# Modelos ARIMA

Gustavo A. Caffaro

Instituto Tecnológico de Santo Domingo

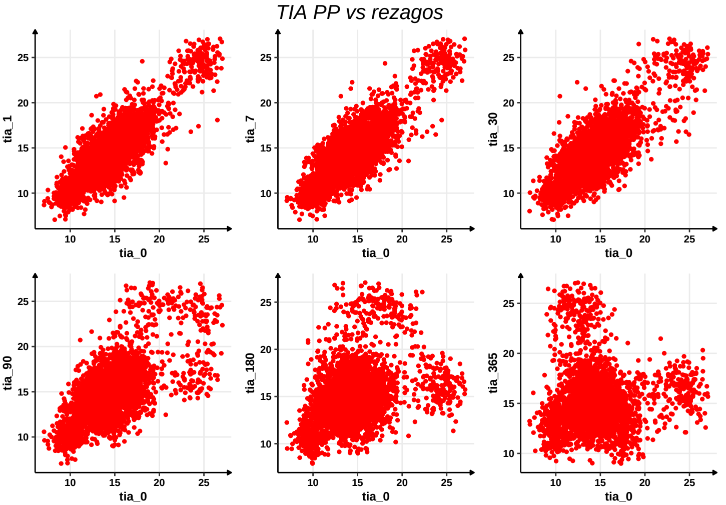
Nov 2021 - Ene 2022

# Introducción

Contenido:

- Modelos autoregresivos (AR)
- Modelos de media móvil (MA)
- Series integradas de orden d, I(d)
- Modelos ARIMA

# Dependencia temporal



# Modelo AR(1)

Un modelo autoregresivo es aquel que modela series que dependen de valores pasados.

El modelo autoregresivo de orden 1, escrito como AR(1), es aquel descrito por:

$$x_t = \alpha + \phi x_{t-1} + w_t,$$

donde $w_t \overset{iid}{\sim} N(0, \sigma^2)$, $E(w_t | x_{t-1}, x_{t-2}, \dots) = 0$ y $\alpha, \phi$ son los parámetros del modelo.

El parámetro $\phi$ se conoce como el término autoregresivo.

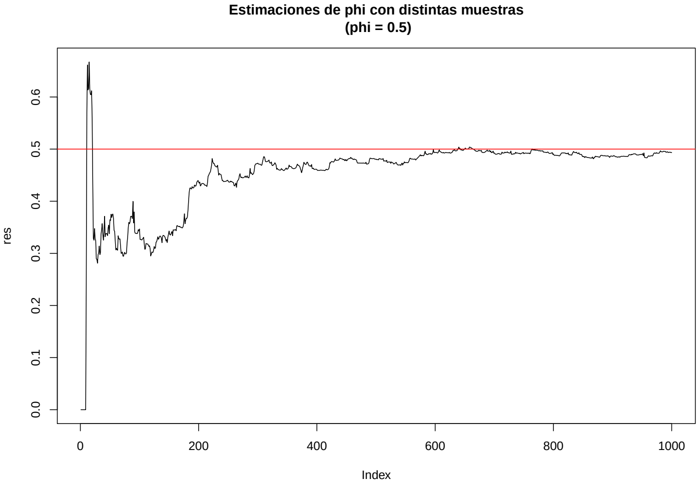
# El término autoregresivo, $\phi$

Recordemos de la clase pasada que dependiendo del valor del término autoregresivo, tendremos una serie explosiva, estable, o persistente:

1. $|\phi| < 1$: $\phi^t \to 0$ cuando $t \to \infty$, por lo que el efecto de un choque disminuye conforme el tiempo pasa. Si esta condición se cumple decimos que el proceso es **estable**.
2. $|\phi| = 1$: $\phi^t = 1, \forall t$, por lo que el efecto es persistente y su varianza crece con el tiempo (caminata aleatoria).
3. $\phi > 1$: los choques se vuelven más importantes con el tiempo.

# El término autoregresivo, $\phi$

# Ley de grandes números



# Ley de grandes números

```
## Simulacion AR(1) y ley de los grandes numeros

n <- 100

simular_ar <- function(phi,n){
  set.seed(1e3)
  w <- rnorm(n)
  x <- numeric(n)

  for (i in 2:(n)){
    x[i] <- phi*x[i-1]+w[i]
  }
  coefficients(lm(x ~ lag(x)-1))
}

simular_ar(0.5,10)
simular_ar(0.5,50)
simular_ar(0.5,100)
simular_ar(0.5,1000)

res <- numeric(1e3)

for (i in 10:length(res)){
  res[i] <- simular_ar(0.5,i)
}
plot(res,type="l")
abline(a=0.5,b = 0, col="red")
res[1e3]
```

# Propiedades del modelo AR(1)

La media, varianza, y FAC de un proceso autoregresivo estable de orden 1 vienen dados por:

- Media

$$E(x_t) = \mu = \frac{\alpha}{1-\phi}$$

- Varianza

$$Var(x_t) = \frac{\sigma_w^2}{1-\phi^2}$$

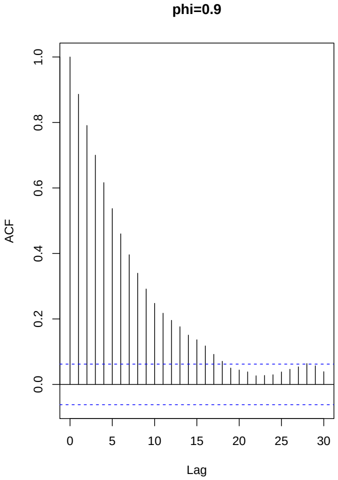
- Correlación

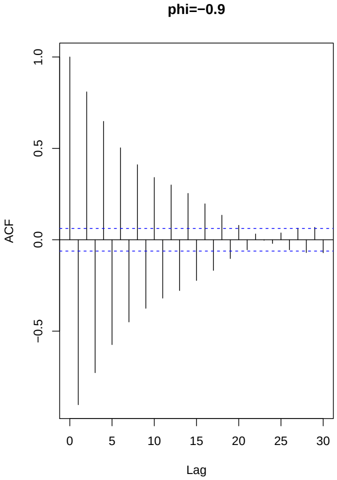
$$\rho(h) = \phi^h$$

# Función de autocorrelación (FAC)

- Cuando el término autoregresivo $\phi$ es positivo, la función de autocorrelación del modelo AR(1) decae de manera exponencial conforme aumenta el horizonte $h$.
- Cuando $\phi$ es negativo, la FAC también decae a cero de manera exponencial, pero el signo de las autocorrelaciones fluctuan entre positivo y negativo.

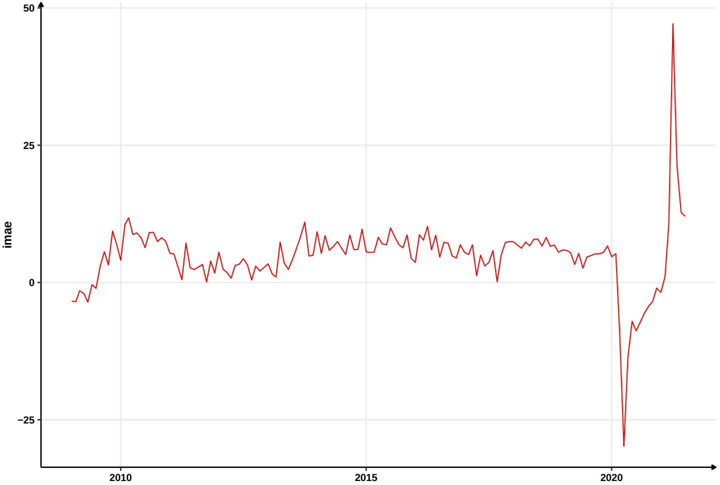
# Función de autocorrelación (FAC)



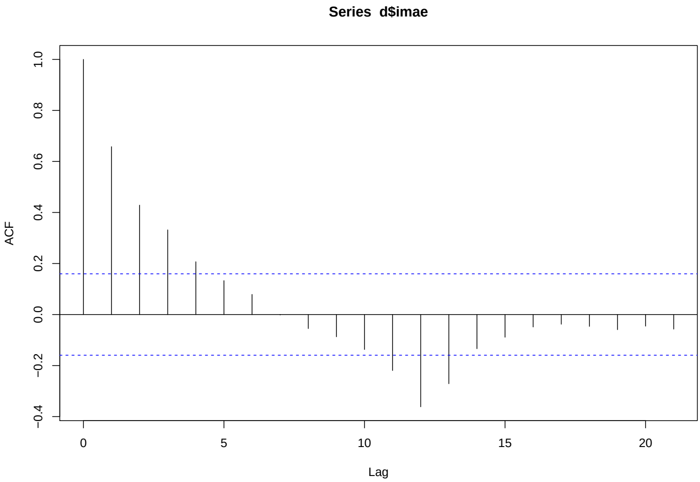



# Ejercicio

Tomen la serie del crecimiento interanual del IMAE y verifiquen que podríamos modelar la serie como un proceso AR(1).



# IMAE - FAC



# IMAE

```
##
## Call:
## lm(formula = d$imae ~ lag(d$imae))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -25.234  -1.674   0.065   1.526  38.415
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.67534    0.48091   3.484 0.000651 ***
## lag(d$imae)  0.66394    0.06135  10.822  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.726 on 148 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.4417, Adjusted R-squared:  0.438
## F-statistic: 117.1 on 1 and 148 DF,  p-value: < 2.2e-16
```

# Modelo AR(p)

Podríamos describir una serie que depende de $p$ rezagos (lags), generalizando el modelo AR(1) para incluir estos rezagos.

Un modelo autoregresivo de orden $p$, también conocido como AR($p$), es aquel descrito por:

$$x_t = \alpha + \sum_{i=1}^{p} \phi_i x_{t-i} + w_t,$$

donde $\alpha, \phi_1, \ldots, \phi_p$ son los parámetros del modelo y $w_t$ ruido blanco.

# Modelo MA(1)

Un modelo de media móvil incluye un rezago de los errores $w_t$, multiplicado por un coeficiente.

Por lo tanto, el modelo MA(1) se escribe como:

$$x_t = \mu + w_t + \theta w_{t-1},$$

donde $w_t \overset{iid}{\sim} N(0, \sigma^2)$.

Fíjense que por definición, $E(w_t | w_{t-1}, w_{t-2}, \dots) = 0$.

# Propiedades del modelo AR(1)

La media, varianza, y FAC de un proceso MA(1) vienen dados por:

- Media

$$E(x_t) = \mu$$

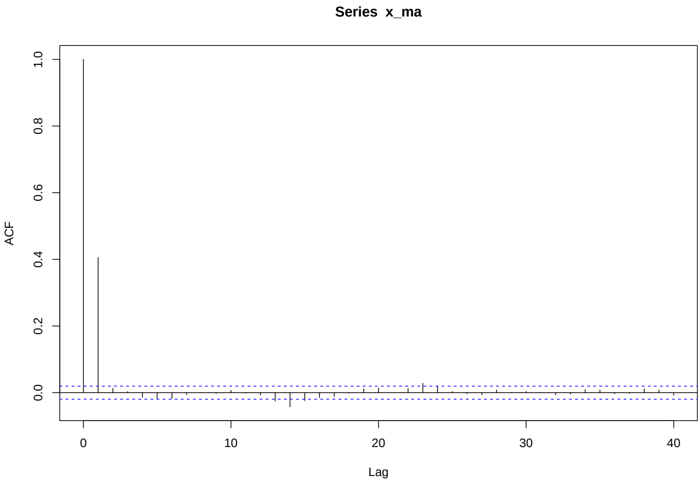
- Varianza

$$Var(x_t) = \sigma_w^2(1 + \theta^2)$$

- Autocorrelación

$$\rho(h) = \frac{\theta}{1 + \theta^2}, \text{ si } h = 1 \text{ y } \rho(h) = 0, \text{ si } h \geq 2$$

# Función de autocorrelación



# Modelo MA(q)

El modelo de media móvil de orden q, MA(q), viene dado por:

$$x_t = \mu + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q}$$
$$x_t - \mu = +w_t + \sum_{i=1}^{q} \theta_i w_{t-i},$$

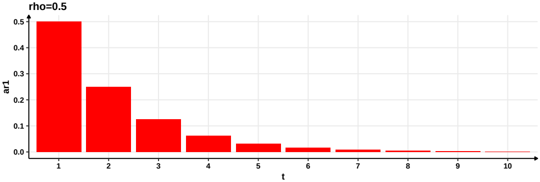
donde $w_t \overset{iid}{\sim} N(0, \sigma^2)$.

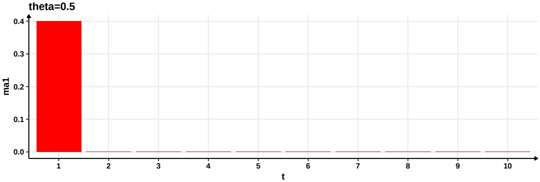
Usualmente, los programas estadísticos añaden la restricción de que los $|\theta_i| \leq 1$.

# Identificando un modelo AR vs MA

Recapitulando:

- La FAC de los modelos autoregresivos decae de manera exponencial, oscilando si algún $\phi_i$ es negativo.
- La FAC de los modelos de media móvil mueren a partir del rezago $q$.



# El modelo AR(1) como MA($\infty$)

Podríamos reescribir un modelo AR(1) en un MA de orden infinito, MA($\infty$):

$$x_t - \mu = w_t + \theta w_{t-1} + \theta^2 w_{t-2} + \cdots = \sum_{j=0}^{\infty} \theta^j w_{t-j}.$$

Esta suma de ruidos blancos es conocida como la **representación causal** de un modelo AR(1).

# Orden de integración

El **orden de integración** de una serie nos dice cuántas veces dicha serie debe ser diferenciada para que sea estacionaria.

Por lo tanto, si $x_t$ es integrada de orden 0, I(0), entonces es estacionaria (no necesita diferenciación).

Ej.:

En clases anteriores vimos que el RW no es estacionario. Si restamos cada valor del RW con el anterior, nos quedamos con una serie de ruidos blancos, que claramente son estacionarios:

$$\nabla x_t = x_t - x_{t-1} = \sum_{i=0}^{t} w_i - \sum_{i=0}^{t-1} w_i = w_t \overset{iid}{\sim} N(0, \sigma^2)$$

Por lo tanto, el RW es integrado de orden 1, I(1).

# Raíz unitaria

Como hemos visto a lo largo de esta clase, muchas series económicas y financieras suelen ser no estacionarias.

Sólo con series estacionarias podemos usar los modelos de la familia AR o MA.

# Raíz unitaria

Consideremos el proceso autoregresivo de orden $p$:

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \epsilon_t$$

Si $m = 1$ es una raíz de la ecuación característica

$$m^p - m^{p-1} - \phi_1 m^{p-2} - \cdots - \phi_p = 0,$$

entonces el proceso $x_t$ tiene raíz unitaria (es integrada de orden 1).

# Raíz unitaria

¿Por qué le llamamos raíz unitaria?

Utilicemos el operador rezago para reescribir el RW como

$$x_t = Lx_t + w_t,$$

por tanto,

$$(1 - L)x_t = w_t.$$

La ecuación

$$1 - L = 0$$

tiene la raíz

$$L = 1,$$

por lo tanto, tiene raíz unitaria.

# Múltiples raíces unitarias

Un serie de tiempo potencialmente podría tener varias raíces unitarias.

Por ejemplo, consideremos la siguiente serie, con dos raíces unitarias:

$$
\begin{aligned}
&(1-L)(1-L)x_t = w_t \\
&\Rightarrow (1-2L+L^2)x_t = w_t \\
&\Rightarrow x_t = 2Lx_t - L^2 x_t w_t \\
&\Rightarrow x_t = 2x_{t-1} - x_{t-2} + w_t,
\end{aligned}
$$

por lo que el proceso AR(2)

$$x_t = 2x_{t-1} - x_{t-2} + w_t$$

tiene raíz unitaria (de multiplicidad 2) y por tanto no es estacionario.

# Pruebas de raíz unitaria

Como en la práctica muchas de las series con las que trabajamos no son I(0), es apropiado tener una forma de verificar estadísticamente si una serie es I(0) o no.

Verificamos la estacionariedad de una serie mediante dos tipos de tests:

- Tests de raíz unitaria
- Tests de estacionariedad

# Test de Dickey-Fuller

El test de Dickey-Fuller intenta verificar si el parámetro $\phi = 1$ en una serie AR(1):

$$x_t = \alpha + \phi x_{t-1} + w_t.$$

Para hacer esto, le restamos $x_{t-1}$ a la ecuación anterior para obtener

$$\Delta x_t = \alpha + \theta x_{t-1} + w_t$$

y establecer $H_0 : \theta = 0$ y $H_1 : \theta < 1$. Rechazamos $H_0$ cuando el t-estadístico es menor que el valor crítico.

# Test de Dickey-Fuller Aumentado

El test de DF, aunque sencillo, es un poco restrictivo. Por ejemplo, el test de DF asume que no autocorrelación de los errores (vimos en la clase pasada que este supuesto se incumple con regularidad).

Por esta razón, en la práctica utilizamos una versión más sofisticada que incluye más rezagos de $x_t$, el test de Dickey-Fuller Aumentado (ADF):

$$\Delta x_t = \mu + \alpha t + \gamma x_{t-1} + \sum_{i=1}^{p} \beta_i \Delta x_{t-i+1} + w_t,$$

donde

$$\gamma = \left( \sum_{i=1}^{p} \phi_i - 1 \right) \quad \text{y} \quad \beta_i = - \sum_{j=1}^{p} \phi_j$$

El parámetro de interés es el $\gamma$.

# Test de Dickey-Fuller Aumentado

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1345 -0.6706  0.0006  0.6751  4.3198
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## z.lag.1    -2.177e-05  2.197e-05  -0.991    0.322
## z.diff.lag -2.281e-03  3.162e-03  -0.721    0.471
##
## Residual standard error: 0.9997 on 99996 degrees of freedom
## Multiple R-squared:  1.508e-05,  Adjusted R-squared:  -4.924e-06
## F-statistic: 0.7538 on 2 and 99996 DF,  p-value: 0.4706
##
##
## Value of test-statistic is: -0.991
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau1 -2.58 -1.95 -1.62
```

# Test de Dickey-Fuller Aumentado

Lo importante para fijarse en la diapositiva anterior es:

- Que la estimación de z.lag.1 sea muy cercana a cero.
- Si el valor del test-statistic es menor al valor crítico deseado.

En el ejemplo anterior, el valor del test-statistic es mayor que todos los valores críticos:

$$-0.991 > -1.62 > -1.95 > -2.58,$$

por lo que no podemos rechazar la hipótesis de que la serie es no estacionaria:

$$\gamma = \phi - 1 = 0 \quad \Longrightarrow \quad \phi = 1$$

# Test de Dickey-Fuller Aumentado

## Código R:

```
library(urca)

set.seed(123)
rw <- cumsum(rnorm(1e5)) # RW

prueba_adf <- ur.df(rw,type="none",lags = 1) # un solo lag y sin intercepto ni tendencia

summary(prueba_adf)
```

# Test de ADF - Tips

- En el caso de que la serie tenga tendencia y/o drift, se puede especificar el parámetro `type="trend` o `type="drift"` en la función `ur.df`.
- Para saber cuál especificación escoger:
    - AIC
    - BIC
    - Del modelo general al específico (comienza con un alto $p$, y estimar hacia atrás)
        - Schwert (1989) sugiere usar $p_{max} = 12 \cdot (T/100)^{1/4}$
- Especificar el modelo correctamente es fundamental.

# Test de Phillips-Perron (PP)

Este test corrige cualquier autocorrelación serial y heteroscedasticidad en los errores, modificando los estadísticos directamente.

$$x_t = \mu + \delta t + \gamma x_{t-1} + u_t,$$

donde $u_t$ es un error I(0) pero puede ser heteroscedástico y autocorrelacionado.

- $H_0 : \gamma = 1$
- Una gran ventaja de el test PP es que no necesitamos escoger $p$.

# Test de Phillips-Perron (PP)

```r
prueba_pp <- ur.pp(rw, type="Z-tau")
summary(prueba_pp)
```

```
###################################
# Phillips-Perron Unit Root Test #
###################################

Test regression with intercept


Call:
lm(formula = y ~ y.l1)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.1346 -0.6712 -0.0003  0.6743  4.3185 

Coefficients:
             Estimate Std. Error   t value Pr(>|t|)    
(Intercept) 6.940e-04  3.175e-03     0.219    0.827    
y.l1        1.000e+00  2.207e-05 45311.448   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9997 on 99997 degrees of freedom
Multiple R-squared:      1,	Adjusted R-squared:      1 
F-statistic: 2.053e+09 on 1 and 99997 DF,  p-value: < 2.2e-16


Value of test-statistic, type: Z-tau  is: -0.9426 

         aux. Z statistics
Z-tau-mu            0.2277

Critical values for Z statistics: 
                     1pct      5pct     10pct
critical values -3.43356 -2.862127 -2.567114
```

# Tests de estacionariedad

- Los tests de ADF y PP son tests de raíz unitaria, por lo que su hipótesis nula corresponde con que el proceso es I(1).
- Los tests de estacionariedad, por otro lado, tienen como $H_0$ que la serie es I(0).

# Test de Kwiatkowski-Phillips-Schidt-Shin (KPSS)

- $H_0$: $x_t$ es estacionaria (en niveles/tendencia).
- Formulación:

$$x_t = \beta' \boldsymbol{D_t} + \mu_t + w_t$$

$$\mu_t = \mu_{t-1} + \epsilon_t, \ \epsilon \sim WN(0, \sigma_\epsilon^2)$$

- $\boldsymbol{D_t}$ contiene componentes determinísticos (constante o constante más tendencia)
- $u_t$ es I(0) y puede ser heteroscedástico.
- $\mu_t$ es un RW
- $H_0 : \sigma_\epsilon^2 = 0$
- El test se lleva cabo vía simulaciones

# Test de KPSS (sin tendencia)

```
library(tseries)

kpss.test(rw,null="Level")
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  rw
## KPSS Level = 228.53, Truncation lag parameter = 22, p-value = 0.01
```

- Como p-value=0.01, rechazamos la hipótesis nula y concluimos que la serie no es estacionaria en niveles.

# Test de KPSS (con tendencia)

```
library(tseries)
x4 <- 1:100 + rnorm(100) # serie estacionaria alrededor de la tendencia
kpss.test(x4,null="Trend")
```

```
##
##  KPSS Test for Trend Stationarity
##
## data:  x4
## KPSS Trend = 0.091487, Truncation lag parameter = 4, p-value = 0.1
```

- Rechazamos $H_0$ con 10 % de significancia (no es trend stationary), pero no podemos rechazarla con 5 % de significancia (trend stationary).
- Si el p-value > nivel de significancia hay estacionariedad (en niveles/tendencial).

# Modelo ARMA

Si combinamos un modelo AR(p) con uno MA(q) obtenemos el modelo ARMA(p,q):

$$x_t = \alpha + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q}$$

el cual, escrito con sumatorias y pasando los términos autoregresivos a la parte izquierda de la ecuación, obtenemos:

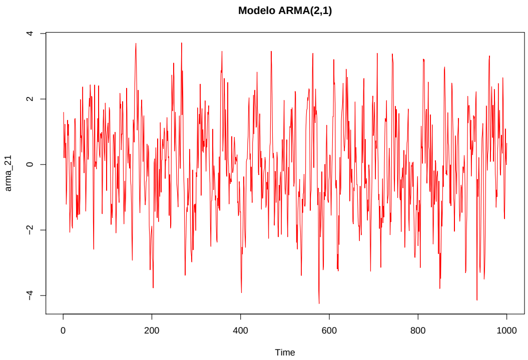
$$x_t - \sum_{i=1}^{p} \phi_i x_{t-i} = \alpha + \sum_{i=1}^{q} \theta_i w_{t-i}.$$

Si reescribimos utilizando el operador de rezago:

$$(1 - \sum_{i=1}^{p} \phi_i L^i) x_t = \alpha + (1 + \sum_{i=1}^{q} \theta_i L^i) w_t$$

# Modelo ARMA

```r
set.seed(1e3)
modelo_arma <- list(ar=c(1/3,1/4),ma=c(1/2)) # ARMA(2,1)
arma_21 <- arima.sim(model=modelo_arma,n=1e3) # simular serie
plot(arma_21, col="red", main="Modelo ARMA(2,1)")
```



# ARIMA

Asumamos que el polinomio autoregresivo de orden $p'$ tiene raíz unitaria de multiplicidad $d$. Entonces podemos reescribirlo como:

$$\left(1 - \sum_{i=1}^{p'} \phi_i L^i\right) = \left(1 - \sum_{i=1}^{p'-d} \beta_i L^i\right)(1 - L)^d$$

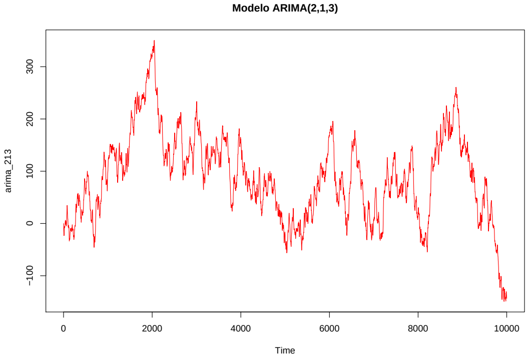
Por lo tanto, un **modelo ARIMA**$(p, d, q)$, con $p = p' - d$ se escribe como

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right)(1 - L)^d x_t = \left(1 + \sum_{i=1}^{q} \theta_i L_i\right) w_t$$

El modelo ARIMA expande el ARMA para incluir series con ordenes de integración superiores a cero.

# ARIMA

```r
set.seed(123)
modelo_arima <- list(order=c(2,1,3),ar=c(1/3,1/4),ma=c(1/2,1/2,2/5)) # ARIMA(2,1,3)
arima_213 <- arima.sim(model=modelo_arima,n=1e4) # simular serie
plot(arima_213, col="red", main="Modelo ARIMA(2,1,3)")
```



# auto.arima

Una función muy útil para automáticamente detectar el órden del ARIMA es `auto.arima`, del paquete `forecast`.

```
library(forecast)
auto.arima(arima_213) # estimamos el orden a partir de las simulaciones
```

```
## Series: arima_213
## ARIMA(2,1,3)
##
## Coefficients:
##          ar1     ar2     ma1     ma2     ma3
##       0.2804  0.2896  0.5371  0.4991  0.4103
## s.e.  0.0349  0.0298  0.0330  0.0113  0.0148
##
## sigma^2 estimated as 0.9979:  log likelihood=-14177.4
## AIC=28366.8   AICc=28366.81   BIC=28410.06
```

# Ejercicio: tasas de interés

```
# Tasas interes Bonos 10 anios EEUU
i_10y <- read.csv("http://research.stlouisfed.org/fred2/series/DTB3/downloaddata/DTB3.csv")

i_10y <- i_10y %>%
  mutate(DATE=as.Date(DATE),VALUE=as.numeric(VALUE)) %>%
  filter(complete.cases(.))

i_10y %>%
  ggplot(aes(x=DATE,y=VALUE)) + geom_line(color="red")

arima_model <- auto.arima(i_10y$VALUE) # seleccionar el mejor modelo arima
arima_order <- arimaorder(arima_model) # orden del arima (p,d,q)

arima_model

test_df <- ur.df(i_10y$VALUE, lags = arima_order[1]) # test de adf
test_pp <- ur.pp(i_10y$VALUE,type="Z-tau", use.lag = arima_order[1]) # test de pp
test_kpss <- kpss.test(rw,null="Trend") # test kpss

summary(test_df)
summary(test_pp)
test_kpss
```