

Series de tiempo

Eloy Alvarado Narváez

Instituto de Estadística
Universidad de Valparaíso



Introducción

Una serie de tiempo es una colección de observaciones realizadas secuencialmente en el tiempo.

Este tipo de datos existen en muchas disciplinas por ejemplo:

- Series de tiempo económicas
- Series de tiempo demográficas
- Procesos binarios
- Procesos puntuales

Una de las características principales de las series de tiempo es el hecho que observaciones sucesivas no son usualmente independientes, por lo que el análisis debe tomar en cuenta el orden temporal de las observaciones

Objetivos del análisis de series de tiempo

Existen distintos objetivos posibles del análisis de tiempo. Estos objetivos pueden ser clasificados en 4 categorías:

- Describir
- Explicar
- Predecir
- Controlar

Modelo aditivo de componentes de Series de tiempo

Dada una serie $Y_t, t = 1, \dots, T$, el modelo aditivo de componentes consiste en asumir que Y_t se puede descomponer en 3 partes.

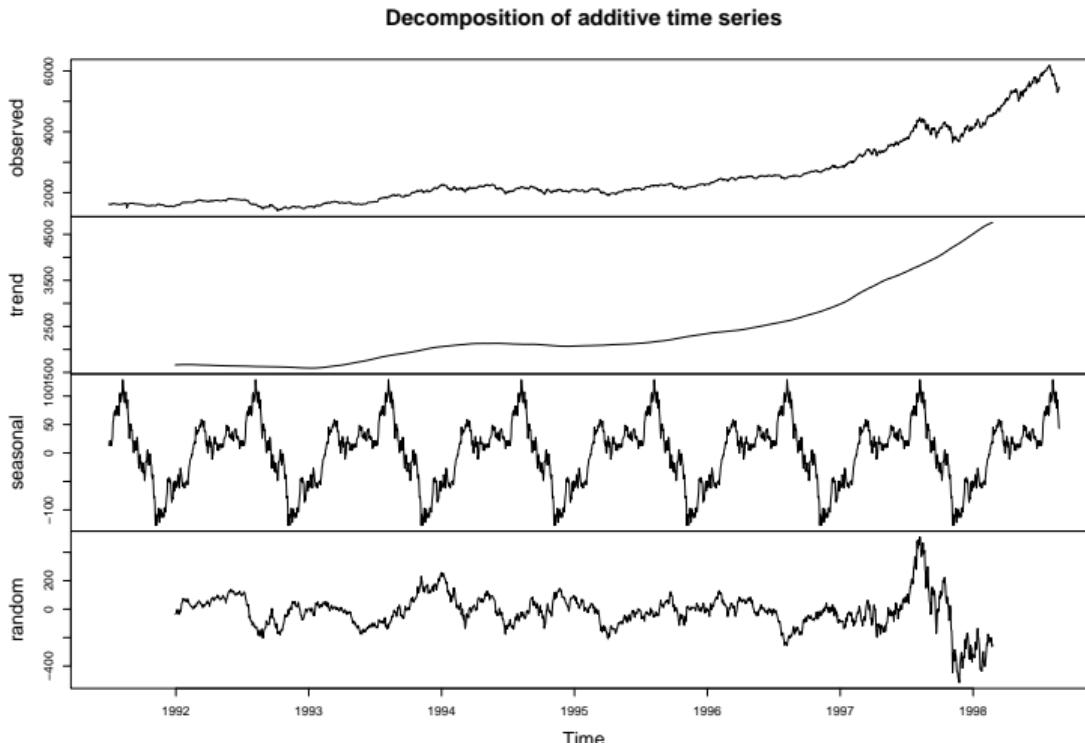
$$Y_t = T_t + S_t + \varepsilon_t$$

Donde T_t es la tendencia, S_t es la componente estacional y ε_t es la componente de errores.

Las primeras dos componentes son funciones determinísticas de t , por lo que su evolución es perfectamente predecible.

En algunos casos, la componente T_t también puede ser una componente estacional, pero de baja frecuencia o, equivalentemente, una componente de período muy grande. Por ejemplo, una serie diaria, S_t puede tener período 30 días, y T_t período 360 días.

```
tsData <- EuStockMarkets[, 1]
decomposedRes2 <- decompose(tsData, type="additive")
plot(decomposedRes2)
```

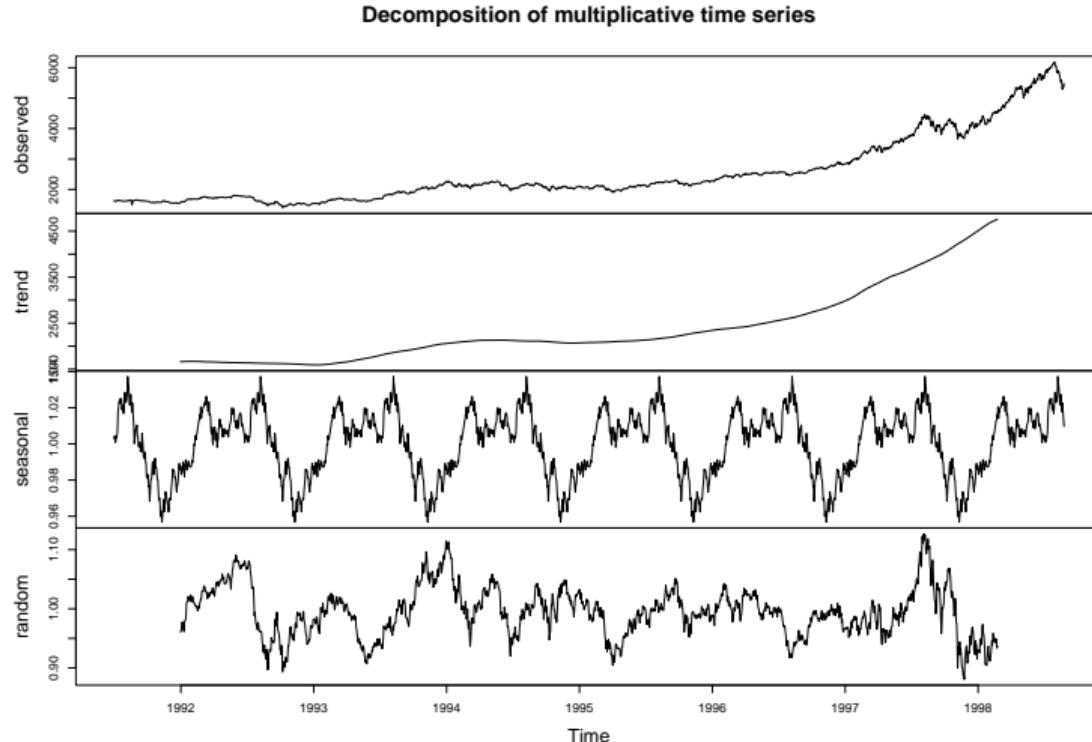


Modelo multiplicativo de componentes de Series de tiempo

El modelo multiplicativo consiste en asumir que Y_t se puede descomponer en tres partes:

$$Y_t = T_t S_t \exp \varepsilon_t$$

```
decomposedRes1 <- decompose(tsData, type="mult")
plot(decomposedRes1)
```



El análisis general consiste en modelar y estimar T_t y S_t para luego extraerlas de Y_t para obtener

$$\hat{\varepsilon}_t = Y_t - \hat{T}_t - \hat{S}_t$$

La serie $\hat{\varepsilon}_t$ se modela y estima para finalmente reconstruir Y_t ,

$$\hat{Y}_t = \hat{T}_t + \hat{S}_t + \hat{\varepsilon}_t$$

y poder realizar el pronóstico

$$\hat{Y}_{T+h} = \hat{T}_{T+h} + \hat{S}_{T+h} + \hat{\varepsilon}_{T+h}$$

utilizando la información disponible Y_1, \dots, Y_T con $h = 1, 2, \dots, m$. Sin embargo, puede suceder que la serie $\hat{\varepsilon}_t$ sea no correlacionada, es decir, $\text{Corr}(\hat{\varepsilon}_t, \hat{\varepsilon}_{t+s}) = 0$, para $s \neq 0$. En este caso $\hat{\varepsilon}_{T+h} = 0, \forall h \geq 0$

Tendencia

Tendencia: Se define como una función T_t que describe la evolución lenta y a largo plazo del nivel medio de la serie. La función T_t depende de parámetros que deben estimarse.

Algunos posibles modelos para T_t

1. Lineal

$$T_t = \beta_0 + \beta_1 t$$

2. Cuadrático

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2$$

3. Cúbico

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3$$

4. Exponencial

$$T_t = \exp(\beta_0 + \beta_1 t)$$

5. Logístico

$$T_t = \frac{\beta_2}{1 + \beta_1 \exp(-\beta_0 t)}$$

En la tendencia cuadrática se puede observar:

1. Si $\beta_1, \beta_2 > 0$, T_t es monótona creciente
2. Si $\beta_1, \beta_2 < 0$, T_t es monótona decreciente
3. Si $\beta_1 > 0$ y $\beta_2 < 0$, T_t es cóncava
4. Si $\beta_1 < 0$ y $\beta_2 > 0$, T_t es convexa.

Modelo log-lineal

El modelo logarítmico lineal o log-lineal se define como:

$$\log Y_t = \beta_0 + \beta_1 t + \varepsilon_t$$

Corresponde a un modelo con **tendencia lineal para el logaritmo de Y_t** . En la ecuación anterior, al tomar exponencial se tiene $Y_t = \exp(\beta_0 + \beta_1 t + \varepsilon_t)$ que es similar al modelo con tendencia exponencial, $Y_t = \exp(\beta_0 + \beta_1 t)$. Sin embargo, son modelos diferentes y se estiman por métodos diferentes.

Estimación de la tendencia

Usualmente, la expresión “suavizar la serie” hace referencia a la extracción de la tendencia de una serie, y ambas equivalen a la estimación de la tendencia.

Para la estimación de los parámetros $\beta = (\beta_0, \beta_1, \beta_2)'$ en los modelos lineal, cuadrático, cúbico y log-lineal se utiliza el **método de mínimos cuadrados ordinarios**. Es decir, el valor $\hat{\beta}$ es el valor en el cual $G(\beta) = \sum_{t=1}^T (Y_t - T_t(\beta))^2$ toma el valor mínimo.

$$\hat{\beta} = \arg \min_{\beta} G(\beta)$$

Para los modelos exponencial y logístico, se usa el método de mínimos cuadrados no lineales, que también minimiza la suma de errores cuadrados $G(\beta) = \sum_{t=1}^T (Y_t - T_t(\beta))^2$, pero $T_t(\beta)$ es una función no lineal de β .

El modelo log-lineal es equivalente, algebráicamente, a

$$Y_t = \exp(\beta_0 + \beta_1 t + \varepsilon_t)$$

Sin embargo, este último modelo es no lineal y no coincide con el modelo exponencial. Es posible estimar los parámetros en este caso, pero estas estimaciones no necesariamente serán iguales.

Aunque la serie tenga una componente estacional S_t , $Y_t = T_t + S_t + \varepsilon_t$, solamente consideramos un modelo de regresión entre Y_t y T_t , tal que $Y_t = T_t + \eta_t$, donde η_t es el término de error, de forma que $\eta_t = S_t + \varepsilon_t$. Por ejemplo:

1. En el caso lineal $\beta = (\beta_0, \beta_1)'$, con $T_t = \beta_0 + \beta_1 t$ se ajusta el modelo de regresión lineal: $Y_t = \beta_0 + \beta_1 t + \eta_t$
2. En el caso cuadrático $\beta = (\beta_0, \beta_1, \beta_2)'$, con $T_t = \beta_0 + \beta_1 t + \beta_2 t^2$ se ajusta el modelo de regresión lineal $Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \eta_t$. Notar que en este caso hay que definir la variable explicativa adicional t^2 .

Pronóstico con base en la tendencia

Supongamos que la serie con tendencia $Y_t = T_t + \eta_t, t = 1, \dots, T$ con (η_t) una sucesión $iid(0, \sigma^2)$. Los pronósticos de Y_t en los tiempos $T+1, T+2, \dots, T+h, h \geq 1$ se definen como

$$\hat{Y}_{T+j} = \hat{T}_{T+j}, j = 1, \dots, h$$

donde \hat{T}_t es la función estimada de la tendencia. Por ejemplo, en el modelo lineal

$$Y_t = \beta_0 + \beta_1 t + \varepsilon_t$$

al reemplazar t por $T+h$ se obtiene

$$Y_{T+h} = \hat{\beta}_0 + \hat{\beta}_1(T+h) + \hat{\varepsilon}_{T+h}$$

Pero el pronóstico $\hat{\eta}_{T+h}$, puede ser o no cero.

La definición general de pronóstico para una serie $Y_t, t \in \mathbb{Z}$ con base en la información Y_1, \dots, Y_T es una esperanza condicional, como sigue

$$\hat{Y}_{T+j} = \mathbb{E}(Y_{T+j} | Y_1, \dots, Y_T), \quad j = 1, \dots, h$$

Modelado y pronóstico de series estacionales

Recordar que definimos la descomposición aditiva de una serie de tiempo Y_t como

$$Y_t = T_t + S_t + \varepsilon_t$$

Siendo T_t la tendencia, ε_t la componente aleatoria y S_t la componente estacional.

Componente Estacional

La componente estacional S_t se define como una función no aleatoria, periódica de período s . Los valores de S_t para $t = 1, \dots, s$ se denominan el *patrón estacional*. El período estacional s es el número mínimo de períodos que tarda el patrón estacional en repetirse.

Propiedades de S_t

1. S_t es una función periódica con período s , $S_{t+s} = S_t$ para $t = 1, 2, \dots$. Por lo que sólo que requiere definir S_t en los primeros s valores.
2. Si $S_{1,t}$ y $S_{2,t}$ son funciones estacionales con período s entonces $aS_{1,t} + bS_{2,t}$, para $a, b \in \mathbb{R}$, es también una función estacional de período s .

Procesos estocásticos

Un **proceso estocástico** puede ser descrito como un fenómeno estadístico que evoluciona en el tiempo de acuerdo a las leyes de probabilidad. Hay ejemplos bastante conocidos de procesos estocásticos, como:

- Longitud de una fila
- Tamaño de una colonia de bacterias
- Temperatura del aire en días sucesivos en una sitio en particular

En la literatura, la noción de proceso estocástico puede ser también llamada **proceso aleatorio**

Matemáticamente, un proceso estocástico puede ser definido como una colección de variables aleatorias que están ordenadas en el tiempo y definidas en un conjunto de puntos que pueden ser continuos o discretos.

Escribiremos la variable aleatoria en el tiempo t como $\mathbf{X}(t)$ si el tiempo es continuo (usualmente, $-\infty < t < \infty$), y por \mathbf{X}_t , si el tiempo es discreto (usualmente, $t = 0, \pm 1, \pm 2, \dots$).

En la mayoría de los problemas de estadística se desea estimar propiedades de la población a partir de una muestra. En el análisis de series de tiempo esto varía un poco, debido a que si bien es posible variar la longitud de la serie de tiempo observada, es usualmente imposible hacer más de una observación en un tiempo determinado. Por lo que tenemos sólo un resultado único del proceso y una observación única de la variable aleatoria en el tiempo t .

Sin embargo, podemos considerar la serie de tiempo observada como un ejemplo del conjunto infinito de series de tiempos que podrían haber sido observadas. Cada elemento de este conjunto es una **realización** del proceso estocástico.

Así, las series de tiempo observadas puede ser pensadas como una realización particular, y será denotado como $x(t)$ para $(0 \leq t \leq T)$ si las observaciones son continuas, y por x_t para $t = 1, \dots, N$ si las observaciones son discretas.

Una forma de describir un proceso estocástico es especificar la distribución de probabilidad conjunta de $X(t_1), \dots, X(t_n)$ para cualquier conjunto de punto t_1, \dots, t_n y cualquier valor de n . Sin embargo, esto es bastante complicado y no es usualmente aplicado en la práctica.

Una forma más sencilla y útil para describir un proceso estocástico es dar los **momentos** del proceso, particularmente el primer y segundo momento, que son llamadas la media, la varianza y la función de autocovarianza. En lo que sigue utilizamos notación para tiempo continuo, pero las definición son análogas para tiempo discreto.

- **Media:** la función de medias $\mu(t)$ está definida como

$$\mu(t) = \mathbb{E}(\mathbf{X}(t))$$

- **Varianza:** La función de varianza $\sigma^2(t)$ está definida como

$$\sigma^2(t) = \mathbb{V}(\mathbf{X}(t))$$

Autocovarianza

La función de varianza por si sola no es suficiente para especificar los segundos momentos de una secuencia de variables aleatorias. Además, debemos definir la función de autocovarianza $\gamma(t_1, t_2)$, que es la covarianza de $\mathbf{X}(t_1)$ y $\mathbf{X}(t_2)$, esto es:

$$\gamma(t_1, t_2) = \mathbb{E}([\mathbf{X}(t_1) - \mu(t_1)][\mathbf{X}(t_2) - \mu(t_2)])$$

Notas que la varianza es un caso especial de la función de autocovarianza cuando $t_1 = t_2$.

Los siguientes momentos de un proceso estocástico pueden ser definidos de manera trivial, pero son rara vez usado en la práctica, ya que el conocer las dos funciones $\mu(t)$ y $\gamma(t_1, t_2)$ es usualmente suficiente.

Proceso estacionario

Una clase importante de procesos aleatorios son los **estacionarios**. Una idea heurística de estacionariedad es que no hayan cambios sistemáticos en la media (tendencia) o en la varianza, y que las variaciones estrictamente periódicas hayan sido eliminadas.

Una serie de tiempo se dice **estrictamente estacionaria** si la distribución de probabilidad de $\mathbf{X}(t_1), \dots, \mathbf{X}(t_n)$ es la misma que la distribución conjunta de $\mathbf{X}(t_1 + \tau), \dots, \mathbf{X}(t_n + \tau)$ para todo t_1, \dots, t_n .

En otras palabras, cambiar el tiempo inicial en una cantidad τ no tiene efecto en la distribución conjunto, por lo que esta debe depender sólo de los intervalos entre t_1, t_2, \dots, t_n . Lo anterior, para cualquier n .

En particular, si $n = 1$, estacionariedad estricta implica que la distribución de $\mathbf{X}(t)$ es la misma para todo t , por lo que, asumiendo que los dos primeros momentos son finitos, se tiene

$$\mu(t) = \mu \quad \sigma^2(t) = \sigma^2$$

que son dos constantes que no dependen del valor de t .

Si $n = 2$, la distribución conjunta de $\mathbf{X}(t_1)$ y $\mathbf{X}(t_2)$ sólo depende de $(t_2 - t_1)$, que es conocido como el **lag** (rezago). Por lo que la función de autocovarianza $\gamma(t_1, t_2)$ también sólo depende de $(t_2 - t_1)$ y puede ser escrita como $\gamma(\tau)$, donde

$$\begin{aligned}\gamma(\tau) &= \mathbb{E}([\mathbf{X}(t) - \mu][\mathbf{X}(t + \tau) - \mu]) \\ &= \text{Cov}[\mathbf{X}(t), \mathbf{X}(t + \tau)]\end{aligned}$$

es llamado el coeficiente de autocovarianza en el **lag** τ .

El tamaño del coeficiente de autocovarianza depende de las unidades en las cuales $\mathbf{X}(t)$ es medido. Por lo que con el propósito de interpretar correctamente, es útil estandarizar la función de autocovarianza para producir una función llamada **función de autocorrelación**, dada por

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}$$

que mide la correlación entre $\mathbf{X}(t)$ y $\mathbf{X}(t + \tau)$.

Proceso débilmente estacionario

En la práctica, es útil definir un tipo de estacionariedad menos restrictiva que la que acabamos de definir. Un proceso es llamado **estacionario de segundo orden o débilmente estacionario** si sus medias son constantes y su función de autocovarianza depende sólo del **lag**, por lo que

$$\mathbb{E}(\mathbf{X}(t)) = \mu$$

y,

$$\text{Cov}(\mathbf{X}(t), \mathbf{X}(t + \tau)) = \gamma(\tau)$$

No se asume nada sobre los momentos mayores al de segundo orden. Notar que si $\tau = 0$, debido a la función de autocovarianza, se tiene que la varianza y la media deben ser constantes. Además, estas dos últimas deben ser finitas.

Función de autocorrelación

Como sabemos la función de autocorrelación está definida por

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}$$

Esta función es una herramienta principal en la descripción de una serie de tiempo, tal como la función de autocorrelación teórica es una herramienta importante para describir las propiedades de proceso estocástico.

Supongamos que tenemos un proceso estocástico $X(t)$ con media μ , varianza σ^2 , autocovarianza $\gamma(\tau)$ y autocorrelación $\rho(\tau)$, por lo que

$$\rho(\tau) = \gamma(\tau)/\sigma^2$$

Propiedades

1. La función de autocorrelación es una función par del **lag**, esto es

$$\rho(\tau) = \rho(-\tau)$$

Esta propiedad dice que la correlación entre $X(t)$ y $X(t + \tau)$ es la misma que la entre $X(t)$ y $X(t - \tau)$.

Para demostrar esta propiedad usamos el hecho que $\gamma(\tau) = \rho(\tau)\sigma^2$ y la estacionariedad de $X(t)$.

$$2. |\rho(\tau)| \leq 1$$

Esta propiedad es ‘usual’ para las correlaciones. Se obtiene notando que

$$\mathbb{V}(\lambda_1 X(t) + \lambda_2 X(t + \tau)) \geq 0$$

Para cualquier constantes λ_1, λ_2 , debido a que la varianza es siempre no negativa.
Esta varianza es igual a

$$\lambda_1^2 \mathbb{V}(X(t)) + \lambda_2^2 \mathbb{V}(X(t + \tau)) + 2\lambda_1 \lambda_2 \text{Cov}(X(t), X(t + \tau)) = (\lambda_1^2 + \lambda_2^2)\sigma^2 + 2\lambda_1 \lambda_2 \gamma(\tau)$$

Cuando $\lambda_1 = \lambda_2 = 1$, vemos que

$$\gamma(\tau) \geq -\sigma^2 \Rightarrow \rho(\tau) \geq -1$$

Cuando $\lambda_1 = 1, \lambda_2 = -1$, vamos que

$$\sigma^2 \geq \gamma(\tau) \Rightarrow \rho(\tau) \leq 1$$

3. Falta de unicidad

A pesar de que un proceso estocástico tenga una estructura de covarianza única, el converso no es verdadero en general.

Incluso para procesos normales estacionarios, que están completamente determinados por su media, varianza y función de autocovarianza, las condiciones de invertibilidad (que veremos más adelante) requerirán que se asegure la unicidad

Ruido blanco

En lo que sigue veremos, distintos tipos de procesos estocásticos que nos serán a lo largo del curso.

Un proceso a tiempo discreto es llamado un proceso puramente aleatorio si consiste de una secuencia de variables aleatorias $\{Z_t\}$ que son mutuamente independientes e idénticamente distribuidas. Desde la definición sigue que el proceso tiene media y varianza constante. y

$$\begin{aligned}\gamma(k) &= \text{Cov}(Z_t, Z_{t+k}) \\ &= 0 \quad \text{para } k = \pm 1, 2, \dots\end{aligned}$$

Como la media y la función de autocovarianza no dependen del tiempo, el proceso es débilmente estacionario. De hecho, el proceso es además estrictamente estacionario y su autocorrelación está dada por

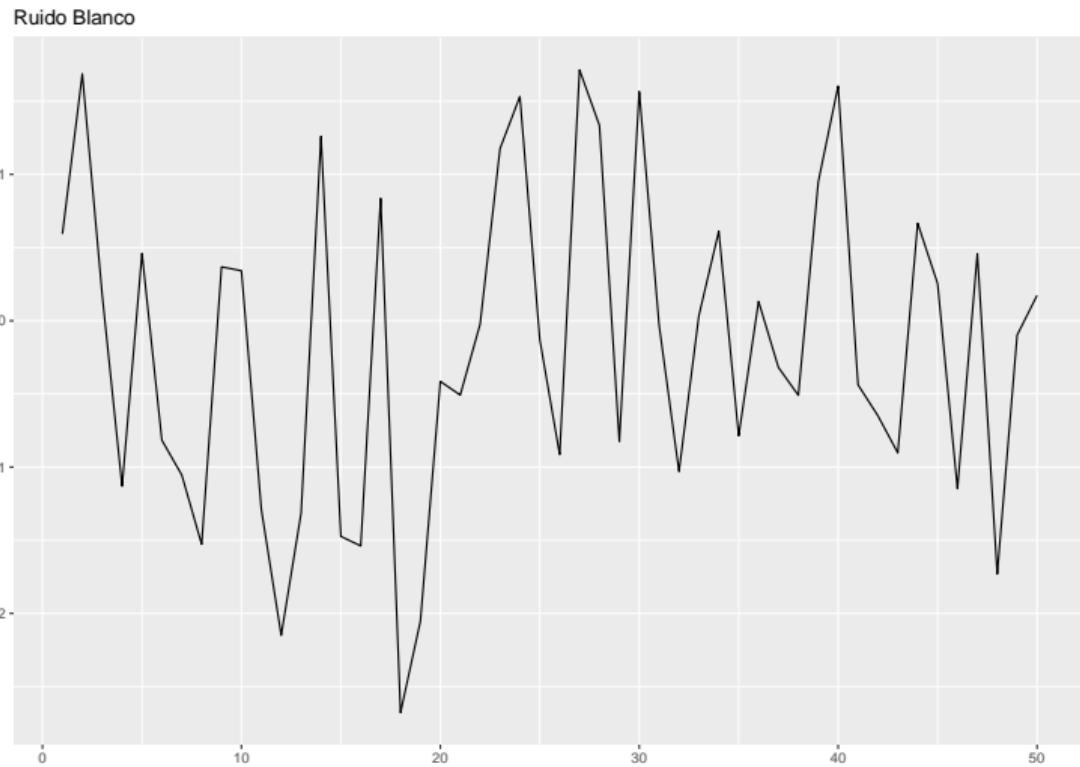
$$\rho(k) = \begin{cases} 1 & k = 0 \\ 0 & k = \pm 1, \pm 2, \dots \end{cases}$$

Un proceso completamente aleatorio es llamado **ruido blanco** o **innovaciones**.

Simulación de un ruido blanco

```
library(ggplot2)
library(ggfortify)
library(gridExtra)
set.seed(414)
y <- ts(rnorm(50))
```

```
autoplott(y) + ggtitle("Ruido Blanco")
```



Paseo Aleatorio

Supongamos que $\{Z_t\}$ es un proceso puramente aleatorio discreto con media μ y varianza σ_Z^2 . Un proceso $\{X_t\}$ se dice que es un **paseo aleatorio** si

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

El proceso es usualmente empezado en 0 cuando $t = 0$, por lo que

$$X_1 = Z_1$$

y,

$$X_t = \sum_{i=1}^t Z_i$$



Así, $E(X_t) = t\mu$ y $V(X_t) = t\sigma_Z^2$. Como la media y la varianza cambian con el tiempo, el proceso no es estacionario.

Sin embargo, notamos que la primera diferencia de un paseo aleatorio está dada por

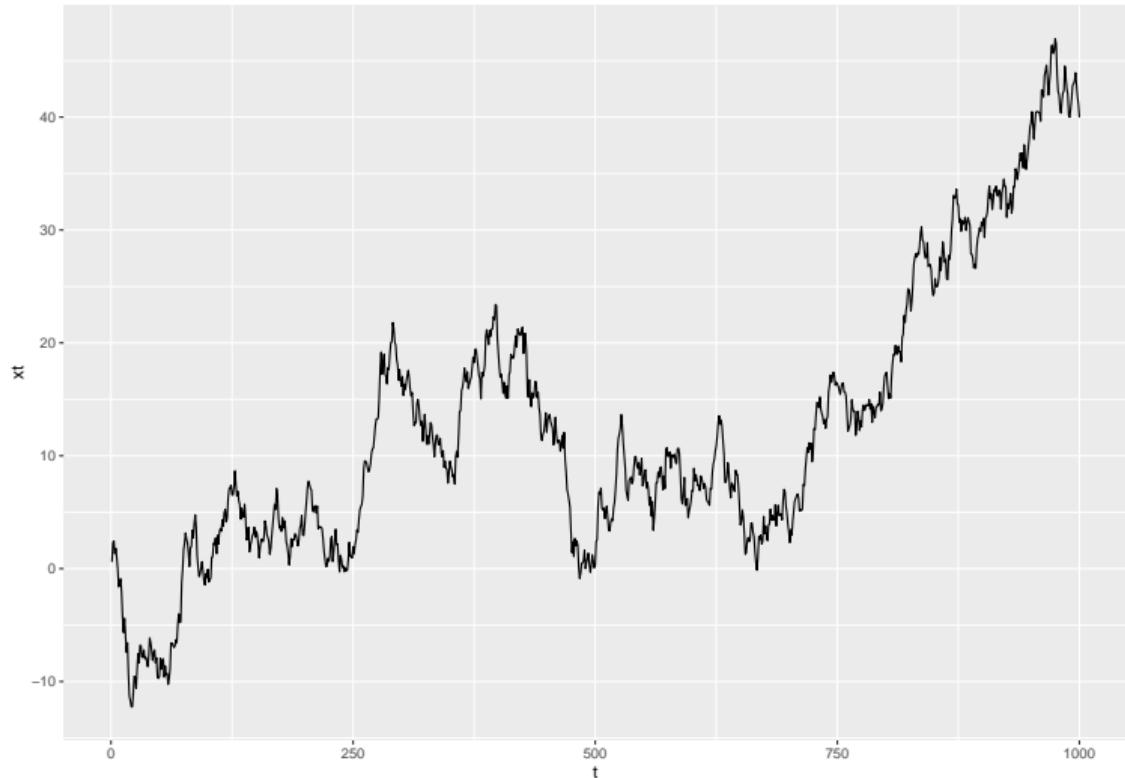
$$\nabla X_t = X_t - X_{t-1} = Z_t$$

que es un proceso puramente aleatorio, que sí es estacionario.

Simulación de un paseo aleatorio

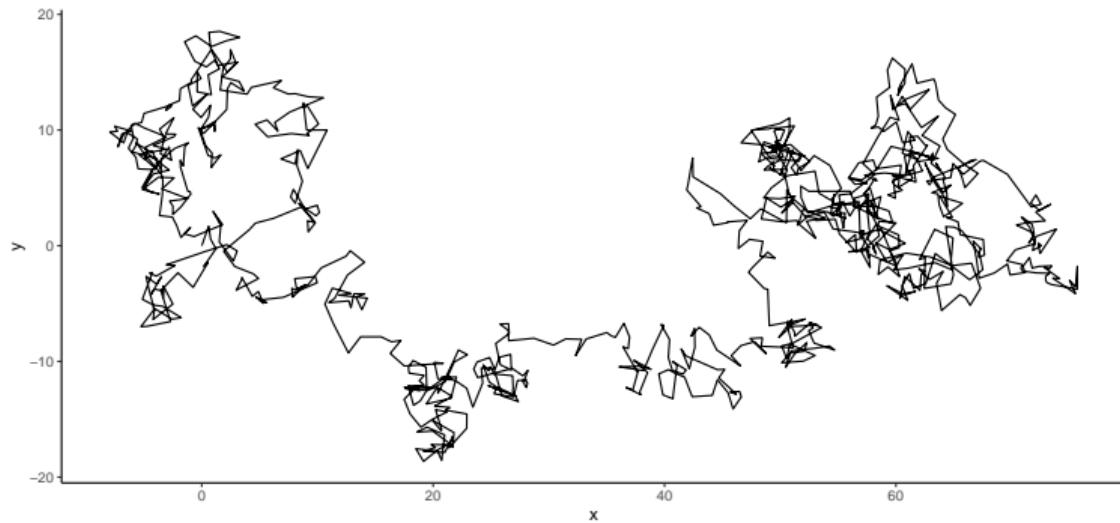
```
set.seed(414)
random_walk <- function(number=1000){
  data.frame(x = rnorm(number),
              t = c(1:1000)) %>%
    mutate(xt = cumsum(x))
}
p <- ggplot() + aes(x = t, y = xt)
```

```
p + geom_line(data = random_walk())
```



```
set.seed(414)
Xt = 0; Yt = 0
for (i in 2:1000)
{
  Xt[i] = Xt[i-1] + rnorm(1,0,1)
  Yt[i] = Yt[i-1] + rnorm(1,0,1)
}
df <- data.frame(x = Xt, y = Yt)
```

```
ggplot(df, aes(x=x, y=y)) + geom_path() + theme_classic()+
  coord_fixed(1)
```



Procesos autorregresivos

Supongamos que $\{Z_t\}$ es un proceso puramente aleatorio con media cero y varianza σ_Z^2 . Entonces, el proceso $\{X_t\}$ se dice que es un **proceso autorregresivo de orden p** si

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + Z_t$$

Es fácil ver que la ecuación anterior corresponde a un modelo de regresión múltiple, pero X_t no depende de variables independientes sino de valores pasados de X_t . Un proceso de orden p lo abreviamos como $AR(p)$.

AR(1)

Supongamos que $p = 1$ por lo que

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

Si $X_0 = h$ y sustituimos sucesivamente en la ecuación anterior, se tiene:

$$X_1 = \alpha h + Z_1$$

$$X_2 = \alpha^2 h + \alpha Z_1 + Z_2$$

⋮

$$X_t = \alpha^t h + \sum_{i=0}^{t-1} \alpha^i Z_{t-i}$$

Si calculamos la esperanza de X_t , como Z_t es un ruido blanco, se tiene:

$$\mathbb{E}(X_t) = \alpha^t h$$

¿Cómo cambiarían estas expresiones si se define X_t con una constante fija?, esto es

$$X_t = \beta + \alpha X_{t-1} + Z_t$$

El proceso $AR(1)$ también puede ser escrito como

$$(1 - \alpha B)X_t = Z_t$$

en donde el término B es el operador de retardo, definido como

$$BX_t = X_{t-1}$$

Por lo que

$$\begin{aligned}X_t &= Z_t / (1 - \alpha B) \\&= (1 + \alpha B + \alpha^2 B^2 + \dots) Z_t \\&= Z_t + \alpha Z_{t-1} + \alpha^2 Z_{t-2} + \dots\end{aligned}$$

Lo anterior debido a que la serie de MacLaurin de $\frac{1}{1-x}$ es

$$\sum_{n=0}^{\infty} x^n = 1 + x + x^2 + \dots$$

Cuando lo expresamos de esta manera, es claro ver que

$$\mathbb{E}(X_t) = 0$$

y,

$$\mathbb{V}(X_t) = \sigma_Z^2(1 + \alpha^2 + \alpha^4 + \dots)$$

De anterior se desprende que la varianza será finita si $|\alpha| < 1$, y en este caso

$$\mathbb{V}(X_t) = \sigma_X^2 = \frac{\sigma_Z^2}{(1 - \alpha^2)}$$

Función de autocovarianza

La función de autocovarianza estará dada por

$$\begin{aligned}\gamma(k) &= \mathbb{E}(X_t X_{t+k}) \\ &= \mathbb{E} \left[\left(\sum_i \alpha^i Z_{t-i} \right) \left(\sum_i \alpha^j Z_{t+k-j} \right) \right] \\ &= \alpha_Z^2 \sum_{i=0}^{\infty} \alpha^i \alpha^{k+i} \quad \text{para } k \geq 0\end{aligned}$$

que converge para $|\alpha| < 1$ a

$$\begin{aligned}\gamma(k) &= \alpha^k \sigma_Z^2 / (1 - \alpha^2) \\ &= \alpha^k \sigma_X^2\end{aligned}$$

Función de autocorrelación

Tarea: Para $k < 0$ se tiene que $\gamma(k) = \gamma(-k)$

Debido a que $\gamma(k)$ no depende de t , un proceso AR de orden 1 es **débilmente estacionario** sujeto a que $|\alpha| < 1$. Su función de autocorrelación estará dada por:

$$\rho(k) = \alpha^k \quad k = 0, 1, 2, \dots$$

Para obtener una función par definida para todos los k enteros, se puede escribir

$$\rho(k) = \alpha^{|k|} \quad k = 0, \pm 1, \pm 2, \dots$$

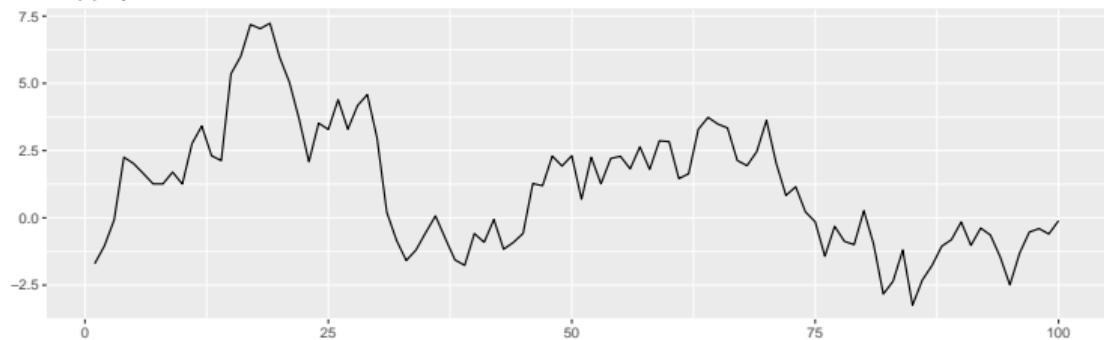
Tarea: Muestre que se tiene la recursión

$$\rho(k) = \alpha \rho(k - 1)$$

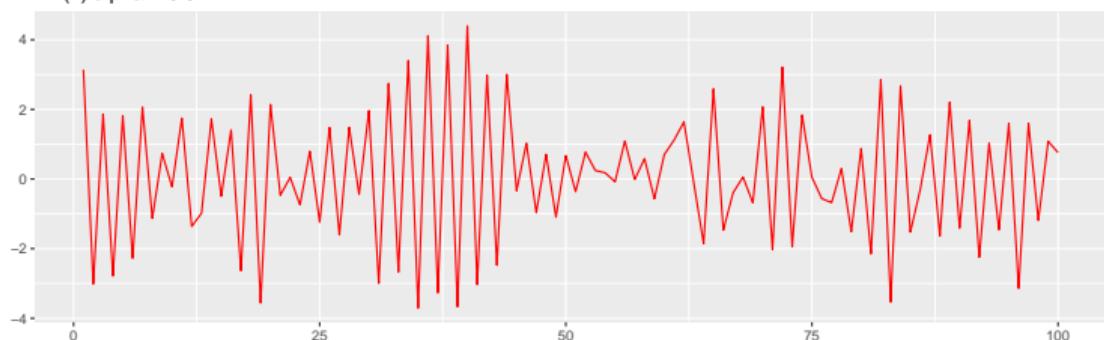
Simulación de AR(1)

```
set.seed(414)
ar1_a<-arima.sim(list(order=c(1,0,0), ar=.9), n=100)
ar1_b<-arima.sim(list(order=c(1,0,0), ar=-.9), n=100)
p1<-autoplot(ar1_a, ts.colour = 'black', main = 'AR(1) alpha=0.9')
p2<-autoplot(ar1_b, ts.colour = 'red', main = 'AR(1) alpha=-0.9')
```

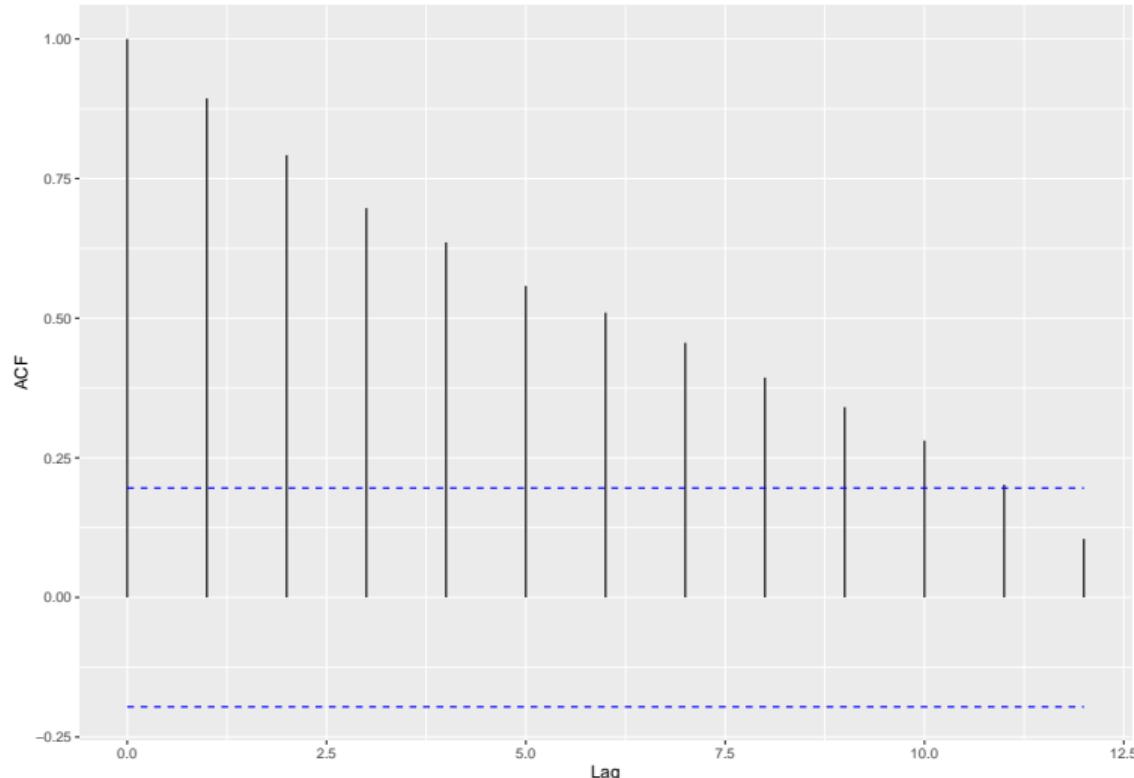
AR(1) alpha=0.9



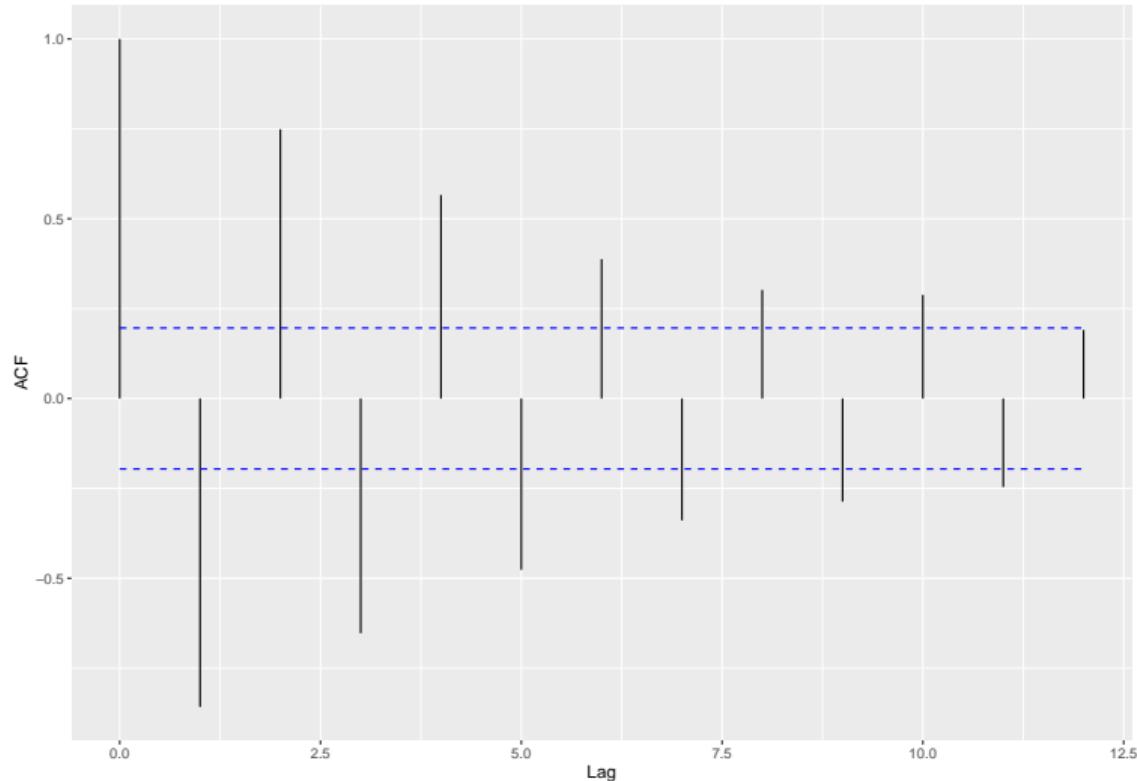
AR(1) alpha=-0.9



```
autofplot(acf(ar1_a, lag.max = 12, plot=FALSE))
```



```
autofplot(acf(ar1_b, lag.max = 12, plot=FALSE))
```



AR(2)

Por definición el modelo autorregresivo de orden 2, que lo denotamos por AR(2), satisface

$$X_t = c + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + Z_t$$

en donde c, α_1 y α_2 son constantes y Z_t es un ruido blanco. Lo anterior lo podemos reescribir en términos del operador de lag

$$(1 - \alpha_1 B - \alpha_2 B^2)X_t = c + Z_t$$

Si tomamos la esperanza en la primera ecuación obtenemos (e imponiendo que la media sea constante)

$$\mu = c + \alpha_1\mu + \alpha_2\mu$$

que implica

$$\mu = \frac{c}{1 - \alpha_1 - \alpha_2}$$

y la condición para que el proceso tenga media finita es

$$1 - \alpha_1 - \alpha_2 \neq 0$$

Si sustituimos c por $\mu(1 - \alpha_1 - \alpha_2)$ y usando $X_t^* = X_t - \mu$ al proceso en desviaciones a su media, entonces

$$X_t^* = \alpha_1 X_{t-1}^* + \alpha_2 X_{t-2}^* + Z_t$$

Para estudiar las propiedades del proceso es conveniente utilizar la notación con operador de lag, esto es:

$$(1 - \alpha_1 B - \alpha_2 B^2)X_t = c + Z_t$$

que tras la formación utilizada se convierte en

$$(1 - \alpha_1 B - \alpha_2 B^2)X_t^* = Z_t$$

El operador $(1 - \alpha_1 B - \alpha_2 B^2)$ puede expresarse como $(1 - G_1 B)(1 - G_2 B)$, donde G_1^{-1} y G_2^{-1} son las raíces de la ecuación del operador considerando B como variable y resolviendo $1 - \alpha_1 B - \alpha_2 B^2 = 0$

Esta ecuación se denomina la **ecuación característica** del operador.

En este caso, la condición de estacionariedad es que $|G_i| < 1, i = 1, 2$. Esta condición es análoga a la estudiada para el $AR(1)$ y es coherente con la condición encontrada para que la media sea finita.

Función de autocovarianza

Tomando como inicio el proceso $AR(2)$ definido por

$$X_t^* = \alpha_1 X_{t-1}^* + \alpha_2 X_{t-2}^* + Z_t$$

elevando al cuadrado y tomando esperanza, obtenemos que su varianza debe satisfacer

$$\gamma(0) = \alpha_1^2 \gamma(0) + \alpha_2^2 \gamma(0) + 2\alpha_1 \alpha_2 \gamma(1) + \sigma^2$$

Para calcular la autocovarianza multiplicamos el proceso inicial por X_{t-1}^* y tomamos esperanza, obteniendo

$$\gamma(k) = \alpha_1 \gamma(k-1) + \alpha_2 \gamma(k-2) \quad k \geq 1$$

Si $k = 1$, como en un proceso estacionario $\gamma(-1) = \gamma(1)$, se obtiene:

$$\gamma(1) = \alpha_1\gamma(0) + \alpha_2\gamma(1) \Rightarrow \gamma(1) = \frac{\alpha_1\gamma(0)}{(1 - \alpha_2)}$$

Luego, sustituyendo en la ecuación de varianza, resulta la fórmula

$$\sigma_{X^*}^2 = \gamma(0) = \frac{(1 - \alpha_2)\sigma^2}{(1 + \alpha_2)(1 - \alpha_1 - \alpha_2)(1 + \alpha_1 - \alpha_2)}$$

Para que el proceso sea estacionario, esta varianza debe ser positiva que sucede cuando el numerador y denominador tienen el mismo signo. Así, los parámetros que hacen que un proceso $AR(2)$ sea estacionario son los incluidos en la región

$$-1 < \alpha_2 < 1, \quad \alpha_1 + \alpha_2 < 1, \quad \alpha_2 - \alpha_1 < 1$$

Función de autocorrelación

De la función general de autocovarianza para el proceso $AR(2)$, al dividir por la varianza, obtenemos la relación entre los coeficientes de autocorrelación

$$\rho(k) = \alpha_1 \rho(k-1) + \alpha_2 \rho(k-2)$$

Así, si $k = 1$, como en un proceso estacionario $\rho(1) = \rho(-1)$, se obtiene:

$$\rho(1) = \frac{\alpha_1}{1 - \alpha_2}$$

y para $k = 2$, utilizando la expresión anterior se obtiene

$$\rho(2) = \frac{\alpha_1^2}{1 - \alpha_2} + \alpha_2$$

Para $k \geq 3$ los coeficientes de autocorrelación pueden obtenerse recursivamente a partir de la ecuación de $\rho(k)$. Es posible mostrar que la solución general de esta ecuación es

$$\rho(k) = A_1 G_1^k + A_2 G_2^k$$

donde G_1 y G_2 son los factores del polinomio característico del proceso, y A_1 y A_2 constantes a determinar a partir de las condiciones iniciales.

AR(2) como suma de innovaciones

Como vimos antes, el proceso $AR(2)$ puede expresarse como

$$(1 - \alpha_1 B - \alpha_2 B^2)X_t^* = Z_t$$

que a su vez puede reescribirse como

$$(1 - G_1 B)(1 - G_2 B)X_t^* = Z_t$$

Invirtiendo estos operadores se tiene

$$X_t^* = (1 + G_1 B + G_1^2 B^2 + \dots)(1 + G_2 B + G_2^2 B^2 + \dots)Z_t$$

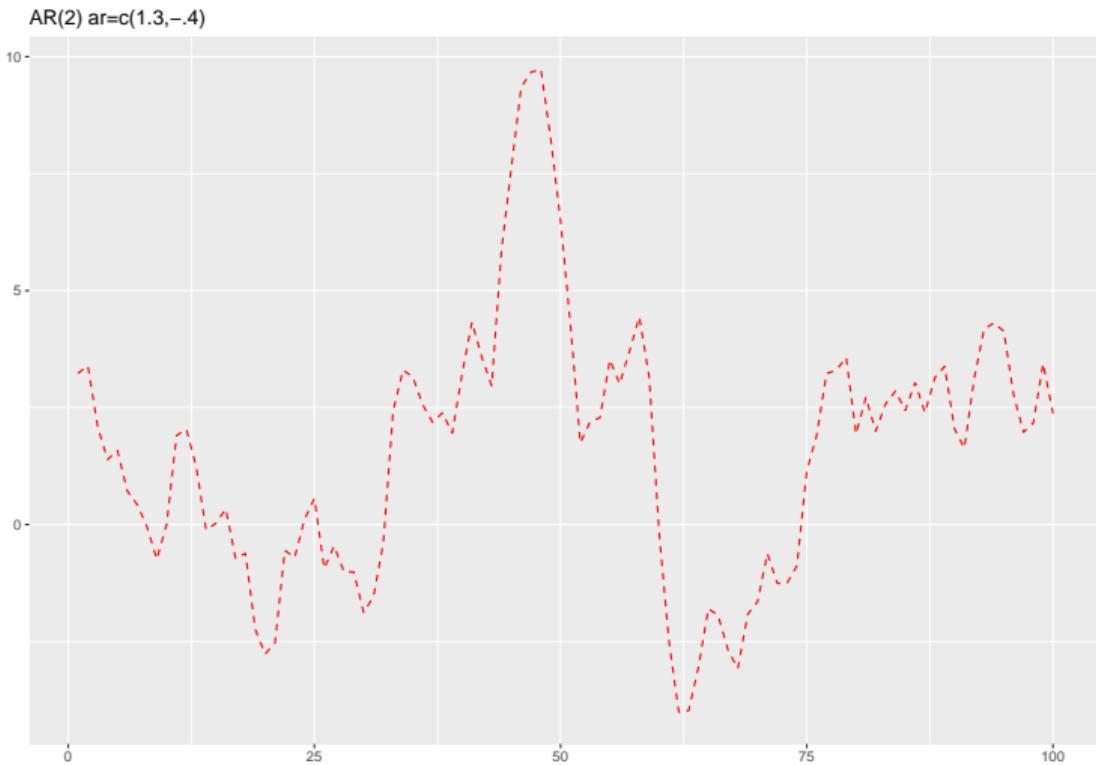
Que conducirá a la expresión del proceso: (que luego llamaremos $MA(\infty)$)

$$X_t^* = Z_t + \psi_1 Z_{t-1} + \psi_2 Z_{t-2} + \dots$$

Los coeficientes ψ_i los podemos obtener como función de las raíces igualando las últimas dos expresiones

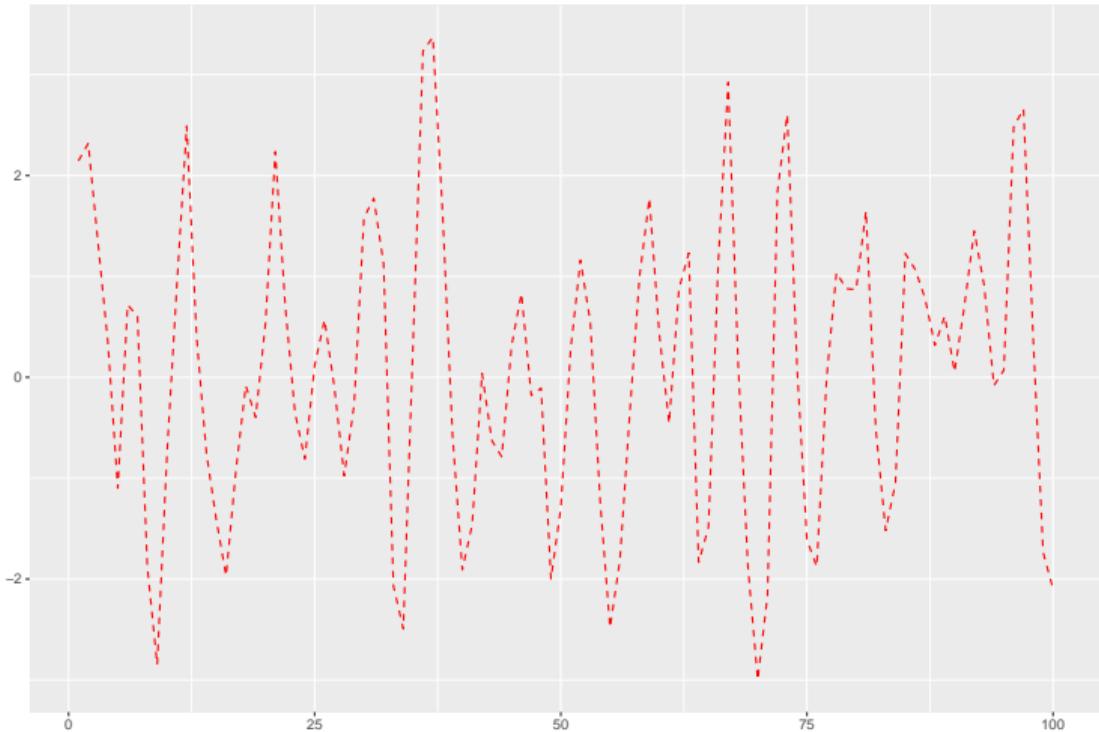
Simulación de AR(2)

```
library(ggfortify)
set.seed(414)
ar2_a<-arima.sim(model=list(ar=c(1.3,-.4)),100)
ar2_b<-arima.sim(model=list(ar=c(.8,-.7)),100)
plot_1<-autoplot(ar2_a, ts.colour = 'red',
                  ts.linetype = 'dashed', main = 'AR(2) ar=c(1.3,-.4)')
plot_2<-autoplot(ar2_b, ts.colour = 'red',
                  ts.linetype = 'dashed', main = 'AR(2) ar=c(.8,-.7)')
```



plot_2

AR(2) ar=c(.8,-.7)



Ejemplo

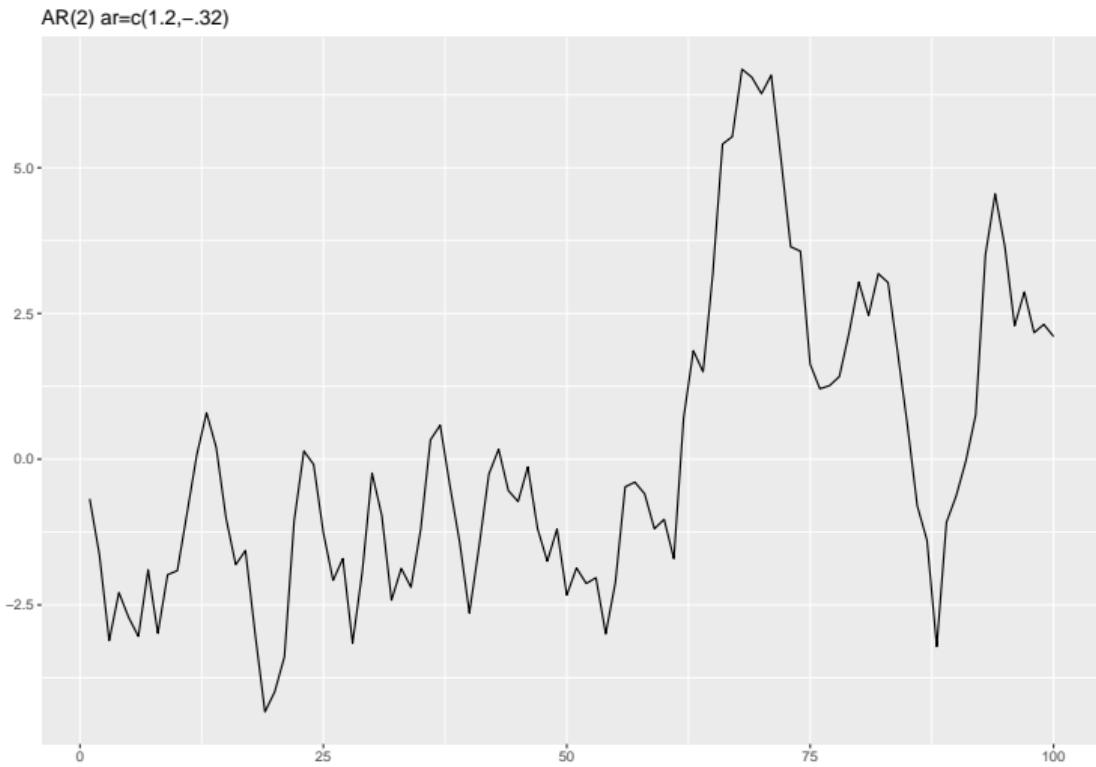
Partamos del proceso $AR(2)$ definido por

$$X_t = 1.2X_{t-1} - 0.32X_{t-2} + Z_t$$

que lo podemos simular como

```
ar2_c<-arima.sim(model=list(ar=c(1.2,-.32)),100)
plot_3<-autoplot(ar2_c, ts.colour = 'black',
                  main = 'AR(2) ar=c(1.2,-.32)')
```

plot_3



La ecuación característica de este proceso es:

$$0.32X^2 - 1.2X + 1 = 0$$

cuya solución es

$$X = \frac{1.2 \pm \sqrt{1.2^2 - 4 * 0.32}}{0.64} = \frac{1.2 \pm 0.4}{0.64}$$

Las soluciones son $G^{-1} = 2.5$ y $G^{-1} = 1.25$ y los factores serán $G_1 = 0.4$ y $G_2 = 0.8$. Así, la ecuación característica puede ser escrita como

$$0.32X^2 - 1.2X + 1 = (1 - 0.4X)(1 - 0.8X)$$

Por lo tanto, el proceso es estacionario con raíces reales y los coeficientes de correlación verifican:

$$\rho(k) = A_1 0.4^k + A_2 0.8^k$$

Para determinar A_1 y A_2 imponemos las condiciones iniciales $\rho(0) = 1$, $\rho(1) = 1.2/(1.32) = 0.91$. Entonces, para $k = 0$

$$1 = A_1 + A_2$$

y para $k = 1$

$$0.91 = 0.4A_1 + 0.8A_2$$

Resolviendo estas ecuaciones se obtiene $A_2 = 0.51/0.4$ y $A_1 = -0.11/0.4$. Por tanto, la función de autocorrelación es

$$\rho(k) = -\frac{0.11}{0.4}0.4^k + \frac{0.51}{0.4}0.8^k$$

Obteniéndose la siguiente tabla:

k	0	1	2	3	4	5	6	7	8
$\rho(k)$	1	0.91	0.77	0.63	0.51	0.41	0.33	0.27	0.21

La representación en función de las innovaciones, escribiendo:

$$(1 - 0.4B)(1 - 0.8B)X_t = Z_t$$

e invirtiendo ambos operadores

$$X_t = (1 + 0.4B + 0.16B^2 + 0.06B^3 + \dots)(1 + 0.8B + 0.64B^2 + \dots)Z_t$$

resulta

$$X_t = (1 + 1.2B + 1.12B^2 + \dots)$$

Tarea: Encuentre la función de autocorrelación para el proceso

$$X_t = X_{t-1} - \frac{1}{2}X_{t-2} + Z_t$$

Diremos que una serie de tiempo X_t estacionaria sigue un proceso autorregresivo de orden p si

$$X_t^* = \alpha_1 X_{t-1}^* + \cdots + \alpha_p X_{t-p}^* + Z_t$$

donde $X_t^* = X_t - \mu$, siendo μ la media del proceso estacionario X_t y Z_t un ruido blanco. Al igual que antes, podemos reescribir este proceso en términos de operador de lag como:

$$(1 - \alpha_1 B - \cdots - \alpha_p B^p) X_t^* = Z_t$$

en donde llamamos $\phi(B) = 1 - \alpha_1 B - \cdots - \alpha_p B^p$ al polinomio de grado p del operador de lag con $p \geq 1$.

Así, la expresión general de un proceso autorregresivo puede ser escrita como:

$$\phi(B)X_t^* = Z_t$$

La **ecuación característica** de este proceso autoregresivo la definimos como

$$\phi(B) = 0$$

en donde consideramos el operador B como variable. Esta ecuación tendrá p raíces $G_i^{-1}, \dots, G_p^{-1}$, en general distintas, por lo que usando el **teorema fundamental del álgebra** podemos reescribir esta función como:

$$\phi(B) = \prod_{i=1}^p (1 - G_i B)$$

Función de autocorrelación

De la forma general del proceso

$$X_t^* = \alpha_1 X_{t-1}^* + \cdots + \alpha_p X_{t-p}^* + Z_t$$

Si multiplicamos la ecuación por el X_{t-k}^* con $k > 0$, tomando esperanzas y luego dividiendo por $\gamma(0)$, es posible obtener la forma general para la autorrelación:

$$\rho(k) = \alpha_1 \rho(k-1) + \cdots + \alpha_p \rho(k-p), \quad k > 0$$

Que tiene la misma forma que en los casos $k = 1, 2$ vistos anteriormente.

Los coeficientes de autocorrelación satisfacen la misma ecuación que el proceso

$$\phi(B)\rho(k) = 0 \quad k > 0$$

En donde la solución general de esta ecuación es:

$$\rho(k) = \sum_{i=1}^p A_i G_i^k$$

en donde A_i son constantes a determinar basado en las condiciones iniciales y los G_i son los factores de la ecuación característica.

Ecuaciones de Yule-Walker

Evaluando la ecuación

$$\rho(k) = \sum_{i=1}^p A_i G_i^k$$

para los distintos $k = 1, \dots, p$, se obtiene un sistema de p ecuaciones que relacionan las p primeras autocorrelaciones con los parámetros del proceso.

Así, llamaremos **ecuaciones de Yule-Walker** al sistema:

$$\rho(1) = \alpha_1 + \alpha_2 \rho(1) + \cdots + \alpha_p \rho(p-1)$$

$$\rho(2) = \alpha_1 \rho(1) + \alpha_2 + \cdots + \alpha_p \rho(p-2)$$

⋮

$$\rho(p) = \alpha_1 \rho(p-1) + \alpha_2 \rho(p-2) + \cdots + \alpha_p$$

Si definimos

$$\phi' = [\alpha_1, \dots, \alpha_p], \quad \rho' = [\rho(1), \dots, \rho(p)]$$

y

$$\mathbf{R} = \begin{bmatrix} 1 & \rho(1) & \dots & \rho(p-1) \\ \vdots & \vdots & & \vdots \\ \rho(p-1) & \rho(p-2) & \dots & 1 \end{bmatrix}$$

El sistema de ecuaciones se escribe matricialmente como

$$\rho = \mathbf{R}\phi$$

y los parámetros se determinan a partir de las autocorrelaciones mediante

$$\phi = \mathbf{R}^{-1}\rho$$

Ejemplo

Obtener los parámetros de un proceso $AR(3)$ cuyas primeras autocorrelaciones son $\rho(1) = 0.9, \rho(2) = 0.8, \rho(3) = 0.5$. ¿Es estacionario el proceso?

Primero planteamos las ecuaciones de Yule-Walker:

$$\begin{bmatrix} 0.9 \\ 0.8 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 1 & 0.9 & 0.8 \\ 0.9 & 1 & 0.9 \\ 0.8 & 0.9 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$$

Cuya solución es

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \begin{bmatrix} 5.28 & -5 & 0.28 \\ -5 & 10 & -5 \\ 0.28 & -5 & 5.28 \end{bmatrix} \begin{bmatrix} 0.9 \\ 0.8 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 0.89 \\ 1 \\ -1.11 \end{bmatrix}$$

Así, el proceso $AR(3)$ con estas correlaciones es

$$(1 - 0.89B - B^2 + 1.11B^3)X_t = Z_t$$

Para comprobar que el proceso es estacionario debemos calcular los factores de la ecuación característica, por lo que debemos obtener las soluciones de

$$X^3 - 0.89X^2 - X + 1.11 = 0$$

y comprobar que todas tienen módulo menor que la unidad.

AR(p) como suma de innovaciones

La forma de proceso $AR(p)$ como suma de innovaciones (que después llamaremos $MA(\infty)$), se obtiene invirtiendo el operador $AR(p)$. Si definimos $\psi(B) = \phi(B)^{-1}$, entonces se tiene

$$(1 - \alpha_1 B - \cdots - \alpha_p B^p)(1 + \psi_1 B + \psi_2 B^2 + \dots) = 1$$

en donde los coeficientes ψ_i se obtienen al igualar las potencias de B a cero. Por lo que, se tienen la relación

$$\psi_k = \alpha_1 \psi_{k-1} + \cdots + \alpha_p \psi_{k-1}$$

que es análoga a la que verifican los coeficientes de autocorrelación del proceso.

Función de autocorrelación parcial

Determinar el orden p de un proceso autorregresivo a partir de su función de autocorrelación usual es difícil, por lo que para poder resolver este problema introduciremos la noción de función de autocorrelación parcial.

En general, un $AR(p)$ presenta efectos **directos** de observaciones separadas por $1, 2, \dots, p$ retardos y los efectos **directos** de las observaciones separadas por más de p retardos son nulos. Esta idea es la clave para la utilización de la función de autocorrelación parcial.

Se define el **coeficiente de autocorrelación parcial** de orden k , ρ_k^P como el coeficiente de correlación entre observaciones separadas k periodos, cuando eliminamos de la relación entre las dos variables la dependencia lineal debida a los valores intermedios. Lo calculamos como:

1. Se elimina de X_t^* el efecto de $X_{t-1}^*, \dots, X_{t-k+1}^*$ mediante la regresión:

$$X_t^* = \beta X_{t-1}^* + \dots + \beta_{k-1} X_{t-k+1}^* + u_t$$

donde la variable u_t recoge la parte de X_t^* no común $X_{t-1}^*, \dots, X_{t-k+1}^*$

2. Se elimina de X_{t-k}^* el efecto de $X_{t-1}^*, \dots, X_{t-k+1}^*$ mediante la regresión:

$$X_{t-k}^* = \gamma_1 X_{t-1}^* + \dots + \gamma_{k-1} X_{t-k+1}^* + v_t$$

donde la variable v_t contiene la parte de X_{t-1}^* no común con las observaciones intermedias.

3. Se calcula el coeficiente de correlación simple entre u_t y v_t , y este será el **coeficiente de autocorrelación parcial de orden k**

Es posible mostrar que las etapas anteriores equivalen a ajustar la regresión múltiple

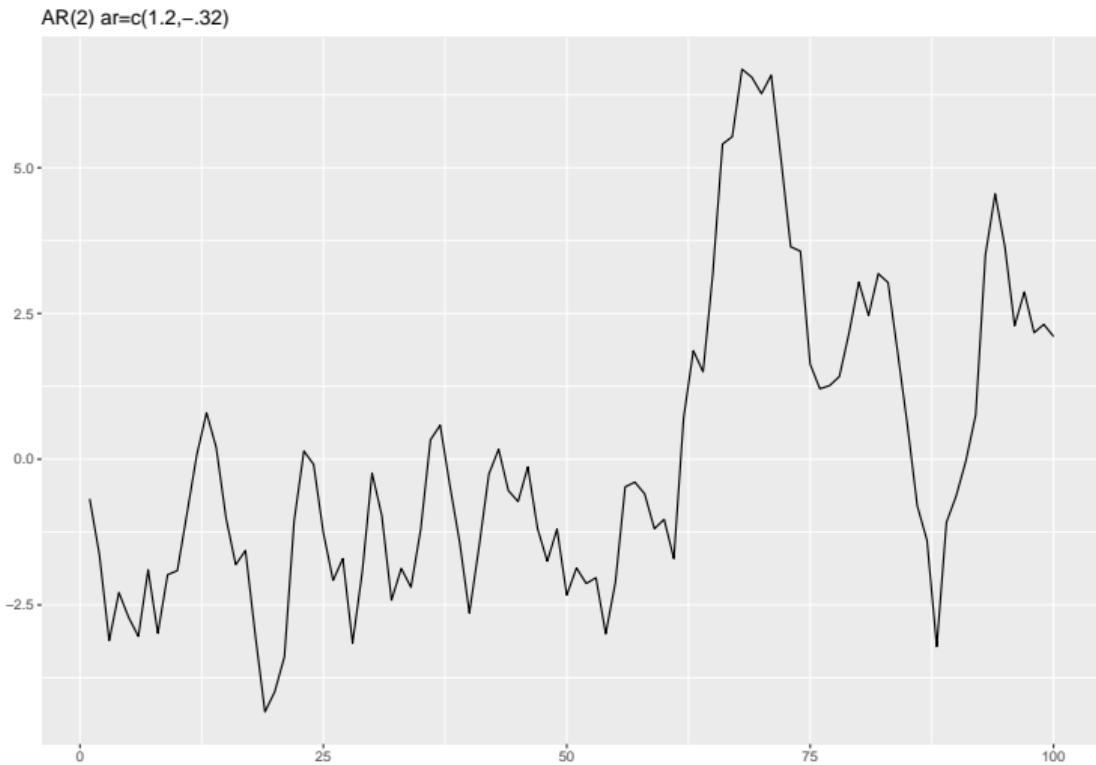
$$X_t^* = \alpha_{k1} X_{t-1}^* + \cdots + \alpha_{kk} X_{t-k}^* + \eta_t$$

y así,

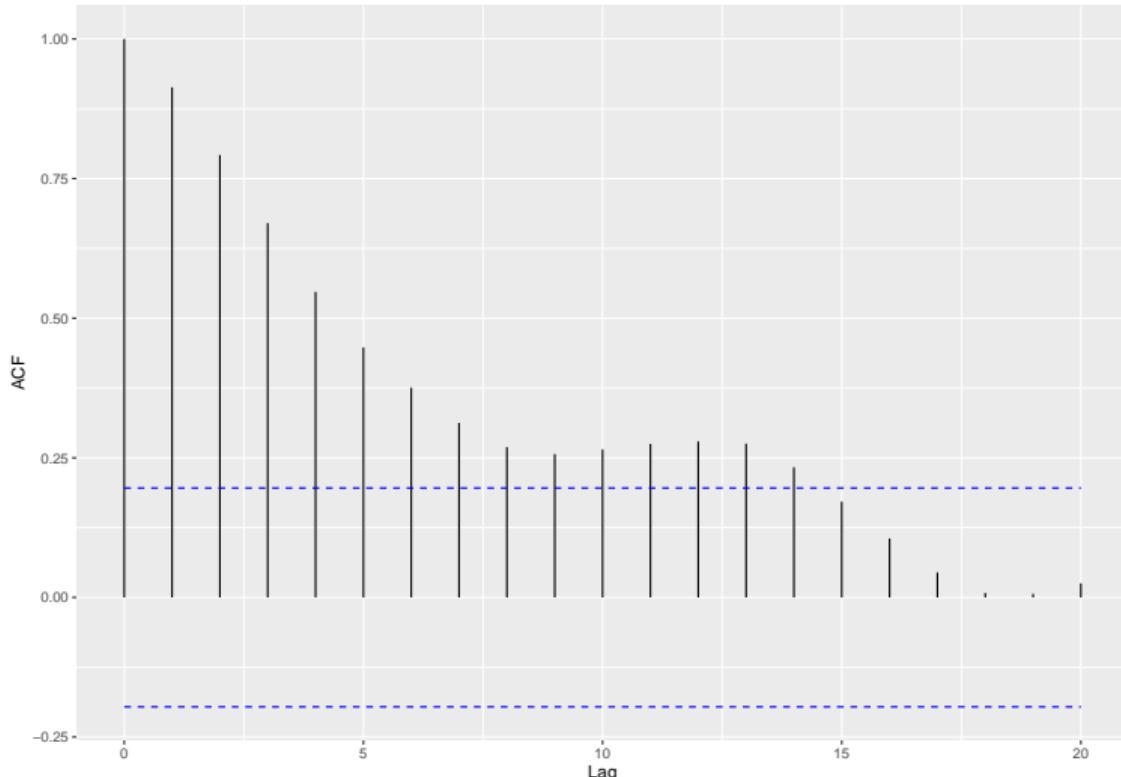
$$\rho_k^p = \alpha_{kk}$$

Es decir, el coeficiente de autocorrelación parcial de orden k es el coeficiente α_{kk} de la variable X_{t-k}^* al ajustar a los datos de la serie un $AR(k)$. Llamamos **función de autocorrelación parcial** a la representación de los coeficientes de autocorrelación parcial en función del retardo (k).

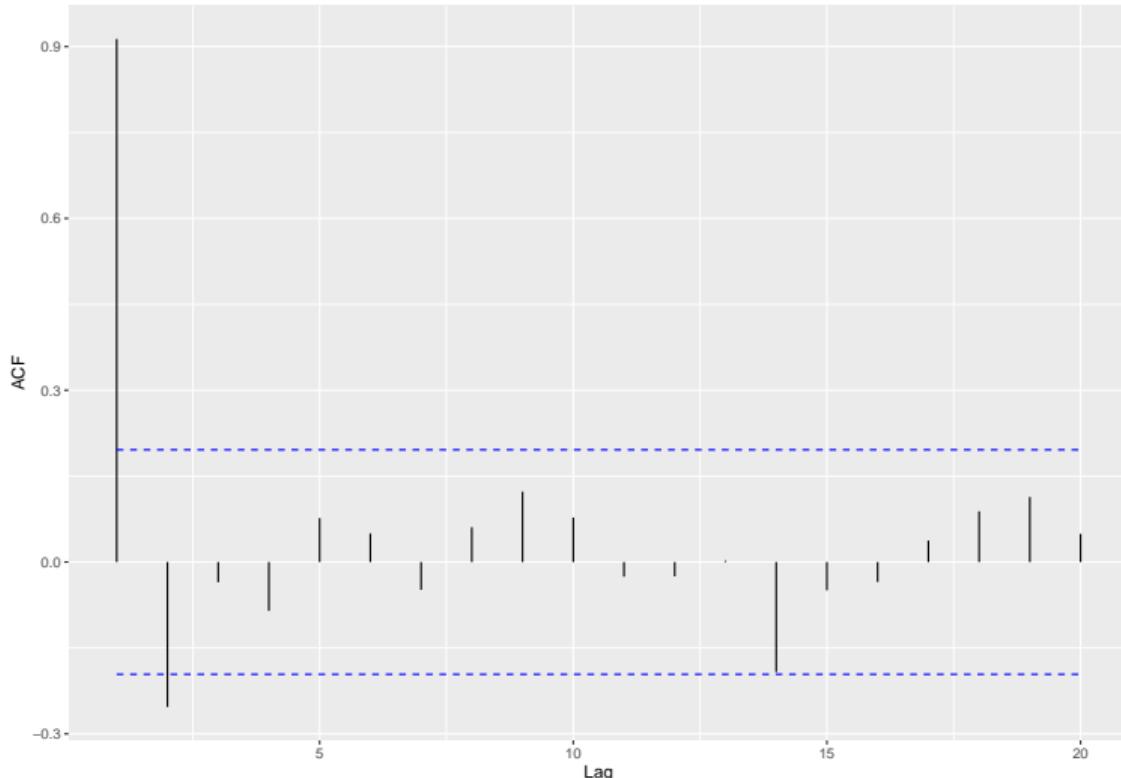
plot_3



```
autofplot(acf(ar2_c, plot=FALSE))
```



```
autoplot(pacf(ar2_c, plot=FALSE))
```



Procesos de media móvil

Los modelos que hemos estudiado antes se caracterizan por tener una memoria relativamente larga, ya que el valor actual está relacionado con todos los anteriores, aunque con coeficientes decrecientes. Esto se traduce en que podemos escribir un proceso autorregresivo como una función lineal de todas las innovaciones que le han generado, con pesos que tienden a cero con el retardo.

Los procesos AR no pueden representar series de tiempo de memoria muy corta, donde el valor actual de la serie sólo está correlacionado con un número pequeño de valores anteriores, de manera que la función de autocorrelación tenga sólo unas pocas correlaciones distintas de cero.

Una familia de procesos que tienen esta propiedad de *memoria corta* son los procesos de media móvil, que le llamamos **MA** por sus siglas en inglés: *moving average*.

Estos procesos **MA** son función de un número finito, y generalmente pequeño, de las innovaciones pasadas.

MA(1)

Un procesos de media móvil de orden 1 está definido como:

$$X_t^* = Z_t - \theta Z_{t-1}$$

donde $X_t^* = X_t - \mu$ siendo μ la media del proceso y Z_t es un proceso de ruido blanco con varianza σ^2 . De igual manera que antes, podemos escribir el proceso usando el operador de lag, como:

$$X_t^* = (1 - \theta B)Z_t$$

Este proceso es la suma de dos procesos estacionarios (Z_t y $-\theta Z_{t-1}$), y por lo tanto, siempre será estacionario, para cualquier valor del parámetro θ (a diferencia de los procesos AR).

En las aplicaciones de este proceso asumiremos que $|\theta| < 1$, de manera que la innovación pasada tenga menos peso que la presente. Entonces, diremos que el proceso es **invertible** y tiene la propiedad que el efecto de los valores pasados decrece en el tiempo. Para justificar eso, reescribiremos Z_{t-1} como función de X_t^* , esto es:

$$X_t^* = Z_t - \theta(X_{t-1}^* + \theta Z_{t-2}) = -\theta X_{t-1}^* - \theta^2 Z_{t-2} + Z_t$$

Si repetimos esta operación para Z_{t-2} , se obtiene:

$$X_t^* = -\theta X_{t-1}^* - \theta^2(X_{t-2}^* + \theta Z_{t-3}) + Z_t = -\theta X_{t-1}^* - \theta^2 X_{t-2}^* - \theta^3 Z_{t-3}^* + Z_t$$

Si seguimos reemplazando, se puede ver que:

$$X_t^* = -\sum_{i=1}^{t-1} \theta^i X_{t-i}^* - \theta^t Z_0 + Z_t$$

Observamos que si $|\theta| < 1$, el efecto X_{t-k}^* tiende a cero y el proceso se denomina **invertible**. Si $|\theta| \geq 1$ se daría la situación paradójica en que el efecto de las observaciones pasadas aumentaría con la distancia, y aunque el proceso seguiría siendo estacionario, es poco adecuado para representar series reales. En lo que sigue, asumiremos que el proceso es invertible.

Como $|\theta| < 1$, existe el operador inverso $(1 - \theta B)^{-1}$ y podemos reescribir el proceso como

$$(1 + \theta B + \theta^2 B^2 + \dots) X_t^* = Z_t$$

que equivale a

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

que es la misma ecuación de antes, asumiendo que el proceso comienza en el pasado infinito. Esta ecuación representa el proceso $MA(1)$ invertible como un $AR(\infty)$ con coeficientes que decrecen en progresión geométrica.

Podemos calcular la varianza del proceso a partir de

$$X_t^* = Z_t - \theta Z_{t-1}$$

Elevando al cuadrado y tomando esperanzas, en donde se obtiene:

$$\mathbb{E}(X_t^2) = \mathbb{E}(Z_t^2) + \theta^2 \mathbb{E}(Z_{t-1}^2) - 2\theta \mathbb{E}(Z_t Z_{t-1})$$

En donde, como $\mathbb{E}(Z_t Z_{t-1}) = 0$ (por ser un proceso de ruido blanco) y $\mathbb{E}(Z_t^2) = \mathbb{E}(Z_{t-1}^2) = \sigma^2$, se tiene que:

$$\sigma_X^2 = \sigma^2(1 + \theta^2)$$

Esta ecuación nos indica que la varianza marginal del proceso, σ_X^2 , es siempre mayor que la varianza de las innovaciones, σ^2 , y aumenta conforme θ^2 crece.

Función de autocorrelación

Para obtener la función de autocorrelación simple de un proceso $MA(1)$ comenzamos calculando las covarianzas. La de primer orden se obtiene multiplicando

$$X_t^* = Z_t - \theta Z_{t-1}$$

por X_{t-1}^* y tomando esperanzas, por lo que se obtiene:

$$\gamma(1) = \mathbb{E}(X_t^* X_{t-1}^*) = \mathbb{E}(Z_t X_{t-1}^*) - \alpha \mathbb{E}(Z_{t-1} X_{t-1}^*)$$

En esta expresión, el primer término es cero debido a que X_{t-1}^* sólo depende de su innovación en el tiempo $t-1$ y $t-2$, y no en el tiempo t . Para calcular el segundo término, sustituimos X_{t-1}^* , obteniéndose:

$$\mathbb{E}(Z_{t-1} X_{t-1}^*) = \mathbb{E}(Z_{t-1} (Z_{t-1} - \alpha Z_{t-2})) = \sigma^2$$

Así,

$$\gamma(1) = -\theta\sigma^2$$

Luego, para calcular la autocovarianza de segundo orden:

$$\gamma(2) = \mathbb{E}(X_t^* X_{t-2}^*) = \mathbb{E}(Z_t X_{t-2}^*) - \alpha \mathbb{E}(Z_{t-1} X_{t-2}^*) = 0$$

Ya que la serie no está correlacionada con sus innovaciones futuras. Siguiendo el mismo procedimiento se pueden obtener las covarianzas de orden superior a dos, ya que al multiplicar la forma general del $MA(1)$ por X_{t-j}^* donde $j > 1$, tendremos productos de la serie por sus innovaciones futuras y las esperanzas de estos son nulas. Por lo que, podemos afirmar que

$$\gamma_j = 0, \quad j > 1$$

Finalmente, al dividir las covarianzas por la varianza del proceso, se obtiene la función de autocorrelación simple, siendo esta:

$$\rho(1) = \frac{-\theta}{1 + \theta^2}, \quad \rho(k) = 0 \quad k > 1$$

Por lo que la función de autocorrelación sólo tiene un valor distinto de cero en el primer retardo.

Como $|\theta| < 1$ el valor del coeficiente de autocorrelación en un $MA(1)$ invertible es siempre menor que 0.5.

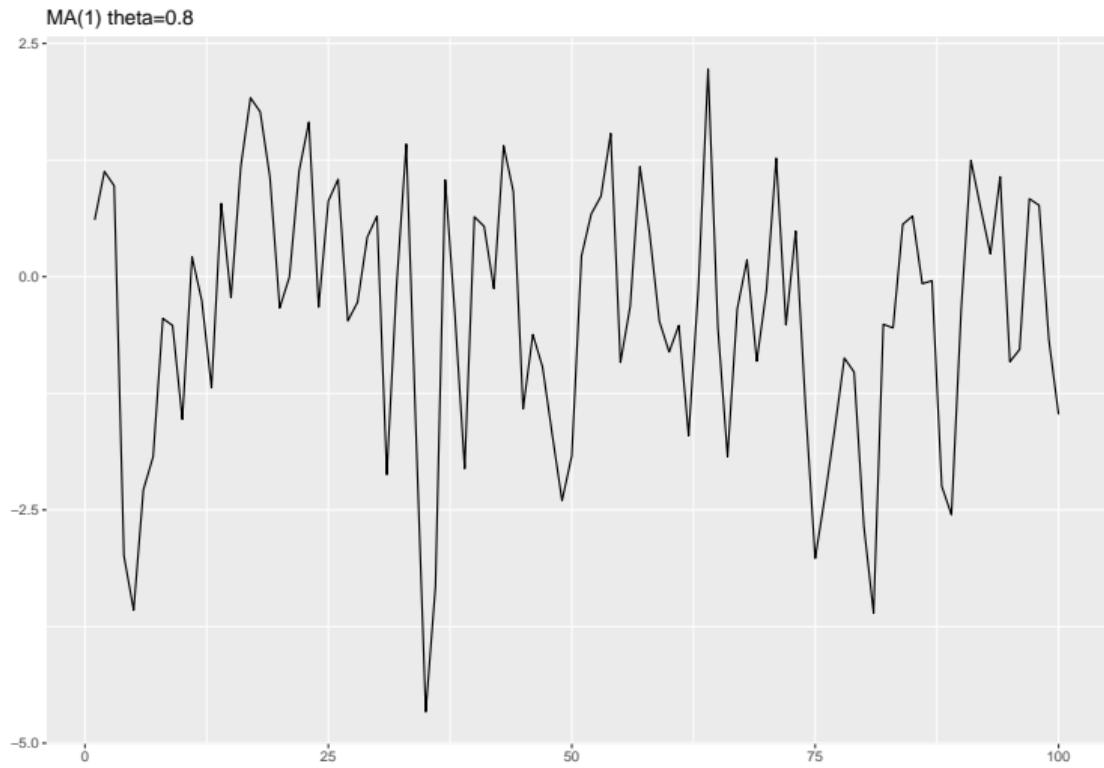
En cuanto a la función de autocorrelación parcial notamos de

$$(1 + \theta B + \theta^2 B^2 + \dots) X_t^* = Z_t$$

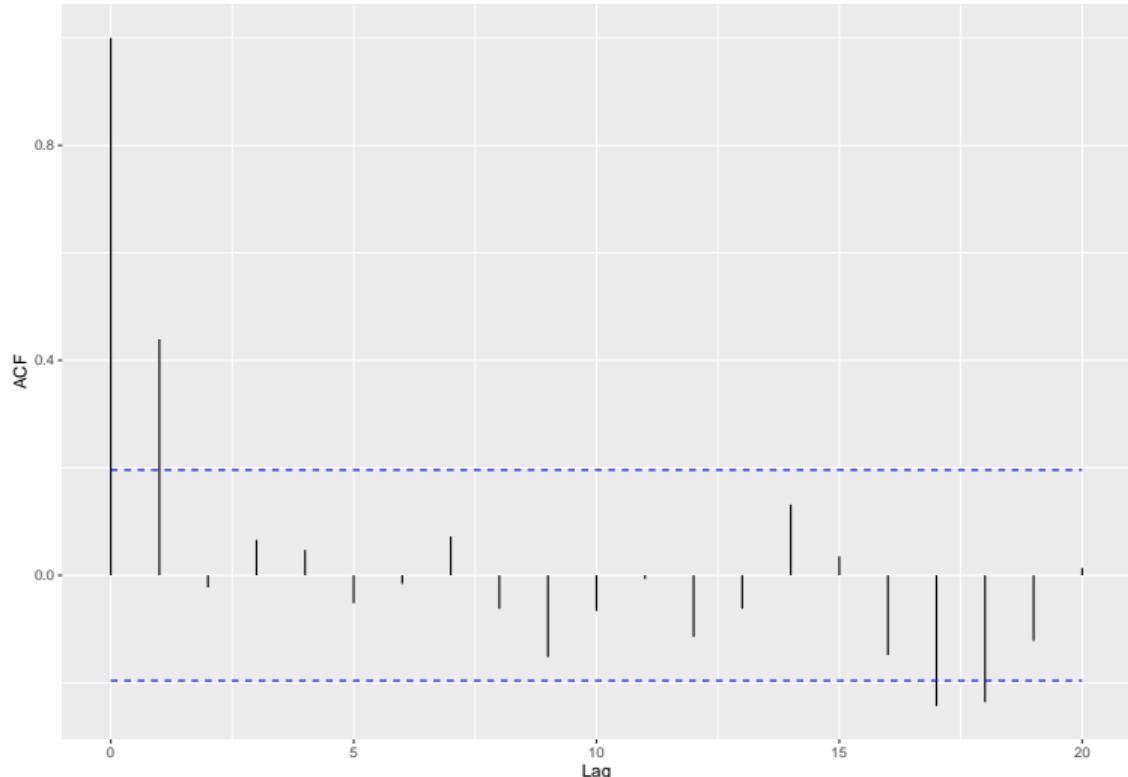
que al escribir un $MA(1)$ en forma autorregresiva, hay un efecto directo de X_{t-k}^* sobre X_t^* de magnitud θ^k , para cualquier k . Por lo que, la función de autocorrelación parcial tendrá todos los coeficientes no nulos y que decrecen geométricamente con k (que es la estructura de una función de autocorrelación de un $AR(1)$). Así, la función de autocorrelación parcial de un $MA(1)$ tiene la misma estructura que la función de autocorrelación de un $AR(1)$.

Simulación de un MA(1)

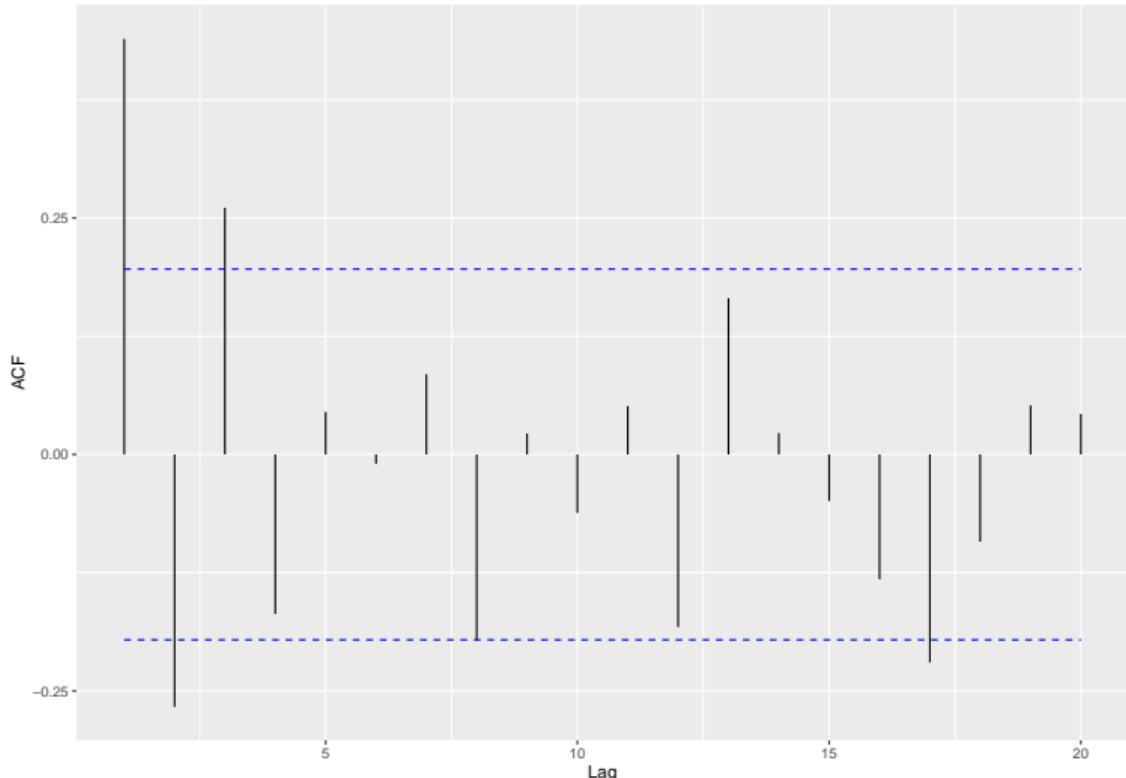
```
ma1 <- arima.sim(n = 100, list(ma=0.8))
plot_4<-autoplot(ma1, ts.colour = 'black', main = 'MA(1) theta=0.8')
```



```
autofplot(acf(ma1, plot=FALSE))
```



```
autoplott(pacf(ma1, plot=FALSE))
```



MA(q)

Generalizando la idea de un $MA(1)$, podemos escribir procesos cuyo valor actual dependa no sólo de la última innovación, sino de las q últimas innovaciones. Un proceso $MA(q)$ está definido como:

$$X_t^* = Z_t - \theta_1 Z_{t-1} - \theta_2 Z_{t-2} - \cdots - \theta_q Z_{t-q}$$

o equivalentemente, en notación de operador de lag:

$$X_t^* = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q) Z_t$$

que lo escribimos de manera más compacta como:

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

Un proceso $MA(q)$ es siempre estacionario, por ser la suma de procesos estacionarios.

Diremos que el proceso es **invertible** si las raíces del operador $\theta(B) = 0$, son (en módulo) mayores que la unidad.

Las propiedades de este proceso las obtenemos a partir de

$$X_t^* = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q) Z_t$$

Multiplicando convenientemente.

Función de autocovarianza

Multiplicando la ecuación anterior

$$X_t^* = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q) Z_t$$

por X_{t-k}^* para $k \geq 0$ y tomando esperanzas, se obtienen las siguientes autocovarianzas:

$$\gamma(0) = (1 + \theta_1^2 + \cdots + \theta_q^2)\sigma^2$$

$$\gamma(k) = (-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q)\sigma^2 \quad k = 1, \dots, q$$

$$\gamma(k) = 0 \quad k > q$$

Por lo que un proceso $MA(q)$ tiene exactamente los q primeros coeficientes de la función de autocovarianzas distintos de cero.

Función de autocorrelación

Dividiendo las covarianzas por $\gamma(0)$ y utilizando una notación más compacta, la función de autocorrelación será:

$$\rho_k = \frac{\sum_{i=0}^{i=q} \theta_i \theta_{k+i}}{\sum_{i=0}^{i=q} \theta_i^2}, \quad k = 1, \dots, q$$
$$\rho_k = 0, \quad k > q,$$

donde $\theta_0 = -1$, y $\theta_k = 0$ para $k \geq q + 1$. En particular, para $q = k = 1$ obtenemos la autocorrelación para el $MA(1)$.

Función de autocorrelación parcial

Para calcular la función de autocorrelación parcial de un $MA(q)$ expresaremos el proceso como un $AR(\infty)$:

$$\theta^{-1}(B)X_t^* = Z_t$$

y llamando $\theta^{-1}(B) = \pi(B)$, donde:

$$\pi(B) = 1 - \pi_1 B - \cdots - \pi_k B^k - \dots$$

y los coeficientes de $\pi(B)$ se obtienen imponiendo que $\pi(B)\theta(B) = 1$.

Suponer que el proceso es invertible implica que las raíces de $\theta(B) = 0$ están fuera del círculo unitario, y la serie $\pi(B)$ será convergente. Igualando las potencias de B a cero, se obtiene que los coeficientes π_i verifican la siguiente relación:

$$\pi_k = \theta_1 \pi_{k-1} + \cdots + \theta_q \pi_{k-q}$$

donde $\pi_0 = -1$ y $\pi_j = 0$ para $j < 0$. La solución de esta ecuación será de la forma $\sum A_i G_i^k$, donde ahora los G_i^{-1} son las raíces del operador de media móvil.

Tras obtener los coeficientes π_i de la representación $AR(\infty)$, podemos escribir el proceso MA como:

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

De la expresión anterior, podemos concluir que la función de autocorrelación parcial de un *MA* será no nula para todo retardo, ya que existe un efecto directo de X_{t-i}^* sobre X_t^* para todo i .

La función de autocorrelación parcial de un proceso *MA* tendrá la misma estructura que la función de autocorrelación de un proceso *AR* del mismo orden.

Así, existe una dualidad entre procesos *AR* y *MA*, de manera que la función de autocorrelación parcial de un *MA*(q) tiene la estructura de la función de autocorrelación de un *AR*(q), y la función de autocorrelación de un *MA*(q) tiene la estructura de la función de autocorrelación parcial de un *AR*(q).

MA(∞), Descomposición de Wold

Los procesos autorregresivos y de media móvil estudiados antes son casos particulares de una representación general de procesos estacionarios obtenida por Wold (1983), quien mostró que todo proceso estocástico débilmente estacionario, X_t , de media finita μ , que no contenga componentes deterministas, puede escribirse como una función lineal de variables aleatorias no correlacionadas, Z_t , como:

$$X_t = \mu + Z_t + \psi_1 Z_{t-1} + \psi_2 Z_{t-2} + \dots = \mu + \sum_{i=0}^{\infty} \psi_i Z_{t-i}, \quad (\psi_0 = 1)$$

donde $\mathbb{E}(X_t) = \mu$, $\mathbb{E}(Z_t) = 0$, $\mathbb{V}(Z_t) = \sigma^2$ y $\mathbb{E}(Z_t Z_{t-k}) = 0$, $k > 1$

Si llamamos, nuevamente, $X_t^* = X_t - \mu$ y utilizamos el operador de lag, podemos escribir:

$$X_t^* = \psi(B)Z_t$$

siendo $\psi(B) = 1 + \psi_1B + \psi_2B^2 + \dots$ un polinomio indefinido en el operador de retardo B . Llamaremos a la ecuación anterior la representación lineal general de un proceso estacionario no determinista.

Esta representación es importante porque nos garantiza que cualquier proceso estacionario admite una presentación lineal.

En general, las variables Z_t forman un proceso de ruido blanco, esto es: no correlacionadas, con media cero y varianza constante. En ciertos casos particulares el proceso puede escribirse en función de variables $\{Z_t\}$ normales independientes. Entonces la variable X_t^* tendrá una distribución normal, y la estacionariedad débil coincidirá con la estricta.

Las propiedades del proceso se obtienen como en los casos de un modelo MA . La varianza de X_t es:

$$X_t = \mu + Z_t + \psi_1 Z_{t-1} + \psi_2 Z_{t-2} + \dots = \mu + \sum_{i=0}^{\infty} \psi_i Z_{t-i}, \quad (\psi_0 = 1)$$

estará dada por:

$$\mathbb{V}(Z_t) = \gamma(0) = \sigma^2 \sum_{i=0}^{\infty} \psi_i^2$$

y para que el proceso tenga varianza finita, la serie $\{\psi_i^2\}$ debe ser convergente. Notar que si los coeficientes ψ_i se anulan a partir de un retardo q , el modelo general se reduce a un $MA(q)$ y la fórmula anterior coincide con la obtenida anteriormente.

Las covarianzas se obtienen con

$$\gamma(k) = \mathbb{E}(X_t^* X_{t-k}^*) = \sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}$$

que para $k = 0$ proporcionan, como caso particular, $\gamma(0)$. Nuevamente, si los coeficientes ψ_i se anulan a partir de un retardo q , esta expresión proporciona las autocovarianzas de un proceso $MA(q)$. Los coeficientes de autocorrelación vendrán dados por:

$$\rho(k) = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+k}}{\sum_{i=0}^{\infty} \psi_i^2}$$

que generaliza la expresión de autocorrelación de un $MA(q)$.

Una consecuencia de la representación de Wold, es que todo proceso estacionario admite también una representación autorregresiva, que puede ser de orden infinito. Esta representación es la inversa de la de Wold, y escribiremos:

$$X_t^* = \pi_1 X_{t-1}^* + \pi_2 X_{t-2}^* + \cdots + Z_t$$

que en notación de operadores de lag, se reduce a:

$$\pi(B)X_t^* = Z_t$$

La representación $AR(\infty)$ es la dual de la $MA(\infty)$ y se verifica:

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

con lo que, igualando las potencias de B a cero, podemos obtener los coeficientes de una representación a partir de los de la otra.

AR, MA y proceso general

Es inmediato comprobar que un proceso *MA* es un caso particular de la representación de Wold. Esto también lo es en los *AR*. Por ejemplo el proceso *AR(1)*:

$$(1 - \alpha B)X_t^* = Z_t$$

puede escribirse (al multiplicar por el operador inverso $(1 - \alpha B)^{-1}$) como:

$$X_t^* = (1 + \alpha B + \alpha^2 B^2 + \dots)Z_t$$

que representa al proceso *AR(1)* como un caso particular de la forma *MA*(∞) del proceso lineal general, con coeficientes ψ_i que decaen en progresión geométrica. La condición de estacionariedad y varianza finita, serie de coeficientes ψ_i^2 convergente, equivale ahora a que $|\alpha| < 1$.

Para procesos *AR* de orden mayor, resulta más cómodo obtener los coeficiente de la representación *MA*(∞) imponiendo la condición de que, al ser resultado de invertir un proceso *AR*, el producto de ambos operadores deber la unidad. Por ejemplo para un *AR*(2) la condición será:

$$(1 - \alpha_1 B - \alpha_2 B^2)(1 + \psi B + \psi_2 B^2 + \dots) = 1$$

Imponiendo la anulación de las potencias de B , se obtienen los coeficientes:

$$\psi_1 = \alpha_1$$

$$\psi_2 = \alpha_1 \psi_1 + \alpha_2$$

$$\psi_i = \alpha_1 \psi_{i-1} + \alpha_2 \psi_{i-2} \quad i \geq 2$$

donde $\psi_0 = 1$.

Análogamente, para un $AR(p)$ los coeficientes ψ_i de la representación general se calculan mediante:

$$(1 - \alpha_1 B - \cdots - \alpha_p B^p)(1 + \psi_1 B + \psi_2 B^2 + \dots) = 1$$

y para $i \geq p$ deben verificar la condición:

$$\psi_i = \alpha_1 \psi_{i-1} + \cdots + \alpha_p \psi_{i-p} \quad i \geq p$$

La condición de estacionariedad implica que las raíces de la ecuación característica del proceso $AR(p)$, $\phi(B) = 0$, deben estar fuera del círculo unitario. Escribiendo el operador $\phi(B)$ como:

$$\phi(B) = \prod_{i=1}^p (1 - G_i B)$$

donde G_i^{-1} son las raíces de $\phi(B) = 0$, se verifica que, desarrollando en fracciones parciales:

$$\phi^{-1}(B) = \sum \frac{k_i}{(1 - G_i B)}$$

será convergente si $|G_i| < 1$.

En síntesis, los procesos *AR* pueden considerarse casos particulares de la representación del proceso lineal general que se caracterizan porque:

1. Todos los ψ_i son distintos de cero
2. Existen restricciones sobre los ψ_i , que dependen del orden del proceso. En general, verifican la secuencia $\psi_i = \alpha_1\psi_{i-1} + \cdots + \alpha_p\psi_{i-p}$, con condiciones iniciales que dependen del orden del proceso.

Esta relación entre los coeficientes ψ_i y los del proceso permiten concluir que los coeficientes ψ_i tienen la misma estructura que los coeficientes de autocorrelación simple. De hecho, para un $AR(p)$ hemos visto que satisfacen la ecuación de un proceso, como ocurría con las autocorrelaciones, y para un $MA(q)$ el número de coeficientes de autocorrelación no nulos es el orden del proceso.

Procesos ARMA

El proceso $ARMA(p, q)$ está definido como:

$$(1 - \alpha_1 B - \cdots - \alpha_p B^p) \tilde{X}_t = (1 - \theta_1 B - \cdots - \theta_q B^q) Z_t$$

o equivalentemente,

$$\phi_p(B) \tilde{X}_t = \theta_q(B) Z_t$$

El proceso será estacionario si las raíces de $\phi_p(B) = 0$ están fuera del círculo unitario, e invertible si lo están las de $\theta_q(B) = 0$. Supondremos además, que no hay raíces comunes que puedan cancelarse entre los operadores AR y MA .

Para obtener los coeficientes ψ_i de la representación general del modelo $MA(\infty)$, escribimos:

$$\tilde{X}_t = \phi_p(B)^{-1} \theta_q(B) Z_t = \psi(B) Z_t$$

e igualamos las potencias de B en $\psi(B)\phi_p(B)$ a las de $\theta_q(B)$. Análogamente, podemos representar un $ARMA(p, q)$ como un modelo $AR(\infty)$ haciendo:

$$\theta_q^{-q}(B)\phi_p(B)\tilde{X}_t = \pi(B)\tilde{X}_t = Z_t$$

y los coeficientes π_i resultarán $\phi_p(B) = \theta_q(B)\pi(B)$.

Para calcular las autocovarianzas, multiplicamos la forma general por \tilde{X}_{t-k} y tomamos esperanzas, quedando:

$$(1 - \alpha_1 B - \cdots - \alpha_p B^p) \tilde{X}_t \tilde{X}_{t-k} = \mathbb{E}(Z_t \tilde{X}_{t-k}) - \theta_1 \mathbb{E}(Z_{t-1} \tilde{X}_{t-k}) - \cdots - \mathbb{E}(Z_{t-q} \tilde{X}_{t-k})$$

para todo $k > q$, todos los términos de la derecha se anulan, y dividiendo por $\gamma(0)$:

$$\rho(k) - \alpha_1 \rho_{k-1} - \cdots - \alpha_p \rho(k-p) = 0$$

es decir:

$$\phi_p(B) \rho(k) = 0 \quad k > q$$

y concluimos que los coeficientes de autocorrelación para $k > q$ seguirán un decrecimiento determinado únicamente por la parte autorregresiva. Los primeros q coeficientes dependen de los parámetros MA y AR , y de ellos p proporcionarán los valores iniciales para el decrecimiento posterior (para $k > q$).

Así, si $p > q$ toda la función de autocorrelación mostrará un decrecimiento dictado por la ecuación anterior. En resumen, la función de autocorrelación:

1. Tendrá $q - p + 1$ valores iniciales con una estructura que depende de los parámetros AR y MA
2. Decrecerá a partir del coeficiente $q - p$ como una mezcla de exponenciales y sinusoides, determinada exclusivamente por la parte autorregresiva.

Para el caso de la autocorrelación parcial, se tendrá una estructura similar.

La función de autocorrelación simple y parcial de los procesos ARMA es el resultado de la superposición de sus propiedades *AR* y *MA*: en la función de autocorrelación simple ciertos coeficientes iniciales que dependen del orden de la parte *MA* y después un decrecimiento dictado por la parte *AR*.

En la función de autocorrelación parcial, valores iniciales dependientes del orden del *AR* seguidos del decrecimiento debido a la parte *MA*.

Esta estructura compleja hace que el orden un proceso ARMA sea difícil de indentificar en la práctica

	F.A. Simple	F.A. parcial
AR(p)	Muchos coeficientes no nulos	Primeros p no nulos, resto 0
MA (q)	Primeros q no nulos, resto 0	Muchos coeficientes no nulos
ARMA (p, q)	Muchos coeficientes no nulos	Muchos coeficientes no nulos

Procesos Integrados

La mayoría de las series con las que nos encontramos en la práctica no son estacionarias, y su nivel medio varía con el tiempo. Una de las formas que podemos solucionar este problema es diferenciar la serie, por ejemplo si tenemos nuestra serie de tiempo original X_t , podemos construir un nuevo modelo como:

$$W_t = \nabla X_t$$

esta serie tiende a oscilar alrededor de una media constante y ser estacionaria. Llamaremos a este tipo de series de tiempo **integradas**. Llamaremos orden de integración al número de diferencias necesarias para obtener un proceso estacionario.

Generalizando, diremos que un proceso es **integrado de orden** $h \geq 0$, y lo representaremos por $I(h)$, cuando al diferenciarlo h veces se obtiene un proceso estacionario.

Así, un proceso estacionario es siempre $I(0)$. En la práctica, la mayoría de las series no estacionario que son integradas tienen un orden $h \leq 3$.

Paseo aleatorio

Analizaremos nuevamente el proceso paseo aleatorio para exemplificar los procesos integrados. Sabemos que los procesos *MA* finitos son siempre estacionarios y que los *AR* sólo lo son si las raíces de $\phi(B) = 0$ están fuera del círculo unitario.

Consideremos el *AR(1)*:

$$X_t = c + \alpha X_{t-1} + Z_t$$

Si $|\alpha| < 1$ el proceso es estacionario. Si $|\alpha| > 1$ es fácil comprobar que se obtiene un proceso explosivo, donde los valores de la variable crecen sin límite hacia el infinito. Un caso interesante es cuando $|\alpha| = 1$, pues el proceso no es estacionario, pero tampoco es explosivo, y pertenece a la clase de procesos integrados de orden uno. Esto debido a que:

$$W_t = \nabla X_t = c + Z_t$$

Una característica importante que diferencia los procesos estacionarios y los no estacionarios es el papel de las constantes. En un proceso estacionario la constante no es importante. Sin embargo, en un proceso no estacionario las constantes, si existen, son muy importantes y presentan alguna propiedad permanente del proceso.

Procesos integrados de orden dos

Muchas series reales con tendencia pueden representarse con un modelo con dos diferencias, o integrado de orden dos. Un modelo simple que aparece en muchas aplicaciones es:

$$\nabla^2 X_t = (1 - \theta B) Z_t$$

Para justificar este modelo, supongamos un paseo aleatorio con una *deriva* que va cambiando en el tiempo, como:

$$\nabla X_t = c_t + u_t$$

donde la media de ∇X_t , que es el crecimiento de X_t , evoluciona en el tiempo. Sustituyendo sucesivamente en la ecuación anterior, suponiendo que el proceso comienza en $t = 0$ y que $X_0 = u_0 = 0$, se obtiene:

Supongamos que la evolución del coeficiente del crecimiento en cada instante, c_t , es suave, de manera que:

$$c_t = c_{t-1} + \varepsilon_t$$

donde ε_t es un proceso de ruido blanco, independiente de u_t . Entonces:

$$\begin{aligned}\nabla^2 X_t &= \nabla X_t - \nabla X_{t-1} \\ &= c_t + u_t - (c_{t-1} + u_{t-1}) \\ &= \varepsilon_t + u_t - u_{t-1} \\ &= (1 - \theta B) Z_t\end{aligned}$$

ya que la suma de un ruido blanco y un $MA(1)$ no invertible será un $MA(1)$ invertible.

Así, el proceso integrado de orden dos es una generalización del paseo aleatorio que permite que la deriva varíe suavemente en el tiempo. En general, los procesos integrados de orden dos puede verse como una generalización de los procesos integrados de orden uno, pero donde la pendiente de la recta de crecimiento en lugar de ser fija va cambiando con el tiempo.

Procesos integrados ARIMA

tbd