son o no significativamente distintas de cero, concluiremos con que muchas de ellas lo son.

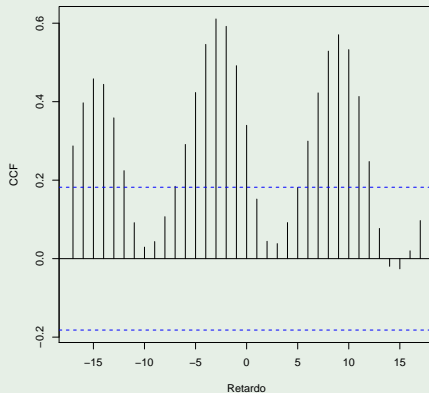
Estaríamos entonces detectando relación lineal entre los procesos involucrados; sin embargo, en ambos casos los procesos (simulados) eran independientes.

Ejemplo

CO2 (Hawái) y Turismo (España)



Correlaciones cruzadas: CO2 (Hawái) y Turismo (España)



En los gráficos de la transparencia anterior se muestran las series mensuales del nivel de CO₂ medido en el Observatorio de Manua Loa (Hawái) y de la cantidad de turistas recibida por España, y sus correlaciones cruzadas.

Al igual que en los dos ejemplos simulados anteriores (AR(1)'s y ARI(1,1)'s), las bandas dibujadas sobre el gráfico de correlaciones cruzadas están situadas a alturas $\pm 1.96/\sqrt{T}$.

Si utilizamos tales bandas para detectar relación lineal entre ambas series, concluiremos con que efectivamente la hay.

Pero... ¿alguien se cree que exista relación entre tales series? ¿Es justificable que un cambio en el turismo en España provoque un cambio en el nivel de CO₂ en Hawái (o viceversa)?

¿Qué es lo que está ocurriendo? ¿Cuál es el motivo por el cual, siendo algunas correlaciones cruzadas “tan altas”, no haya relación lineal entre los procesos?

Lo que ocurre es que, si ninguno de los procesos es ruido blanco (o, siendo ruido blanco uno de ellos, el otro no es estacionario), la variabilidad (desviación típica) de los estimadores de las correlaciones cruzadas puede ser muy superior a $1/\sqrt{T}$.

Por lo tanto, esas correlaciones cruzadas que decíamos que eran “tan altas”, en realidad no lo son (su magnitud debería ser comparada con su variabilidad, que en este caso no es $1/\sqrt{T}$).

El comparar la magnitud de las correlaciones cruzadas con $1/\sqrt{T}$ en lugar de con su verdadera variabilidad puede dar lugar a la detección de falsas relaciones lineales (correlación espuria).

Por ejemplo, en el caso particular en que ambos procesos sean AR(1) e independientes entre sí, se tiene que

$$\hat{\rho}_k(X, Y) \approx N \left(0, \sqrt{\frac{1 + \phi_x \phi_y}{T(1 - \phi_x \phi_y)}} \right).$$

Por lo tanto, en este caso las bandas a utilizar debieran ser

$$\left(-1.96 \sqrt{\frac{1 + \phi_x \phi_y}{T(1 - \phi_x \phi_y)}}, 1.96 \sqrt{\frac{1 + \phi_x \phi_y}{T(1 - \phi_x \phi_y)}} \right).$$

Nótese que, dependiendo de los valores de ϕ_x y ϕ_y , la amplitud de tales bandas puede ser mucho mayor que la de las bandas correspondientes a procesos de ruido blanco.

Proponemos seguir los siguientes pasos para detectar relaciones lineales entre dos procesos $\{X_t\}$ e $\{Y_t\}$:

- ➊ Transformar el proceso $\{X_t\}$ en un proceso de ruido blanco $\{\tilde{X}_t\}$:
 $\tilde{X}_t = \pi(B)X_t$.
- ➋ Aplicar la misma transformación al proceso $\{Y_t\}$: $\tilde{Y}_t = \pi(B)Y_t$.

Asumiendo que $\{\tilde{Y}_t\}$ es estacionario:

- ➌ Proponer relación lineal entre $\{X_t\}$ e $\{Y_t\}$ si, para algún k , se verifica:

$$|\hat{\rho}_{-k}(\tilde{X}, \tilde{Y})| \geq 1.96/\sqrt{T}.$$

En tal caso, sugeriríamos k como valor para el retardo r en el modelo:

$$Y_t = \beta_0 + \beta_1 X_{t-r} + e_t.$$

Nota: Puesto que $\pi(B)$ es un operador lineal, cualquier relación lineal entre las series originales $\{X_t\}$ e $\{Y_t\}$ se mantendrá entre $\{\tilde{X}_t\}$ e $\{\tilde{Y}_t\}$.

El proceso de transformar $\{X_t\}$ en $\{\tilde{X}_t\}$ a través de la aplicación del filtro $\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots$ se conoce como blanqueado o preblanqueado.

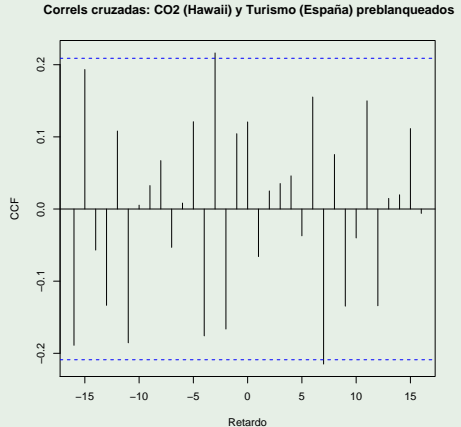
En la práctica, a efectos de detectar la existencia de relación lineal entre los procesos $\{X_t\}$ e $\{Y_t\}$ y, en caso de que la hubiese, proponer el valor de r , si $\{X_t\}$ e $\{Y_t\}$ no fuesen estacionarios es usual transformarlos (diferenciarlos) para que lo sean y aplicar un preblanqueo sobre los procesos transformados basado en la expresión AR (del proceso $\{X_t\}$ transformado).

```
TSA::prewhiten(· , ·)
```

Ejemplo

En el gráfico de la derecha se muestran las correlaciones cruzadas de las series **preblanquedas** procedentes del nivel de CO₂ en Hawaii y de la cantidad de turismo en España (preblanqueo aplicado a las series originales diferenciadas regular y estacionalmente).

El gráfico sugiere ausencia de relación lineal.



Una vez detectada la existencia de relación lineal y sugerido el retardo, r , se procede a la propuesta de un modelo para los errores, $\{e_t\}$, del modelo de regresión:

$$Y_t = \beta_0 + \beta_1 X_{t-r} + e_t.$$

Nótese que, puesto que estamos tratando con series de tiempo, es natural esperar que los errores $\{e_t\}$ estén correlados (por lo que debemos modelizarlos).

Incluso, puesto que los procesos $\{X_t\}$ e $\{Y_t\}$ pueden no ser estacionarios, es posible que los errores tampoco lo sean.

Ambas posibilidades deben ser tenidas en cuenta en la modelización y posterior estimación.

Los pasos a seguir son:

- ❶ Sugerir un ARMA para los errores ($\{e_t\}$) del modelo de regresión y estimar todos los parámetros, tanto los de la función de regresión (β_0 y β_1) como los del ARMA.
- ❷ Analizar los residuos $\{\hat{a}_t\}$:
 - Si superan el análisis, finalizar.
 - Si no lo superan pero son estacionarios, sugerir otro ARMA. Ir al paso 1.
 - Si no son estacionarios, diferenciar los procesos y construir con ellos el modelo. Ir al paso 1.

Podemos utilizar 2 opciones para sugerir el modelo para los errores $\{e_t\}$:

- (a) Obtener los residuos $\{\hat{e}_t\}$ sin tener en cuenta la posible correlación de los errores (esto es, estimar la regresión bajo el (posiblemente falso) supuesto de que los errores están incorrelados), y sugerir un modelo ARMA apropiado para dichos residuos.
- (b) Sugerir el modelo ARMA que dé lugar al menor valor de un criterio de información (AIC, AICc, BIC) en la estimación del modelo de regresión completo (regresión + ARMA).

Aplicación a datos reales

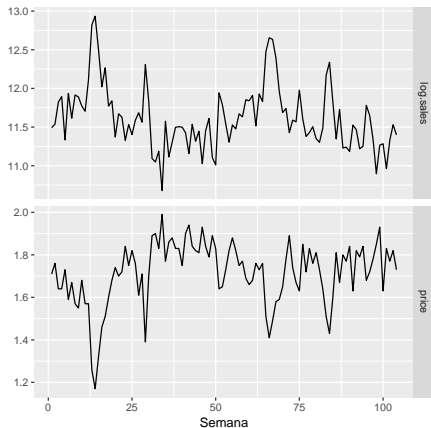
A continuación ilustramos la regresión con series de tiempo a través de un ejemplo con datos reales.

Las series de tiempo que utilizaremos se corresponden con la **venta** (en unidades logarítmicas) **y el precio semanales de las patatas fritas de la marca “Bluebird”** (Nueva Zelanda) observados en el período 20/09/1998 - 10/09/2000.

El objetivo es la construcción de un modelo de regresión lineal que explique el logaritmo de las ventas (variable respuesta) a través del precio (variable explicativa).

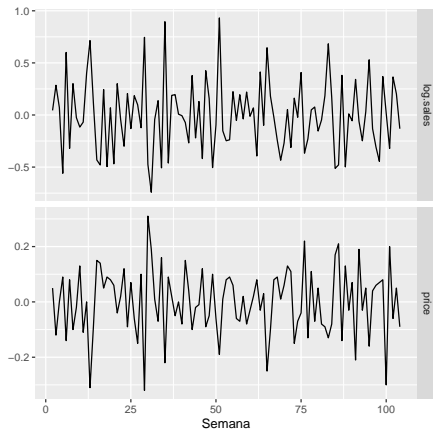
Para detectar la posible existencia de relación lineal transformaremos las series iniciales en series estacionarias (a no ser que ya lo fuesen): preblanquearemos dichas series y en base a sus correlaciones cruzadas sugeriremos un modelo para las series iniciales.

El gráfico de la derecha sugiere que las series en estudio no son estacionarias.



```
autoplot(· , facets=TRUE)
```

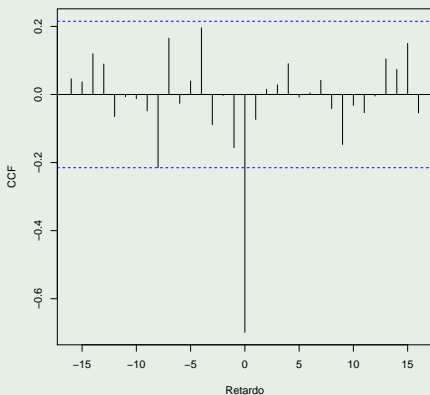

Series diferenciadas



El gráfico de la izquierda sugiere que las series diferenciadas regularmente son estacionarias.

Las correlaciones cruzadas de tales series preblanqueadas nos servirán para proponer un modelo de regresión lineal.

Series diferenciadas preblanqueadas: correlaciones cruzadas



El gráfico de la izquierda sugiere:

- Existencia de relación lineal:

$$Y_t = \beta_0 + \beta_1 X_{t-r} + e_t,$$

donde $Y_t = \log(\text{Ventas}_t)$ y
 $X_{t-r} = \text{Precio}_{t-r}$;

- $r = 0$: relación contemporánea (o instantánea).

`TSA::prewhiten(· , ·)`

El siguiente paso consiste en sugerir un modelo ARMA para los errores $\{e_t\}$ y estimar todos los parámetros del modelo (β_0 , β_1 y también los parámetros del ARMA).

Dicha sugerencia podría hacerse a través del modelo ARMA que dé lugar al menor valor de un criterio de información (AIC, AICc, BIC) en la estimación del modelo de regresión planteado (aquel que modeliza tanto a la regresión como a los errores $\{e_t\}$).

```
auto.arima(..., xreg=..., ...)
```

El criterio AICc propuso un modelo MA(4) para $\{e_t\}$.

El resultado de ajustar el modelo de regresión correspondiente,

$$Y_t = \beta_0 + \beta_1 X_t + e_t, \text{ siendo } \{e_t\} \text{ un MA(4)}$$

o, equivalentemente, el modelo

$$Y_t = \beta_0 + \beta_1 X_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \theta_3 a_{t-3} + \theta_4 a_{t-4} + a_t$$

siendo $\{a_t\}$ ruido blanco, es:

$$Y_t = 15.86 - 2.47X_t + 0.29a_{t-2} + 0.54a_{t-4} + a_t,$$

siendo $\hat{\sigma}_a^2 = 0.02728$. No se ha detectado la presencia de atípicos.

Nota: Recuerdese que $Y_t = \log(\text{Ventas}_t)$ y $X_t = \text{Precio}_t$.

`Arima(..., xreg=..., fixed=..., ...)`

El modelo ajustado superó el análisis de residuos (incluyendo normalidad en las innovaciones $\{a_t\}$). Por lo tanto puede ser utilizado para expresar el efecto que ejercen los precios sobre $\log(\text{ventas})$, y también para predecir valores de dichas ventas.

Así, según nuestro modelo,

$$Y_t = 15.86 - 2.47X_t + 0.29a_{t-2} + 0.54a_{t-4} + a_t,$$

se tiene que si el precio aumenta en 1 unidad, el $\log(\text{ventas})$ disminuirá, en media, en 2.47 unidades.

Si queremos utilizarlo para predecir el valor de $\log(\text{ventas})$ a la semana siguiente de la última observación, sabiendo que el precio en dicha semana será 1.9, se tendría $\hat{y}_T(1) = 11.05$ (equivalentemente, la predicción para las ventas sería $\exp(11.05)$).

```
forecast(..., xreg=1.9, h=1)
```