

# Procesos de memoria larga

## 1 Introducción

Vimos en el capítulo anterior que los procesos  $AR(p)$  tiene una función de autocorrelación  $\rho(h)$  que tiende a cero muy rápidamente (exponencialmente) a cero cuando  $h \rightarrow +\infty$ . Este resultado se puede generalizar a todo proceso  $ARMA(p, q)$ . En la vida real existen muchas series de tiempo pertenecientes a diversas áreas de las ciencias en las cuales la función de autocorrelación tiende a cero de manera más lenta, a ese tipo de procesos es a los que llamaremos de “memoria larga”. Son ejemplos en los cuales cuando uno observa las función de autocorrelación, la misma tiende a cero mucho más lentamente que las que se obtienen ajustandole cualquier modelo  $ARMA$ . El fenómeno de la memoria larga se encuentra en ejemplos como los de aportes hidrológicos, precipitaciones, anillos de crecimiento en los árboles, fenómenos bursátiles sobre valor de rescate de ciertas acciones, grosor de las capas de barro en los lagos, velocidad del viento, etc. El término memoria larga tiene que ver con el hecho de que podemos tener  $\rho(h)$  no muy pequeño para valores moderadamente altos de  $h$ , dicho de otra forma puede existir una correlación importante entre un dato actual y un dato de un pasado remoto ( $h$  períodos anteriores), de modo que la serie tiene “memoria” a largo plazo. Modelar una serie de tiempo de estas características a través de procesos  $ARMA$  cuyo  $\rho$  tiende a cero exponencialmente, no sería lo más adecuado.

## 2 Preliminares

En esta sección haremos un recordatorio de algunos conceptos matemáticos básicos que se necesitan para este capítulo.

### 2.1 Recordatorio de algunas definiciones y resultados sobre sumas infinitas (series)

1. Una serie  $\sum_{n=1}^{+\infty} a_n$  es convergente si y sólo si  $\lim_{n \rightarrow +\infty} \sum_{i=1}^n a_i = L$  número finito y en ese caso se dice que la suma infinita es igual a  $L$ , o sea  $\sum_{n=1}^{+\infty} a_n = L$ .
2. Una serie  $\sum_{n=1}^{+\infty} a_n$  se dice absolutamente convergente si y sólo si  $\sum_{n=1}^{+\infty} |a_n|$  es convergente (es decir que converge en valor absoluto).
3. Si  $\sum_{n=1}^{+\infty} |a_n|$  es convergente entonces  $\sum_{n=1}^{+\infty} a_n$  es convergente (es decir que si una serie es absolutamente convergente, entonces es convergente).
4. No vale el recíproco en general del teorema anterior, es decir que puede pasar que  $\sum_{n=1}^{+\infty} a_n$  sea convergente pero  $\sum_{n=1}^{+\infty} |a_n|$  no sea convergente.

5. Serie armónica.

$$\sum_{n=1}^{+\infty} \frac{1}{n^\alpha} \begin{cases} \text{converge si } \alpha > 1 \\ \text{no converge si } \alpha \leq 1 \end{cases} .$$

6. Condición necesaria de convergencia. Si  $\sum_{n=1}^{+\infty} a_n$  es convergente entonces  $\lim_{n \rightarrow +\infty} a_n = 0$ .

7. No vale en general el recíproco de la propiedad anterior (por ejemplo  $a_n = \frac{1}{n}$  cumple la condición  $\lim_{n \rightarrow +\infty} a_n = 0$  y sin embargo  $\sum_{n=1}^{+\infty} a_n = \sum_{n=1}^{+\infty} \frac{1}{n}$  no converge por la armónica).

Observemos que cuando una sucesión cumple las condiciones  $\lim_{n \rightarrow +\infty} a_n = 0$  y  $\sum_{n=1}^{+\infty} a_n$  no converge es porque  $a_n$  converge “lentamente” a cero. Por ejemplo observamos que las sucesiones  $\frac{1}{n}$  y  $\frac{1}{n^2}$  convergen ambas a 0 pero  $\sum_{n=1}^{+\infty} \frac{1}{n} = +\infty$ , mientras que  $\sum_{n=1}^{+\infty} \frac{1}{n^2}$  converge, claramente  $\frac{1}{n^2}$  tiende a cero más rápidamente que  $\frac{1}{n}$ .

## 2.2 La Función Gamma

La función Gamma es la extensión del operador factorial (que se aplica sólo a los números naturales) para valores reales.

**Definición 1** Si  $\alpha > 0$ , entonces  $\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$ .

Se prueba que para  $\alpha > 0$ , la función está bien definida porque la integral es convergente.

Listamos a continuación algunas de las propiedades que cumple dicha función.

1. Si  $\alpha > 1$ , entonces  $\Gamma(\alpha) = (\alpha - 1) \Gamma(\alpha - 1)$ .
2. Si  $n \in \mathbb{N}$ ,  $\Gamma(n) = (n - 1)!$ .
3.  $\Gamma(1/2) = \sqrt{\pi}$ .

**Nota.** Para valores negativos salvo los enteros negativos, la función Gamma se define como  $\Gamma(\alpha) = \frac{1}{\alpha} \prod_{n=1}^{+\infty} \frac{(1+1/n)^\alpha}{1+\alpha/n}$ .

La expresión anterior es válida también para valores de  $\alpha > 0$ , es decir que se puede probar que cuando  $\alpha > 0$  la expresión del miembro derecho de la igualdad anterior coincide con  $\int_0^{+\infty} x^{\alpha-1} e^{-x} dx$ .

### 2.3 Desarrollo en serie de potencias de una función

Cuando una función  $f$  se puede derivar infinitas veces en un entorno de 0, es posible desarrollarla como la suma de una serie de potencias de la forma  $f(x) = \sum_{n=1}^{+\infty} a_n x^n$  para  $x$  variando en un entorno de la forma  $(-r, r)$ . El resultado más preciso lo da el siguiente teorema.

**Teorema 2** *Si  $f : \mathbb{R} \rightarrow \mathbb{R}$  es infinitamente derivable en un entorno centrado en el origen, entonces existe  $r > 0$  tal que*

$$f(x) = \sum_{n=0}^{+\infty} \frac{f^{(n)}(0)}{n!} x^n \text{ para todo } x \in (-r, r).$$

Un desarrollo como el que indica el teorema anterior, es el llamado desarrollo de Taylor de la función  $f$  alrededor de 0.

Aplicando el teorema anterior a la función  $f(x) = (1-x)^d$ , se llega (derivando sucesivamente hasta llegar a una fórmula explícita para  $f^{(n)}(0)$ ) a que su desarrollo de Taylor viene dado por la siguiente fórmula:

$$(1-x)^d = \sum_{n=0}^{+\infty} b_n x^n \text{ donde } b_n = (-1)^n \prod_{k=1}^n \frac{d-k+1}{k} = (-1)^n \frac{\Gamma(d+1)}{\Gamma(d-n+1)\Gamma(n+1)}$$

igualdad válida para  $x \in (-1, 1)$ .

## 3 Procesos de memoria corta y de memoria larga

Vimos en el capítulo anterior que la ACF de un modelo es información de gran utilidad en el modelado de una serie de tiempo. Si un proceso cumple que  $\rho(h) \rightarrow 0$  como habitualmente ocurre en la práctica (en la medida en que dos observaciones estén muy distanciadas en el tiempo se influirán mutuamente muy poco), distinguiremos si la memoria del proceso es corta o larga según la función  $\rho(h)$  tienda a cero rápidamente o lentamente, en términos de que la serie sea convergente absolutamente o no.

**Definición 3** *Un proceso estocástico  $\{X_t\}_{t \geq 0}$  en  $L^2$ , débilmente estacionario se dice que es de memoria larga si y sólo si*

$$\rho(h) \xrightarrow{h \rightarrow +\infty} 0 \text{ y } \sum_{h=1}^{+\infty} |\rho(h)| = +\infty$$

*y diremos que es de memoria corta si y sólo si*

$$\rho(h) \xrightarrow{h \rightarrow +\infty} 0 \text{ y } \sum_{h=1}^{+\infty} |\rho(h)| < +\infty.$$

Existen otras definiciones de procesos de memoria larga, no todas equivalentes entre sí, pero todas ellas tienen en común el hecho de que  $\rho(h)$  tienda lentamente a cero. En nuestra definición la tendencia rápida o lenta a cero la realizamos a través de ver el comportamiento de la serie si es convergente o no respectivamente. Los procesos ARMA son procesos de memoria corta, debido al siguiente teorema que formaliza lo dicho anteriormente sobre el decrecimiento exponencial de su función de autocorrelación.

**Teorema 4**  $\{X_t\}_{t \in \mathbb{Z}}$  es un proceso ARMA( $p, q$ ), entonces existen constantes  $c > 0$  y  $0 < r < 1$  tales que  $|\rho(h)| \leq cr^{|h|}$  para todo  $h \in \mathbb{Z}$ .

Al ser  $r$  un valor en  $(0, 1)$  resulta que la serie de los  $\rho(h)$  es menor o igual que la serie de una geométrica con razón  $r$  por lo que la serie es convergente y el proceso queda de memoria corta.

A partir de este resultado se puede deducir que tanto los ARIMA y los SARIMA son de memoria corta, lo cual es esperable porque son los ARMA a los que se les quita la tendencia y/o la componente estacional.

## 4 Procesos ARFIMA (o también FARIMA)

Aprovechando el hecho que  $(1 - B)^d$  puede ser expresado en serie de potencias, el operador  $(1 - B)^d$  tiene sentido ser aplicado para valores de  $d$  fraccionales, por lo que veremos a continuación que un proceso ARFIMA( $p, d, q$ ) es como el proceso ARIMA( $p, d, q$ ) donde se permite que el parámetro  $d$  sea fraccionario.

**Definición 5** Si  $\{X_t\}_{t \in \mathbb{Z}}$  es un proceso estocástico estacionario y centrado, se dice que es un proceso ARFIMA( $p, d, q$ ) si y sólo si existen  $d \in (-1/2, 1/2)$ ,  $p$  y  $q$  naturales tales que el proceso  $\{T_t\}_{t \in \mathbb{Z}}$  definido como  $Y_t = (1 - B)^d X_t$  es un proceso ARMA( $p, q$ ).

Más explícitamente,  $Y_t = \sum_{s=0}^{+\infty} b(s)X_{t-s}$  donde  $b(s) = (-1)^s \prod_{k=1}^s \frac{d-k+1}{k} = \frac{(-1)^s \Gamma(d+1)}{\Gamma(d-s+1)\Gamma(s+1)}$ .

Entonces  $\{X_t\}_{t \in \mathbb{Z}}$  es un proceso ARFIMA( $p, d, q$ ) si y sólo si existen polinomios  $\phi$  de grado  $p$ ,  $\theta$  de grado  $q$  y  $d \in (-1/2, 1/2)$  tales que

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

donde  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  es ruido blanco.

**Observación 6** ARFIMA( $p, 0, q$ )=ARMA( $p, q$ ), es decir que cuando  $d = 0$  el ARFIMA queda un ARMA, de modo que se puede pensar a los procesos ARFIMA como una generalización de los procesos ARMA.

El siguiente teorema nos explica el por qué considerar  $d \in (-1/2, 1/2)$  en la definición de todo ARFIMA( $p, d, q$ ).

**Teorema 7** Si  $\{X_t\}_{t \in \mathbb{Z}}$  es un proceso que cumple la igualdad funcional  $\phi(B)(1-B)^d X_t = \theta(B)\varepsilon_t$  siendo  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  ruido blanco, entonces

1.  $\{X_t\}_{t \in \mathbb{Z}}$  es estacionario y causal cuando  $d < 1/2$  y  $\phi$  no tiene raíces en el círculo unidad.
2.  $\{X_t\}_{t \in \mathbb{Z}}$  es invertible cuando  $d > -1/2$  y  $\theta$  no tiene raíces en el círculo unidad.

#### 4.1 ACF y PACF de un ARFIMA

Dado que en los procesos ARMA no existen fórmulas explícitas para la función de autocorrelación o para la función de autocorrelación parcial, en los ARFIMA no puede haberlas.

Se puede probar que en un proceso ARFIMA( $p, d, q$ ) la función de autocorrelación cumple la siguiente equivalencia asintótica válida para  $d \in (-1/2, 1/2)$ ,  $d \neq 0$ :

$$\rho(h) \sim \frac{c}{|h|^{1-2d}} \text{ cuando } |h| \rightarrow +\infty \text{ siendo } c \text{ una constante}$$

donde la constante  $c$  es constante como función de  $h$ , pero sí depende de los parámetros  $\theta, \phi, d$ .

Utilizando el criterio de la serie armónica, deducimos que  $\sum_{h=1}^{+\infty} |\rho(h)|$  es convergente si y sólo si  $1 - 2d > 1$ , o sea cuando  $d < 0$ . Teniendo en cuenta que  $-1/2 < d < 1/2$  por definición, deducimos que

$$\text{ARFIMA}(p, d, q) \text{ es de memoria corta cuando } -1/2 < d \leq 0$$

y

$$\text{ARFIMA}(p, d, q) \text{ es de memoria larga cuando } 0 < d < 1/2.$$

En cuanto a la PACF se puede probar que cuando  $|h| \rightarrow +\infty$  y  $d \in (0, 1/2)$ , entonces la PACF es equivalente a  $\frac{d}{|h|}$ . En la Figura 1, vemos la función de autocorrelación muestral para un proceso AR(1) junto con la de un ARFIMA(1,  $d$ , 0) para  $d = 0.2$ ,  $d = 0.3$  y  $d = 0.4$ . Vemos cómo el proceso AR(1) tiende a cero muy rápidamente mientras que los ARFIMA tienden a cero mucho más lentamente, y que la convergencia es cada vez más lenta en la medida que aumentamos el valor de  $d$ . En el gráfico se marca el coeficiente de Hurst asociado al proceso ARFIMA definido mediante  $H = d + 1/2$ .

#### 4.2 Representación de un ARFIMA mediante un AR( $\infty$ ) y/o un MA( $\infty$ )

**Teorema 8** Si  $\{X_t\}_{t \in \mathbb{Z}}$  es un ARFIMA( $p, d, q$ ) de la forma  $\phi(B)(1-B)^d X_t = \theta(B)\varepsilon_t$  donde los polinomios  $\phi$  y  $\theta$  no tienen raíces en el disco cerrado  $\{z \in \mathbb{C}: |z| \leq 1\}$ , entonces el proceso es causal e invertible. Es decir que podemos obtener la representación

$$X_t = \phi^{-1}(B)\theta(B)(1-B)^{-d}\varepsilon_t = \psi(B)\varepsilon_t = \sum_{i=0}^{+\infty} \psi_i \varepsilon_{t-i}$$

Función de autocorrelación para  
5000 datos simulados.

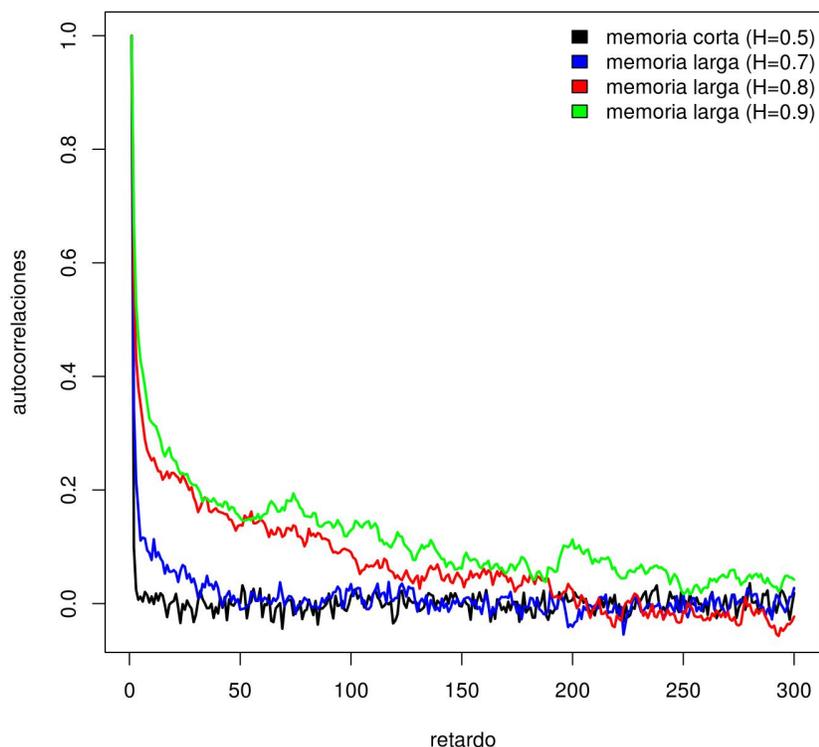


Figura 1: Simulación de la acf de AR(1) (en negro) junto con la de un ARFIMA(1,  $d$ , 0) para  $d = 0.2$  (en azul),  $d = 0.3$  (en rojo) y  $d = 0.4$  (en verde).

donde los coeficientes  $\psi_i$  surgen de la igualdad  $\sum_{i=0}^{+\infty} \psi_i z^i = \frac{\theta(z)}{\phi(z)(1-z)^d}$  para todo  $|z| \leq 1$  y también la representación

$$\varepsilon_t = \phi(B)(1-B)^d \theta^{-1}(B) X_t = \pi(B) X_t = \sum_{i=0}^{+\infty} \pi_i X_{t-i}$$

donde los coeficientes  $\pi_i$  surgen de la igualdad  $\sum_{i=0}^{+\infty} \pi_i z^i = \frac{\phi(z)(1-z)^d}{\theta(z)}$  para todo  $|z| \leq 1$ .

Del teorema anterior, se deduce de manera inmediata que todo ARFIMA( $p, d, q$ ) causal e invertible podemos escribirlo como un AR( $\infty$ ) y como un MA( $\infty$ ) ya

que del cálculo de los coeficientes  $\psi_i$  y  $\pi_i$  se deduce que  $\psi_0 = \pi_0 = 1$ . Entonces,

la igualdad

$$X_t = \sum_{i=0}^{+\infty} \psi_i \varepsilon_{t-i} = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \psi_3 \varepsilon_{t-3} + \dots$$

la podemos leer como que el proceso  $\{X_t\}_{t \in \mathbb{Z}}$  es un  $MA(\infty)$ , mientras que la igualdad

$$\varepsilon_t = \sum_{i=0}^{+\infty} \pi_i X_{t-i} = X_t + \pi_1 X_{t-1} + \pi_2 X_{t-2} + \pi_3 X_{t-3} + \dots$$

que la leemos como  $X_t = -\pi_1 X_{t-1} - \pi_2 X_{t-2} - \pi_3 X_{t-3} \dots + \varepsilon_t$  que sería un  $AR(\infty)$ .

La representación del ARFIMA como un  $AR(\infty)$  da interesante información respecto a la dependencia de un valor  $X_t$  respecto de los anteriores. Como la serie es convergente, los  $\pi_i$  tienden a cero, por lo que a partir de un momento en adelante los  $\pi_i$  van a ser cercanos a cero. Esos primeros valores indican el peso en el modelo que tienen los correspondientes valores pasados en la observación  $X_t$ .

**Ejemplo 9** *En el ejemplo de los aportes complexivos a las represas en Uruguay, luego de aplicada una función que gaussianiza los datos y quitada la componente estacional, el modelo ajustado fue el siguiente ARFIMA(3, d, 1) :*

$$(1 - 1.211B + 0.599B^2 - 0.17B^3) (1 - B)^{0.251} X_t = (1 - 0.377B) \varepsilon_t.$$

*La única raíz del polinomio  $\theta(z) = 1 - 0.377z$  es  $z = \frac{1}{0.377} = 2.6525$  que al estar fuera del disco cerrado  $|z| \leq 1$  nos permite afirmar que el proceso es invertible, por lo que podemos escribir al mismo como un  $AR(\infty)$ . Para hallar los coeficientes desarrollamos en series de potencias la función*

$$\psi(z) = \frac{(1 - 1.211z + 0.599z^2 - 0.17z^3) (1 - z)^{0.251}}{1 - 0.377z} =$$

$1.0 - 1.085z + 0.39992z^2 - 0.11055z^3 - 0.02661z^4 - 0.0995084z^5 - 0.0049818z^6 - 0.0036802z^7 - 0.0032314z^8 - 0.0029941z^9 - 0.0028029z^{10} + \dots$  *Por lo tanto el proceso lo podemos desarrollar como*

$$\begin{aligned} X_t = & 1.085X_{t-1} - 0.39992X_{t-2} + 0.11055X_{t-3} + 0.02661X_{t-4} + \\ & 0.0995084X_{t-5} + 0.0049818X_{t-6} + 0.036802X_{t-7} + 0.0032314X_{t-8} \\ & + 0.0029941X_{t-9} + 0.0028029X_{t-10} + \dots + \varepsilon_t. \end{aligned}$$

De este desarrollo podemos ver el peso que tienen las variables pasadas en la observación al tiempo  $t$ , por ejemplo el peso de la variable  $X_{t-1}$  es casi el triple de la  $X_{t-2}$ , es casi 10 veces mayor a la contribución de  $X_{t-3}$ , etc.

El proceso también es causal porque las raíces del polinomio  $\phi(z) = 1 - 1.211z + 0.599z^2 - 0.17z^3 = 0$  son  $z_1 = 1.4227$ ,  $z_2 = 1.0504 - 1.7411i$  y  $z_3 = 1.0504 + 1.7411i$  y todas ellas caen fuera del disco  $|z| \leq 1$  porque  $|z_1| = 1.4227$ ,  $|z_2| = |z_3| = \sqrt{1.0504^2 + 1.7411^2} = 2.0334$  por lo que el proceso podemos representarlo como un  $MA(\infty)$  con cálculo similar. Ahora desarrollamos en series de potencias la función  $\pi(z) = \frac{\phi(z)(1-z)^d}{\theta(z)} = \frac{1-0.377z}{(1-1.211z+0.599z^2-0.17z^3)(1-z)^{0.251}} = 1.0 + 1.085z + 0.77731z^2 + 0.52002z^3 + 0.39992z^4 + 0.35026z^5 + 0.31357z^6 + 0.27467z^7 + 0.23802z^8 + 0.20815z^9 + 0.18516z^{10} + \dots$  por lo que escribimos

$$\begin{aligned} X_t = & \varepsilon_t + 1.085\varepsilon_{t-1} + 0.77731\varepsilon_{t-2} + 0.52002\varepsilon_{t-3} + 0.39992\varepsilon_{t-4} + \\ & 0.35026\varepsilon_{t-5} + 0.31357\varepsilon_{t-6} + 0.27467\varepsilon_{t-7} + 0.23802\varepsilon_{t-8} + \\ & 0.20815\varepsilon_{t-9} + 0.18516\varepsilon_{t-10} + \dots \end{aligned}$$

## 5 Estimación de parámetros

Mientras que en el ARIMA se deben ingresar previamente los valores de  $p$  y  $q$  para ajustar, así como el valor de  $d$ , en los ARFIMA( $p, d, q$ ) es necesario ingresar sólo los valores de  $p$  y  $q$ , ya que el parámetro  $d$  correspondiente a la memoria larga debe ser estimado también. Al igual que con los ARIMA, los valores de  $p$  y  $q$  más adecuados se pueden encontrar minimizando el AIC o BIC por ejemplo (o por algún otro criterio).

## 6 Paquetes en R para procesos de memoria larga

Hay varios paquetes en R que incluyen a los ARFIMA. Entre ellos se encuentran por ejemplo arfima, fArma, longmemo, FGN.

El paquete “arfima” funciona en forma similar al “arima”. Por ejemplo, dado el vector de observaciones  $x$  (que como siempre deben ser datos provenientes de un proceso estacionario, o sea que debo centrar y quitar la tendencia y la componente estacional en caso de tenerlas), mediante la sentencia

```
fit <- arfima(x, order = c(p, 0, q), numeach = c(3, 3))
```

ajustamos un ARFIMA( $p, d, q$ ) al vector de datos  $x$ . En `order=c(p,0,q)`, se incluyen los valores de  $p$  y  $q$ , mientras que el valor del medio va en 0, porque en el parámetro  $d$  será estimado. Como siempre mediante la sentencia `?arfima` encontramos toda la información con respecto a dicha función. El argumento `numeach`, es el número de puntos iniciales para la estimación de los parámetros. El primero de ellos es para estimar los parámetros correspondientes al AR y al MA del modelo mientras que el segundo es para estimar el parámetro fraccional

d. “fit” nos ajusta varios modelos con verosimilitudes similares y nos dará los coeficientes de todos ellos.

`resid([1])#` nos da los residuos del modelo 1 ajustado.

Si los residuos no se comportan como ruido blanco, se puede probar con los residuos del modelo 2 por ejemplo mediante `resid([2])`.

Hacemos predicciones nuevamente con `predict`, por ejemplo `predict(fit,2)` me dará las dos predicciones siguientes para todos los modelos ajustados en `fit`.

`predict(fit,2)[1]$Forecast` nos da las 2 predicciones del modelo 1.

`plot(predict(fit,2))` nos grafica los últimos datos de la serie, a la que le agrega las dos predicciones junto con sus intervalos de confianza al 95%.

## 7 Procesos de Ornstein-Uhlenbeck fraccionarios

En esta sección y en las próximas, veremos procesos indizados en tiempo continuo, esto es, procesos donde la variable tiempo varía en un intervalo. Modelar series de tiempo a través de procesos estocásticos a tiempo continuo es más natural, ya que muchos fenómenos son observados en momentos de tiempo en particular, pero son observaciones que están sumergidas en un proceso a tiempo continuo. Dicho de otra forma, muchas veces uno tiene un objeto a tiempo continuo (por ejemplo la velocidad del viento a lo largo de cada instante de un día) del cual sólo obtiene un conjunto finito de observaciones. Por otro lado, como en la práctica todo conjunto de observaciones son una cantidad finita de datos, se puede modelar sin problemas con procesos a tiempo discreto como todos los vistos hasta ahora. Comenzamos definiendo el movimiento Browniano fraccional.

**Definición 10** *Se dice que un proceso  $\{B_H(t)\}_{t \in \mathbb{R}}$  es un movimiento browniano fraccional (mBf) de parámetro  $H \in (0, 1)$  si y sólo si:*

- es gaussiano;
- tiene trayectorias continuas c.s.;
- es centrado (esto es  $\mathbb{E}(B_H(t)) = 0$  para todo  $t$ );
- $\mathbb{E}(B_H(t)B_H(s)) = \frac{1}{2} \left( |t|^{2H} + |s|^{2H} - |s - t|^{2H} \right)$  para todos  $s, t$ .

**Observación 11**  $B_H(0) = 0$ .

**Observación 12** *Como se ve de la función de autocovarianzas de un mBf, el proceso es sí no es estacionario, pero se construye un proceso estacionario tomando incrementos.*

**Teorema 13** *Los incrementos de un proceso mBf son estacionarios y en el caso en que  $H > 1/2$ , son de memoria larga.*

**Proof.** Si le llamamos  $X_t = B_H(t) - B_H(t-1)$ , entonces

$$\begin{aligned} \mathbb{E}(X_t X_{t+n}) &= \\ \mathbb{E}((B_H(t) - B_H(t-1))(B_H(t+n) - B_H(t+n-1))) &= \\ \frac{1}{2} \left[ (n+1)^{2H} + (n-1)^{2H} - 2n^{2H} \right] &= \\ \frac{n^{2H}}{2} \left[ \left(1 + \frac{1}{n}\right)^{2H} + \left(1 - \frac{1}{n}\right)^{2H} - 2 \right] &\sim \\ H(2H-1)n^{2H-2} & \end{aligned}$$

por lo que si  $H > 1/2$  se tiene que  $\rho(n) \rightarrow 0$  y  $\sum_{n=1}^{+\infty} \rho(n) = +\infty$ . ■

A partir de aquí se deduce además que, para  $n$  suficientemente grande

- si  $H > 1/2$  los incrementos del mBf están positivamente correlacionados,
- si  $H < 1/2$  los incrementos del mBf están negativamente correlacionados,
- si  $H = 1/2$ , los incrementos son no correlacionados y por lo tanto independientes.

Dado un mBf  $\{B_H(t)\}_{t \in \mathbb{R}}$  cuando consideramos el proceso de los incrementos con  $t$  variando en los enteros, obtenemos un proceso estacionario a tiempo discreto que se llama ruido Gaussiano fraccional (FGN).

**Definición 14** Si  $\{B_H(t)\}_{t \in \mathbb{R}}$  es un mBf, entonces el proceso  $\{Y_t\}_{t \in \mathbb{Z}}$  definido mediante

$$Y_t = B_H(t) - B_H(t-1) \text{ para todo } t \in \mathbb{Z}$$

se llama ruido Gaussiano fraccional.

Cuando  $H = 1/2$  el movimiento Browniano fraccional se transforma en el movimiento Browniano estándar.

**Definición 15** Se dice que un proceso  $\{B(t)\}_{t \in \mathbb{R}}$  es un movimiento Browniano (mB) si y sólo si:

- es gaussiano;
- tiene trayectorias continuas c.s.;
- es centrado (esto es  $\mathbb{E}(B(t)) = 0$  para todo  $t$ );
- $\mathbb{E}(B(t)B(s)) = \frac{1}{2}(|t| + |s| - |s-t|) = s \wedge t$  para todos  $s, t$ .

Las trayectorias de un mBf son continuas pero no diferenciables con probabilidad 1 casi seguramente) como lo demuestra el siguiente teorema.

**Teorema 16** Dado un mBf  $\{B_H(t)\}_{t \in \mathbb{R}}$ , entonces

$$\limsup_{t \rightarrow t_0} \left| \frac{B_H(t) - B_H(t_0)}{t - t_0} \right| = +\infty \text{ c.s.}$$

A partir de la definición del mBf podemos definir el proceso de Ornstein-Uhlenbeck fraccionario.

**Definición 17** Dados  $\lambda, \sigma > 0$  y  $\{B_H(t)\}_{t \in \mathbb{R}}$  movimiento browniano fraccional, si para cada  $t$  definimos  $X_t = \sigma \int_{-\infty}^t e^{-\lambda(t-s)} dB_H(s)$ , entonces se dice que  $\{X_t\}_{t \in \mathbb{R}}$  es un proceso de Ornstein-Uhlenbeck fraccionario.

Notación:  $\{X_t\}_{t \in \mathbb{R}} \sim FOU(\lambda, \sigma, H)$ .

La integral de la definición anterior respecto del mBf, es la de Riemann-Stieltjes, y su existencia está justificada en un trabajo publicado por Cheridito, Kawaguchi & Maejima (2003). Más explícitamente, dicha integral se define como

$$\int_{-\infty}^t e^{-\lambda(t-s)} dB_H(s) = \lim_{u \rightarrow -\infty} \int_u^t e^{-\lambda(t-s)} dB_H(s)$$

siendo

$$\int_u^t e^{-\lambda(t-s)} dB_H(s) = \lim_{\substack{n \rightarrow +\infty \\ s_i - s_{i-1} \rightarrow 0 \\ \text{para todo } i}} \sum_{i=1}^n e^{-\lambda(t-s_i)} (B_H(s_i) - B_H(s_{i-1})).$$

En el límite anterior, los  $s_i$  son una partición del intervalo  $[u, t]$  es decir que  $u = s_0 < s_1 < s_2 < \dots < s_n = t$ .

Al ser los incrementos  $B_H(s_i) - B_H(s_{i-1})$  centrados, estacionarios y Gaussianos, el pasaje al límite (cuando estamos en variables gaussianas) mantiene dichas propiedades.

**Propiedad importante.**

Todo  $FOU(\lambda, \sigma, H)$  es un proceso centrado, Gaussiano y estacionario por ser límite de incrementos del mBf que es centrado, Gaussiano y estacionario.

Se puede ver que dicha definición, es la única solución estacionaria de la ecuación diferencial estocástica

$$dX_t = -\lambda X_t dt + \sigma dB_H(t).$$

Observar que intuitivamente la parte  $-\lambda X_t dt$  es la que termina haciendo aparecer la exponencial en la definición y la parte  $\sigma dB_H(t)$  es la que genera el ruido en la ecuación diferencial que juega el papel del ruido blanco (en algún sentido).

La no derivabilidad de las trayectorias de un mBf, traslada dicha propiedad a los FOU, como lo muestra el siguiente teorema.

**Teorema 18** Si  $\{X_t\}_{t \in \mathbb{R}} \sim FOU(\lambda, \sigma, H)$ , entonces sus trayectorias son no derivables en todo punto con probabilidad 1.

En la Figura 2 se observa la trayectoria de un proceso  $FOU(\lambda = 0.3, \sigma = 1, H = 0.7)$  en el intervalo  $[0, 100]$  observado en  $n = 10.000$  puntos, se visualiza la continuidad de las trayectorias mientras que la cantidad de irregularidades muestra la no derivabilidad.

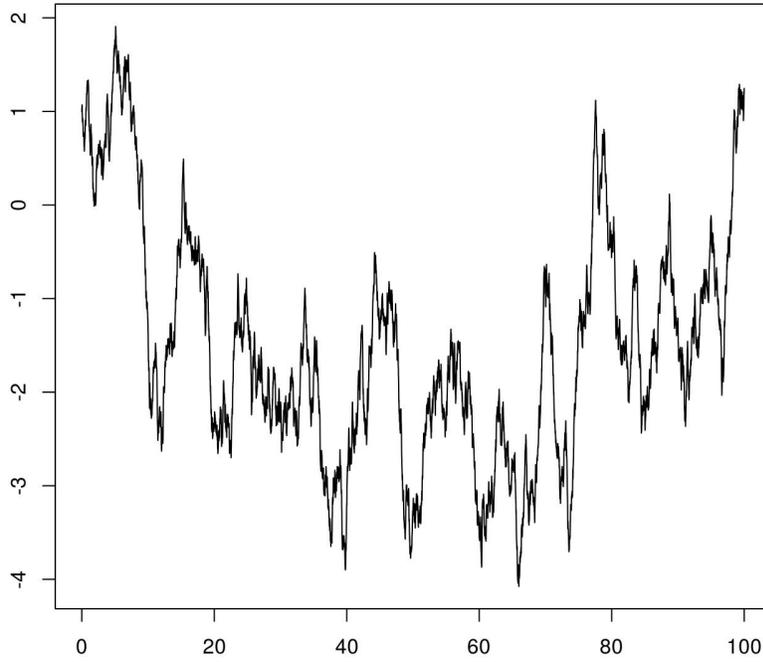


Figura 2: Trayectoria de un proceso  $FOU(\lambda = 0.3, \sigma = 1, H = 0.7)$  en el intervalo  $[0, 100]$  observado en  $n = 10.000$  puntos.

### 7.1 Función de autocovarianzas de un FOU

Se prueba que la función de autocovarianzas de un  $FOU(\lambda, \sigma, H)$  viene dada por

$$\mathbb{E}(X_0 X_t) = \frac{\sigma^2 H f_H(\lambda t)}{2\lambda^{2H}}$$

siendo  $f_H$  la función definida mediante

$$f_H(x) := e^{-x} \left( \Gamma(2H) - \int_0^x e^s s^{2H-1} ds \right) + e^x \left( \Gamma(2H) - \int_0^x e^{-s} s^{2H-1} ds \right).$$

Tomando  $t = 0$  en la igualdad anterior y teniendo en cuenta que  $f_H(0) = 2\Gamma(2H)$ , obtenemos la varianza de las variables que componen el proceso:

$$\mathbb{V}(X_t) = \frac{\sigma^2 \Gamma(2H + 1)}{2\lambda^{2H}}.$$

Se puede probar que  $f_H(x) \sim 2(2H - 1)x^{2H-2}$  cuando  $x \rightarrow +\infty$  a partir de la que deducimos de manera inmediata que un FOU( $\lambda, \sigma, H$ ) es de memoria corta cuando  $H \leq 1/2$  y de memoria larga cuando  $H > 1/2$  por aplicación de la serie armónica. En la Figura 3 se observa el gráfico de las funciones  $f_H$  para distintos valores de  $H$  donde se muestra cómo converge cada vez más lentamente a cero en la medida que  $H$  aumenta.

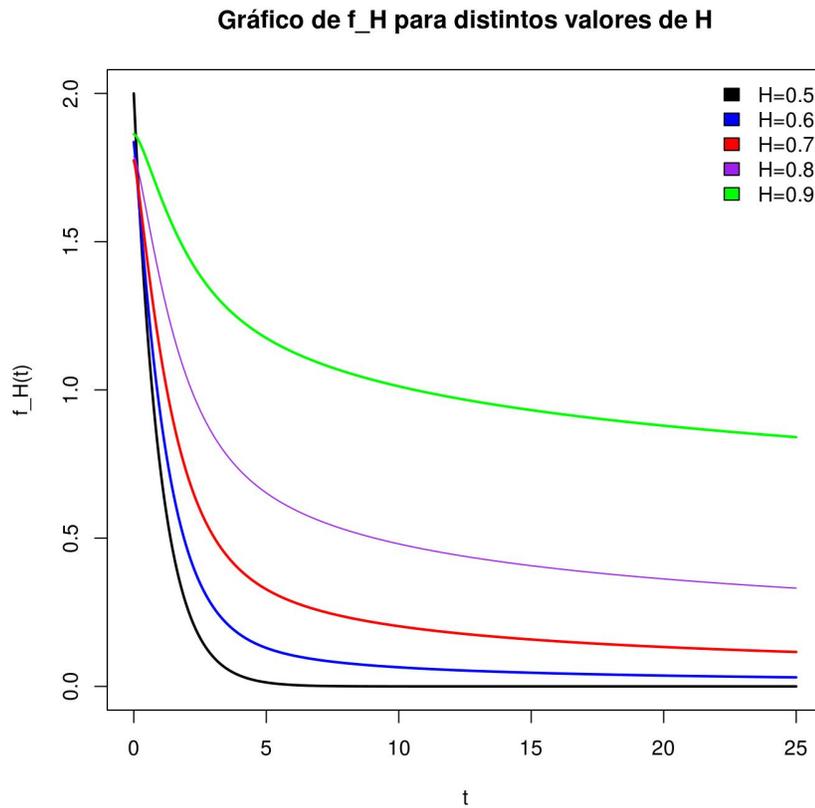


Figura 3: Gráfico de distintas  $f_H$  variando  $H$ .

## 7.2 Estimación de parámetros en un FOU

Dado un conjunto de observaciones equiespaciadas en el tiempo provenientes de un proceso FOU( $\lambda, \sigma, H$ ) en determinado intervalo  $[0, T]$  es posible estimar los 3 parámetros, los estimadores además cumplen con propiedades de consistencia y normalidad asintótica. Explicamos en las líneas que siguen cómo se estiman los parámetros. La metodología de estimación (junto con las propiedades teóricas) que presentaremos es la que presentan Brouste & Iacus en un paper de 2013. Para ello debemos comenzar con la definición de filtro.

**Definición 19** *Se dice que  $a = (a_0, a_1, \dots, a_k)$  es un filtro de longitud  $k + 1$  y orden  $L \geq 1$  si y sólo si, se cumplen las siguientes condiciones:*

- $\sum_{i=0}^k a_i i^l = 0$  para todo  $0 \leq l \leq L - 1$ .
- $\sum_{i=0}^k a_i i^L \neq 0$ .

*Si además el filtro cumple que  $\sum_{i=0}^k (-1)^{i-1} a_i = 1$ , se dice que el filtro está normalizado.*

**Ejemplo 20**  $(-1, 1)$  es un filtro de orden 1 y longitud 2.

**Ejemplo 21**  $(-1, 2, -1)$  es un filtro de orden 2 y longitud 3.

**Ejemplo 22** Más en general  $(-1, k, -C_2^k, \dots, C_2^k, -k, 1)$  es un filtro de orden  $k$  y longitud  $k + 1$ .

Dado un filtro  $a$  de orden  $L$  y longitud  $k + 1$ , se puede construir uno de orden  $L$  y longitud  $2k + 1$  que se llama filtro  $a$  dilatado y se define como

$$a^2 := (a_0, 0, a_1, 0, a_2, 0, \dots, 0, a_k).$$

Supongamos que tenemos  $n$  observaciones de un proceso FOU( $\lambda, \sigma, H$ ) en el intervalo  $[0, T]$  a tiempos equiespaciados

$$X_\Delta, X_{2\Delta}, \dots, X_{i\Delta}, \dots, X_{n\Delta} = X_T.$$

Dado un filtro  $a$  de longitud  $k + 1$  y orden  $L$ , los autores definen

$$V_{n,a} := \frac{1}{n} \sum_{i=0}^{n-k} \left( \sum_{j=0}^k a_j X_{(i+j)\Delta_n} \right)^2$$

siendo  $\Delta_n := T_n/n$  luego  $H$  y  $\sigma$  se estiman mediante

$$\hat{H} = \frac{1}{2} \log_2 \left( \frac{V_{n,a^2}}{V_{n,a}} \right),$$

$$\hat{\sigma}^2 = \frac{-2V_{n,a}}{\Delta_n^{2\hat{H}} \sum_{k=0}^k \sum_{l=0}^k a_k a_l |k-l|^{2\hat{H}}}.$$

A partir de estos estimadores, se estima  $\lambda$  mediante

$$\hat{\lambda} = \left( \frac{2 \sum_{i=1}^n X_{i\Delta}^2}{n \hat{\sigma}^2 \Gamma(2\hat{H} + 1)} \right)^{\frac{-1}{2\hat{H}}}.$$

Las fórmulas para  $\hat{H}$  y  $\hat{\sigma}$  (así como las pruebas de consistencia y normalidad asintótica) están basados de un trabajo de Istas & Lang (1997), sobre procesos gaussianos de incrementos estacionarios donde  $H$  es el índice local de Hölder del proceso. Las propiedades asintóticas de los estimadores se resumen en el siguiente teorema debido a Brouste & Iacus (2013).

**Teorema 23** *Si el filtro  $a$  tiene orden  $L \geq 2$ ,  $T \rightarrow +\infty$  y  $\Delta_n \rightarrow 0$  cuando  $n \rightarrow +\infty$ , entonces*

1.  $(\hat{H}, \hat{\sigma}) \xrightarrow{c.s.} (H, \sigma).$
2.  $\sqrt{n}(\hat{H} - H) \xrightarrow{w} N(0, \Gamma_1(H, \sigma, a))$
3.  $\frac{\sqrt{n}}{\log n}(\hat{\sigma} - \sigma) \xrightarrow{w} N(0, \Gamma_2(H, \sigma, a)).$

Ajustando un poco la relación entre el tamaño de muestra y  $T$  se obtienen las propiedades asintóticas de  $\hat{\lambda}$  que vienen dadas en el siguiente teorema.

**Teorema 24** *Si  $H \in (1/2, 3/4)$  y existe  $p > 1$  tal que  $n\Delta_n^p \rightarrow 0$  cuando  $n \rightarrow +\infty$ , entonces*

1.  $\hat{\lambda} \xrightarrow{c.s.} \lambda.$
2.  $\sqrt{T_n}(\hat{\lambda} - \lambda) \xrightarrow{w} N\left(0, \lambda \left(\frac{\sigma_H}{2H}\right)^2\right).$

$$\text{siendo } \sigma_H^2 = (4H - 1) \left(1 + \frac{\Gamma(1-4H)\Gamma(4H-1)}{\Gamma(2-2H)\Gamma(2H)}\right).$$

**Observación 25** *La consistencia de  $\hat{\lambda}$  a  $\lambda$  está probada únicamente para cuando el valor de  $H$  está entre  $1/2$  y  $3/4$ . Para valores mayores a  $3/4$  es posible que el resultado sea cierto también y pueda ser probado por otro lado.*

**Observación 26** *Los teoremas de consistencia como se ve, exigen que  $n \rightarrow +\infty$  y que  $T \rightarrow +\infty$ , pero  $T$  debe tender más lentamente a infinito que el tamaño de muestra  $n$ .*

**Observación 27**  *$T_n = \ln(n)$  verifica las distintas condiciones de los teoremas de consistencia.*

### 7.3 Modelación mediante procesos FOU

Dada una serie de datos, una vez desestacionalizada y centrada, la propiedad anterior puede ser utilizada para modelar dicha serie de datos resultante. Como vemos, en la definición sus parámetros tienen una interpretación clara.  $H$  es el parámetro de Hurst, del cual sabemos que  $H \leq 1/2$  significa que el proceso es de memoria corta, mientras que si  $H > 1/2$ , el proceso es de memoria larga, y su función de autocovarianza tiende a cero más lentamente en la medida en que  $H$  aumenta.  $\sigma$  es un parámetro que al estar multiplicando sirve para ajustar la dispersión del modelo. El parámetro  $e^{-\lambda}$  juega un papel similar al del  $\phi_1$  en un proceso AR(1) debido al siguiente cálculo Si  $X_t = \int_{-\infty}^t e^{-\lambda(t-s)} dB_H(s)$ , entonces

$$\begin{aligned} \int_{t-1}^t e^{-\lambda(t-s)} dB_H(s) &= \int_{-\infty}^t e^{-\lambda(t-s)} dB_H(s) - \int_{-\infty}^{t-1} e^{-\lambda(t-s)} dB_H(s) = \\ X_t - e^{-\lambda} \int_{-\infty}^{t-1} e^{-\lambda(t-1-s)} dB_H(s) &= X_t - e^{-\lambda} X_{t-1}. \end{aligned}$$

Si le llamamos  $\varepsilon_t = \int_{t-1}^t e^{-\lambda(t-s)} dB_H(s)$ , entonces hemos probado que  $X_t - e^{-\lambda} X_{t-1} = \varepsilon_t$ , o sea que

$$X_t = e^{-\lambda} X_{t-1} + \varepsilon_t$$

donde  $\{\varepsilon_t\}$  es gaussiano, centrado y estacionario.

Por lo tanto tenemos una analogía entre el FOU( $\lambda, \sigma, H$ ) y un proceso AR(1) donde el parámetro  $\phi$  es  $e^{-\lambda}$ , por lo que puede tomarse como un parámetro que gobierna la memoria corta del proceso.

**Nota.** De los teoremas de convergencia de los estimadores de los parámetros a los verdaderos valores de los mismos, observamos que a la hora de implementar un modelo de este tipo a una serie de datos observada en el tiempo, se asume que las observaciones del proceso son equiespaciadas en el tiempo en determinado intervalo de la forma  $[0, T]$ .

### 7.4 El paquete “yuima” en R

El paquete “yuima” en R, es un paquete muy completo que tiene el potencial de realizar muchas cosas, en cuanto a la modelación mediante ecuaciones diferenciales estocásticas. Una de las funciones del paquete, es la de ajustar un modelo FOU( $\lambda, \sigma, H$ ) a un conjunto de observaciones medidas en un intervalo equiespaciado de tiempo. Para obtener las estimaciones de los 3 parámetros utilizamos las siguientes sentencias:

```
X=window(x)#siendo x el vector de observaciones (que deben estar desestacionalizados ya que los FOU son procesos estacionarios)
mod<-setModel(drift="-x*lambda",hurst=NA,diffusion="theta")
serieyuima=setYuima(data=setData(X),model=mod)
```

mmfrac(serieyuima)#con esta sentencia se estiman los parámetros.

H=mmfrac(serieyuima)\$coefficients[1]#nos devuelve la estimación de  $H$ .  
o2=mmfrac(serieyuima)\$coefficients[2]#estimación de  $\sigma$ .  
l=mmfrac(serieyuima)\$coefficients[3]#estimación de  $\lambda$ .

## 8 Procesos iterados de Ornstein–Uhlenbeck fraccionarios

Los procesos AR(1) si bien son muy intuitivos adolecen del defecto de tener sólo dos parámetros, lo cual le da poca versatilidad para captar la complejidad de muchas series de tiempo que existen en la vida real. Esta dificultad es superada mediante los procesos AR( $p$ ) o los ARMA( $p, q$ ) o los ARFIMA( $p, d, q$ ) si se trata de modelar series de memoria larga. En tiempo continuo, el proceso de Ornstein–Uhlenbeck fraccionario adolece del mismo defecto, ya que tiene un parámetro de dispersión  $\sigma$  que jugaría el papel de la varianza del ruido blanco en el AR(1),  $\lambda$  que juega un papel similar al del parámetro  $\phi$  para el AR(1) y agrega el parámetro de Hurst  $H$  como parámetro que sirve como indicar de presencia o no de memoria larga.

Los procesos iterados de Ornstein–Uhlenbeck fraccionarios, buscan agregar más parámetros del tipo  $\lambda$  manteniendo  $\sigma$  y  $H$  para lograr mayor versatilidad. La idea está basada en la siguiente idea. Si dado un proceso  $Y = \{Y_t\}_{t \in \mathbb{R}}$  definimos el operador  $T_\lambda(Y) = \int_{-\infty}^t e^{-\lambda(t-s)} dY(s)$ , entonces el FOU( $\lambda, \sigma, H$ ) se puede definir mediante  $T_\lambda(\sigma B_H)$  siendo  $\{B_H(t)\}_{t \in \mathbb{R}}$  un mBf.

Si componemos el operador  $T_\lambda$  para  $\lambda_1, \lambda_2, \dots, \lambda_p > 0$ , es decir  $T_{\lambda_1} \circ T_{\lambda_2} \circ \dots \circ T_{\lambda_p}$  y se lo aplicamos a  $\{\sigma B_H(t)\}_{t \in \mathbb{R}}$ , tendremos el proceso FOU( $p$ ) que los llamaremos FOU de orden  $p$  ( $p$  iteraciones del operador  $T_\lambda$  aplicadas a un mBf). Vamos a la definición de los FOU( $p$ ).

**Definición 28** Si  $\{\sigma B_H(s)\}_{s \in \mathbb{R}}$  es un mBf con parámetro de Hurst  $H$ , y parámetro de escala  $\sigma$ . Si además  $0 < \lambda_1 < \lambda_2 < \dots < \lambda_q$  y  $p_1, p_2, \dots, p_q \in \mathbb{N}$  son tales que  $p_1 + p_2 + \dots + p_q = p$ . Definimos  $\{X_t\}_{t \in \mathbb{R}}$  mediante

$$X_t := \prod_{i=1}^q T_{\lambda_i}^{p_i}(\sigma B_H)(t).$$

**Notación 29**  $\{X_t\}_{t \in \mathbb{R}} \sim FOU(\lambda_1^{(p_1)}, \lambda_2^{(p_2)}, \dots, \lambda_q^{(p_q)}, \sigma, H)$ , o más simplemente,  $\{X_t\}_{t \in \mathbb{R}} \sim FOU(p)$ .

La notación anterior indica que componemos el operador  $T_\lambda$   $p$  veces de la siguiente forma:  $p_1$  veces  $T_{\lambda_1}$ ,  $p_2$  veces  $T_{\lambda_2}, \dots, p_q$  veces  $T_{\lambda_q}$ , la productoria en la definición significa composición y los exponentes  $p_i$  indica que se compone  $p_i$  veces.

**Observación 30**  $FOU(1) = FOU$ . Es decir que si aplicamos una sólo composición de  $T_\lambda$  nos queda el FOU( $\lambda, \sigma, H$ ) de modo que los FOU de orden  $p$  generalizan los FOU.

## Propiedades

1. Todo FOU( $p$ ) tiene trayectorias continuas y no derivables en ningún punto con probabilidad 1.
2. Si  $p = 1$  y  $H > 1/2$  el proceso es de memoria larga.
3. Si  $p \geq 2$  entonces todo FOU( $p$ ) es de memoria corta, cualquiera sea  $H$ .
4. Todo FOU( $p$ ) cuyos  $\lambda$ 's son todos distintos entre sí, se puede escribir como combinación lineal de FOU de orden 1.

En particular, si tenemos un FOU(2) con dos  $\lambda$ 's distintos, es decir  $0 < \lambda_1 < \lambda_2$ , o sea  $\{X_t\}_{t \in \mathbb{R}} \sim \text{FOU}(\lambda_1, \lambda_2, \sigma, H)$  entonces si le llamo  $X_t^{(1)} = \sigma \int_{-\infty}^t e^{-\lambda_1(t-s)} dB_H(s)$ ,  $X_t^{(2)} = \sigma \int_{-\infty}^t e^{-\lambda_2(t-s)} dB_H(s)$  es decir que  $\{X_t^{(1)}\}_{t \in \mathbb{R}} \sim \text{FOU}(\lambda_1, \sigma, H)$  y  $\{X_t^{(2)}\}_{t \in \mathbb{R}} \sim \text{FOU}(\lambda_2, \sigma, H)$  conducidos por el mismo movimiento Browniano fraccional, entonces

$$X_t = \frac{\lambda_1}{\lambda_1 - \lambda_2} X_t^{(1)} + \frac{\lambda_2}{\lambda_2 - \lambda_1} X_t^{(2)}.$$

5. Fórmula cerrada para la densidad espectral del proceso.

Si  $X = \{X_t\}_{t \in \mathbb{R}} \sim \text{FOU}(\lambda_1^{(p_1)}, \lambda_2^{(p_2)}, \dots, \lambda_q^{(p_q)}, \sigma, H)$  donde  $p_1 + p_2 + \dots + p_q = p$ , entonces la densidad espectral del proceso es

$$f^{(X)}(x) = \frac{\sigma^2 \Gamma(2H + 1) \sin(H\pi) |x|^{2p-1-2H}}{2\pi \prod_{i=1}^q (\lambda_i^2 + x^2)^{p_i}}.$$

**Observación 31** *De la propiedad 2 y observando que si tomamos  $\lambda_1 \rightarrow 0$  en la igualdad de la propiedad 4, deducimos que todo  $\text{FOU}(\lambda_1, \lambda_2, \sigma, H) \rightarrow \text{FOU}(\lambda_2, \sigma, H)$  por lo que en el caso en el que  $H > 1/2$  tenemos procesos de memoria corta que tienden a procesos de memoria larga, de modo que los  $\text{FOU}(\lambda_1, \lambda_2, \sigma, H)$  para valores pequeños de  $\lambda_1$  pueden ser considerados para modelar tanto memoria larga como corta.*

### 8.1 Estimación de parámetros en un FOU( $p$ )

La estimación de los parámetros se lleva a cabo en dos pasos al igual que la de los FOU. En primer lugar se estiman  $H$  y  $\sigma$  independientemente del proceso y luego como segundo paso se estiman los parámetros  $\lambda_i$  por el método de Whittle aprovechando el uso de la fórmula cerrada para la densidad espectral del proceso.

### 8.1.1 Estimación de $H$ y $\sigma$

La estimación de  $H$  y  $\sigma$  se hace de manera idéntica a la realizada para los FOU( $\lambda, \sigma, H$ ), es decir que las fórmulas para  $\hat{H}$  y  $\hat{\sigma}$  son las mismas y el Teorema 21 con un pequeño cambio en las hipótesis sigue siendo válido.

**Teorema 32** Si  $X_\Delta, X_{2\Delta}, \dots, X_{i\Delta}, \dots, X_{n\Delta} = X_T$  es una muestra equiespaciada de un proceso  $\{X_t\}_{t \in \mathbb{R}} \sim FOU(p)$  y el filtro  $a$  es de orden  $L \geq 2$ , si  $\Delta_n = n^{-\alpha}$  para cierto  $\alpha$  tal que  $0 < \alpha < \frac{1}{2(2H-1)}$  y  $T = n\Delta_n \rightarrow +\infty$  cuando  $n \rightarrow +\infty$ , entonces, cuando  $n \rightarrow +\infty$ ,  $T \rightarrow +\infty$  y  $\Delta_n \rightarrow 0$ , tenemos

1. 
$$\left( \hat{H}, \hat{\sigma} \right) \xrightarrow{c.s.} (H, \sigma).$$
2. 
$$\sqrt{n} \left( \hat{H} - H \right) \xrightarrow{w} N(0, \Gamma_1(H, \sigma, a))$$
3. 
$$\frac{\sqrt{n}}{\log n} (\hat{\sigma} - \sigma) \xrightarrow{w} N(0, \Gamma_2(H, \sigma, a))$$

### 8.1.2 Estimación de los $\lambda_i$

El próximo teorema nos da el procedimiento para estimar los  $\lambda_i$  si conocemos los valores de  $H$ ,  $\sigma$  y tenemos observado el proceso completamente en el intervalo  $[0, T]$ , también nos asegura la consistencia y la normalidad asintótica cuando  $T \rightarrow +\infty$ .

**Teorema 33** Dado  $\{X_t\}_{t \in \mathbb{R}} \sim FOU(\lambda_1^{(p_1)}, \lambda_2^{(p_2)}, \dots, \lambda_q^{(p_q)}, \sigma, H)$  donde  $\sigma$  y  $H$  son conocidos. Supongamos además que  $\lambda^0 = (\lambda_1^0, \lambda_2^0, \dots, \lambda_q^0) \in \Lambda \subset \mathbb{R}^q$  es el verdadero valor del parámetro, con  $\lambda^0 \in \text{int}(\Lambda)$  donde  $\Lambda$  es compacto y el proceso es observado en  $[0, T]$  para algún  $T > 0$ . Definimos la función

$$U_T(\lambda) = \frac{1}{4\pi} \int_{-\infty}^{+\infty} \left( \log f^{(X)}(x, \lambda) + \frac{I_T(x)}{f^{(X)}(x, \lambda)} \right) w(x) dx$$

donde  $f^{(X)}(x, \lambda)$  es la densidad espectral del proceso,  $I_T(x)$  es el periodograma de segundo orden definido mediante

$$I_T(x) = \frac{1}{2\pi T} \left| \int_0^T X_t e^{-itx} dt \right|^2$$

y  $w(x) = \frac{|x|}{1+|x|^b}$  donde  $b > 2$ . Entonces  $\hat{\lambda}_T = \arg \min_{\lambda \in \Lambda} U_T(\lambda)$  satisface

- $\hat{\lambda}_T \xrightarrow{P} \lambda^0$  cuando  $T \rightarrow +\infty$  y
- $\sqrt{T} \left( \hat{\lambda}_T - \lambda^0 \right) \xrightarrow{D} N_q \left( 0, W_1^{-1}(\lambda^0) W_2(\lambda^0) W_1^{-1}(\lambda^0) \right)$  cuando  $T \rightarrow +\infty$

donde  $N_q(\cdot, \cdot)$  denota la distribución Gaussiana  $q$ -dimensional y las matrices  $W_1(\lambda^0)$  y  $W_2(\lambda^0)$  son definidas mediante

$$W_1(\lambda) = \left( w_{ij}^{(1)}(\lambda) \right)_{i,j=1,\dots,q} \quad \text{y} \quad W_2(\lambda) = \left( w_{ij}^{(2)}(\lambda) \right)_{i,j=1,\dots,q}$$

donde

$$\begin{aligned} w_{ij}^{(1)}(\lambda) &= \frac{1}{4\pi} \int_{-\infty}^{+\infty} w(x) \frac{\partial}{\partial \lambda_i} \log f^{(X)}(x, \lambda) \frac{\partial}{\partial \lambda_j} \log f^{(X)}(x, \lambda) dx \\ w_{ij}^{(2)}(\lambda) &= \frac{1}{4\pi} \int_{-\infty}^{+\infty} w^2(x) \frac{\partial}{\partial \lambda_i} \log f^{(X)}(x, \lambda) \frac{\partial}{\partial \lambda_j} \log f^{(X)}(x, \lambda) dx. \end{aligned}$$

**Nota.** En la práctica, es imposible tener observado un proceso en los infinitos puntos de un intervalo y tampoco son conocidos los parámetros  $H$  ni  $\sigma$ . Es posible demostrar que con un pequeño cambio en las condiciones sobre la función de pesos  $w$  y de la relación en la que tienden a infinito  $T$  y  $n$ , si sustituimos  $H$  y  $\sigma$  por  $\hat{H}$  y  $\hat{\sigma}$  y si las  $n$  observaciones están equiespaciadas en el intervalo  $[0, T]$ , entonces este método conduce a la consistencia de los estimadores de los  $\lambda_i$ . Si el proceso es observado en una cantidad finita de puntos equiespaciados del intervalo  $[0, T]$ , los valores de  $\sigma$  y  $H$  son sustituidos por sus estimadores, y la función a optimizar es una versión discretizada de  $U_T$ , es posible obtener (al menos) la consistencia del estimador considerando una función adecuada de pesos  $w$  e imponiendo una relación de tendencia a infinito entre  $T$  y el tamaño de muestra  $n$ .

La normalidad asintótica no queda garantizada al considerar el caso discretizado.

**Definición 34** Para cada  $T$  fijo definimos las funciones

$$\begin{aligned} U_T(\lambda, \sigma, H) &= \int_0^T h_T(x, \lambda, \sigma, H) dx \quad \text{y} \\ U_T^{(n)}(\lambda, \sigma, H) &= \frac{T}{n} \sum_{i=1}^n h_T^{(n)}(iT/n, \lambda, \sigma, H) \end{aligned}$$

donde  $h_T$  y  $h_T^{(n)}$  se definen como

$$\begin{aligned} h_T(x, \lambda, \sigma, H) &= \frac{1}{2\pi} \left( \log f^{(X)}(x, \lambda, \sigma, H) + \frac{I_T(x)}{f^{(X)}(x, \lambda, \sigma, H)} \right) w(x), \\ h_T^{(n)}(x, \lambda, \sigma, H) &= \frac{1}{2\pi} \left( \log f^{(X)}(x, \lambda, \sigma, H) + \frac{I_T^{(n)}(x)}{f^{(X)}(x, \lambda, \sigma, H)} \right) w(x) \end{aligned}$$

siendo

**Definición 35**

$$I_T(x) = \frac{1}{2\pi T} \left| \int_0^T e^{itx} X_t dt \right|^2 \quad \text{e} \quad I_T^{(n)}(x) = \frac{T}{2\pi} \left| \frac{1}{n} \sum_{j=1}^n e^{\frac{ijTx}{n}} X_{\frac{jT}{n}} \right|^2$$

el periodograma y el periodograma discretizado respectivamente. Definimos

$$\widehat{\lambda}_T = \arg \min_{\lambda \in \Lambda} U_T(\lambda, \sigma^0, H^0), \quad (1)$$

$$\widehat{\lambda}_T^{(n)} = \arg \min_{\lambda \in \Lambda} U_T^{(n)}(\lambda, \widehat{\sigma}, \widehat{H}) \quad (2)$$

donde  $\widehat{\sigma}$ ,  $\widehat{H}$  son los estimadores propuestos y  $\sigma^0$ ,  $H^0$  son los verdaderos valores de los parámetros.

**Teorema 36** Si  $X_{\Delta}, X_{2\Delta}, X_{3\Delta}, \dots, X_{n\Delta}$  es muestra equiespaciada en  $[0; T]$  de  $\{X_t\}_{t \in \mathbb{R}} \sim FOU(\lambda_1^{(p_1)}, \dots, \lambda_q^{(p_q)}, \sigma, H)$  donde  $p_1 + \dots + p_q = p$ . Suponemos que  $(\lambda, \sigma, H) \in \Lambda \times [\sigma_1, \sigma_2] \times [h_1, h_2]$  donde  $\sigma_1 > 0$ ,  $0 < h_1 < h_2 < 1$ , y  $\Lambda \subset \{\lambda \in \mathbb{R}^q : 0 < \lambda_1 < \lambda_2 < \dots < \lambda_q\}$  es compacto. Le llamamos  $(\lambda^0, \sigma^0, H^0) \in \overset{\circ}{\Lambda} \times (\sigma_1, \sigma_2) \times (h_1, h_2)$  al vector verdadero de parámetros. Definimos  $w(x) = \frac{|x|^\alpha}{1+|x|^b}$  donde  $a \geq 2p$  y  $b \geq a + 3$ .

Si el mínimo de  $U_T(\lambda, \sigma^0, H^0)$  se alcanza en un nico punto aislado  $\widehat{\lambda}_T$ . Si además se cumple (A) o (B) siendo

(A) Si  $1/2 < H < 5/6$  y  $T_n = n^{1-\alpha}$  donde  $\frac{3}{4} < \alpha < \min\{\frac{1}{2(2H-1)}, 1\}$ .

(B) Si  $H < 1/2$  y  $T_n = n^{1-\alpha}$  donde  $\max\{\frac{1}{H+1}, \frac{3}{4}\} < \alpha < 1$ .

Entonces

$$\lim_{n \rightarrow +\infty} \widehat{\lambda}_{T_n}^{(n)} \stackrel{P}{=} \lambda^0.$$

## 8.2 Modelación mediante procesos FOU( $p$ )

Para modelar mediante procesos FOU( $p$ ) (FOU de orden  $p$ ) se deben tomar las mismas precauciones que para modelar mediante el proceso FOU. Dado que los FOU( $p$ ) son estacionarios y centrados, a nuestro vector de datos debemos centrarlos y quitarles la componente de tendencia y estacional. Al igual que en el caso de los FOU, se supone que la serie de datos observada corresponden a observaciones equiespaciadas en el tiempo observadas en  $[0, T]$ . Un detalle a tener en cuenta, mirando los teoremas de consistencia y normalidad asintótica es que para lograr las convergencias es necesario que  $n$  y  $T$  tiendan a infinito, pero que  $T$  lo haga más lentamente que  $n$  (ya que en los teoremas se pide  $\Delta = T/n \rightarrow 0$ ). De modo que a la hora de realizar el ajuste, es conveniente elegir un valor de  $T$  como una fracción de  $n$ . La fracción a usar más conveniente depende de cada serie de datos en particular de modo que es útil tomar el valor de  $T$  como un análogo al  $p$  y  $q$  en los ARMA o ARFIMA, es decir conviene tomar distintos posibles valores de  $T$  y elegir aquel valor de  $T$  que optimice alguna medida, sea el AIC o un conjunto de predicciones.

## 9 Implementación

Como vimos, para modelar un conjunto observado de datos en el tiempo provenientes de un proceso estacionario mediante un proceso FOU( $p$ ) se necesitan valores de  $n$  y  $T$  grandes tales que  $T/n$  sea pequeño.  $n$  es el tamaño de la muestra y se supone que las observaciones son en determinado intervalo de tiempo  $[0, T]$ . La unidad de medida del tiempo podemos elegirla arbitrariamente.

Podemos tomar ventaja de este hecho, eligiendo el valor de  $T$  que optimice alguna medida de bondad del ajuste.

Por ejemplo podemos elegir aquel valor de  $T$  que minimice el MAE o RMSE de un determinado porcentaje de predicciones, o elegir aquel  $T$  que optimice el AIC o BIC del modelo. En todas las definiciones que siguen  $m$  representa la cantidad de predicciones.

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (X_{n-m+i} - \hat{X}_{n-m+i})^2};$$

$$MAE = \frac{1}{m} \sum_{i=1}^m |X_{n-m+i} - \hat{X}_{n-m+i}|$$

el índice de Willmott

$$W_2 = 1 - \frac{\sum_{i=1}^m (X_{n-m+i} - \hat{X}_{n-m+i})^2}{\sum_{i=1}^m (|\hat{X}_{n-m+i} - \bar{X}(m)| + |X_{n-m+i} - \bar{X}(m)|)^2}$$

y el índice  $L^1$  de Willmott

$$W_1 = 1 - \frac{\sum_{i=1}^m |X_{n-m+i} - \hat{X}_{n-m+i}|}{\sum_{i=1}^m (|\hat{X}_{n-m+i} - \bar{X}(m)| + |X_{n-m+i} - \bar{X}(m)|)};$$

donde  $\bar{X}(m) := \frac{1}{m} \sum_{i=1}^m X_{n-m+i}$ , y  $X_1, X_2, \dots, X_n$  (o  $X_{T/n}, X_{2T/n}, \dots, X_T$ ) son las observaciones reales, mientras que  $\hat{X}_i$  son las predicciones realizadas por el modelo.

## 10 Ejercicios

1. Probar que todo proceso FOU( $\lambda, \sigma, H$ ) es de memoria corta cuando  $H \leq 1/2$  y es de memoria larga cuando  $H > 1/2$ .
2. Verificar que todo proceso FOU( $\lambda, \sigma, H$ ) tiene varianza dada por la fórmula  $\mathbb{V}(X_t) = \frac{\sigma^2 \Gamma(2H+1)}{2\lambda^{2H}}$ .
3. Probar que la fórmula que da el estimador de  $\lambda$  en un FOU( $\lambda, \sigma, H$ ), surge de sustituir  $\mathbb{V}(X_t)$  por  $\frac{1}{n} \sum_{i=1}^n X_i^2$ ,  $H$  por  $\hat{H}$  y  $\sigma$  por  $\hat{\sigma}$  en la fórmula de varianza obtenida en el ejercicio anterior y despejar  $\lambda$ .

4. Dentro del paquete arfima se encuentra la serie de datos llamada “tmpyr” que son las temperaturas medias anuales en Inglaterra entre 1659 y 1976.
  - (a) Quitarle la tendencia y centrar los datos.
  - (b) Ajustarles el mejor ARFIMA posible para valores de  $p$  y  $q$  variando entre 0 y 2 de acuerdo al criterio de AIC. Investigar si el modelo elegido ajusta bien.
  - (c) Ajustarles el mejor ARMA posible para valores de  $p$  y  $q$  variando entre 0 y 2 de acuerdo al criterio de AIC. Investigar si el modelo elegido ajusta bien.
  - (d) Hallar las predicciones para los años 1977 al 1985 para ambos modelos ajustados en las partes anteriores junto con sus intervalos de predicción al 95% y al 80%.
  
5. Trabajamos con la misma serie de datos del ejercicio anterior.
  - (a) Ajustarles el mejor ARFIMA posible para valores de  $p$  y  $q$  variando entre 0 y 2 de acuerdo al criterio de minimizar la suma de los valores absolutos de los errores de predicción a un paso correspondientes a los años 1946 a 1976. Investigar si el modelo elegido ajusta bien.
  - (b) Ajustarles el mejor ARMA posible para valores de  $p$  y  $q$  variando entre 0 y 2 de acuerdo al criterio de minimizar la suma de los valores absolutos de los errores de predicción a un paso correspondientes a los años 1946 a 1976. Investigar si el modelo elegido ajusta bien.
  - (c) Graficar los valores observados, y los predichos por ambos modelos para los años 1946 a 1976.
  
6. En este ejercicio utilizamos la serie de datos del archivo seriesA.R. Este conjunto de datos es una serie de lecturas de concentración de un proceso químico tomadas cada dos horas.
  - (a) Ajustarle un modelo FOU a los datos luego de quitada la tendencia (si es que la tiene) y de centrarlos, indicar el valor de los parámetros estimados y realizar las predicciones a un paso para las últimas 30 observaciones. Graficar las últimas 30 observaciones junto con sus predicciones.
  - (b) Realizar lo mismo que en el punto anterior pero ajustando modelos  $\text{FOU}(\lambda_1, \lambda_2)$ ,  $\text{FOU}(\lambda^{(2)})$  y  $\text{FOU}(\lambda^{(3)})$ . Considerar para estos casos el valor  $T = 10, 20, 50$  y  $100$ .
  - (c) De acuerdo a lo calculado en las partes anteriores ¿cuál de todos los modelos ajustados resultó mejor?