Series de tiempo no independientes

Karlo Jair Guevara Diaz

CIMAT & UG

2021

Introducción

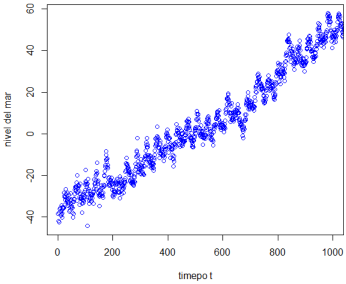
En la presentación veremos lo que son las series de tiempo, algunas características de algunas series de tiempo y algunos metodos para trabajar con las series de tiempo,

Series de tiempo

Una serie de tiempo es un proceso estocástico uninvariante que consiste en un conjunto de variables aleatoreas indexadas por el tiempo. Si hay $T \subset [0, \infty)$ variables, se denota por $\{X_t | t \in T\}$ o $\{X_t\}_{t \in T}$.

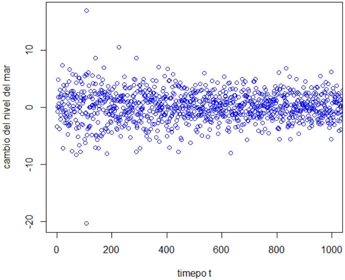
Un ejemplo de una serie de tiempo univariadas seria, el nivel medio global del mar al tiempo t. Por lo que X_t seria el nivel medio global del mar al tiempo t y $\{X_t(w)\}_{t\in[n]}$ (con w fijo) seria una muestra de la serie de tiempo, donde su gráfica esta dada por

Ejemplos de series de tiempo



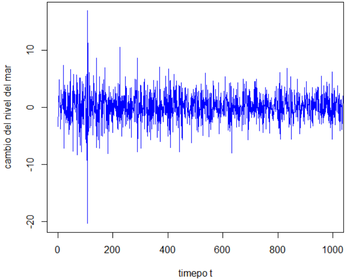
Otra serie de tiempo que podemos obtener de los mismos muestra o datos seria la de las diferencias, con esto nos referimos al aumento del nivel medio global del mar. Lo que se obtiene de la siguiente manera $X_t(w) = X_{t+1}(w) - X_t(w)$, donde su gráfica esta dada por

Ejemplos de series de tiempo



Ejemplos de series de tiempo

La forma de usual de representar estas gráficas es,



Procesos AR(p), AM(p) y ARMA(p,q)

Buscamos presentar los procesos AR(p), AM(q) y ARMA(p,q) para tratar con series de tiempo no independientes.

Ruido blanco

Una serie de tiempo muy útil y poco complicada es el ruido blanco, cuya definición esta dada por

Una serie de tiempo $\{a_t\}_{t \in T}$ se dice ruido blanco si cada a_t tiene media 0 y covarianza constante para toda $t \in T$.

Ejemplos de series de tiempo

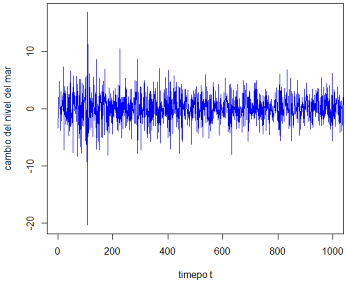


Figure:

Funcion de autocorrelación y correlogramas

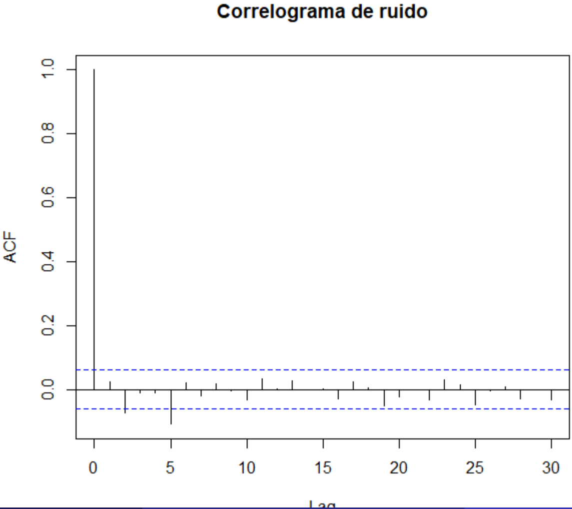
La función de autocorrelación esta dada por,

$$p(k) = \frac{cov(X_t, X_{t+k})}{\sqrt{V(X_t)V(X_{t+k})}}.$$

Esta función nos indica la dependencia lineal del que existe entre X_t y X_{t+k}.

Funcion de autocorrelación y correlogramas

El correlogramas es la gráfica de la función de autocorrelación. El siguiente es el correlograma del ruido blanco.



AR(p)

Una serie de tiempo $\{X_t\}_{t\in T}$ es un proceso $AR(p)$ si $X_t = c + \sum_{j=1}^{p} \theta_j X_{t-j} + a_t$, para constantes $c, \theta_1, .., \theta_j \in \mathbb{R}$ y a_t es ruido blanco para toda t.

Esperanza de AR(p)

Calculemos la esperanza de un proceso AR(p) estacionario. Notemos que

$$\begin{aligned} E(X_t) &= E(c + \sum_{j=0}^{p} \theta_j X_{t-j} + a_t) \\ &= E(c) + \sum_{j=0}^{p} \theta_j E(X_{t-j}) + E(a_t) \\ &= c + \sum_{j=0}^{p} \theta_j E(X_t) + 0, \end{aligned}$$

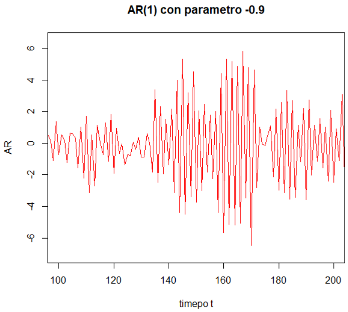
luego,

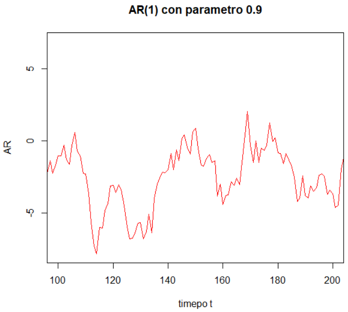
$$(1 - \sum_{j=0}^{p} \theta_j) E(X_t) = c$$

Por lo que, $E(X_t) = \frac{c}{1-\sum_{j=0}^{n} \theta_j}$

Ejemplo de AR(p)

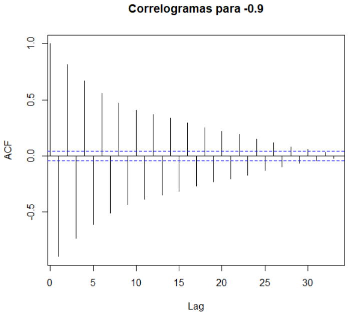
Presentamos 2 proceso AR(1), el primero es uno con $\theta = 0.9$ y con el segundo con $\theta = -0.9$, los dos con ruido blanco norma $(0, 1)$.

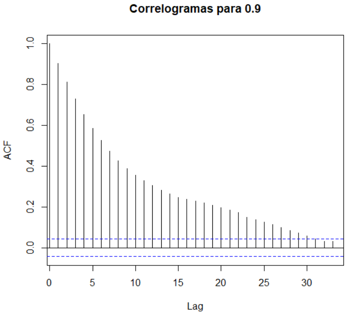




Ejemplo de AR(p)

Correlograma de los proceso AR anteriores.





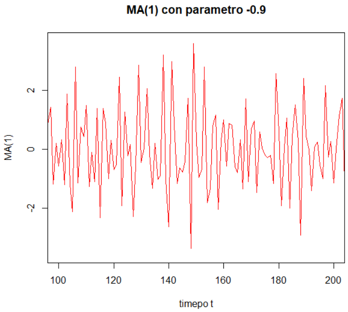
MA(q)

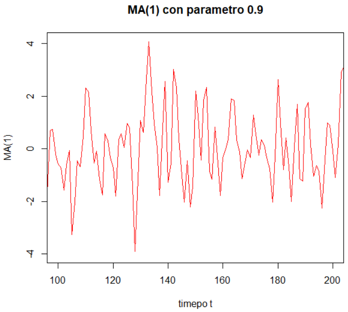
Un proceso $\{X_t\}_{t\in[n]}$ es un proceso de medias móviles de orden n si es de la forma $X_t = c + \sum_{j=1}^{n} \theta_j a_{t-j}$ donde a_t es ruido blanco y c es una constante.

$$\begin{array}{ccccccc}
Y_t & & Y_{t+1} & & Y_{t+2} & & \dots \\
\uparrow & \nearrow & \uparrow & \nearrow & \uparrow & \nearrow & \\
a_t & & a_{t+1} & & a_{t+2} & & \dots
\end{array}$$

Ejemplos de MA(q)

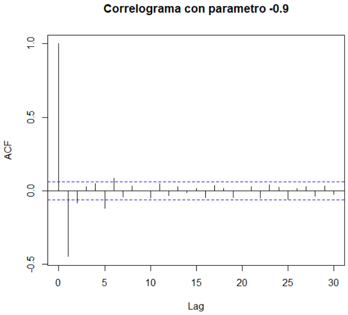
Presentamos 2 proceso MA(1), el primero es uno con $\theta = 0.9$ y con el segundo con $\theta = -0.9$, los dos con ruido blanco norma $(0, 1)$.

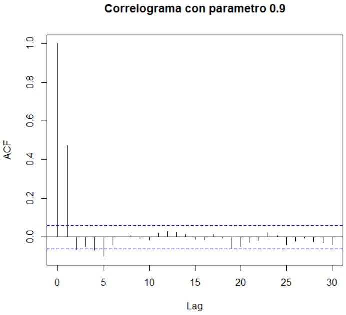




Ejemplos de MA(q)

Los correlogramas correspondientes a los procesos anteriores son,





ARMA(p,q)

Una serie de tiempo $\{X_t\}_{t \in T}$ es un proceso autorregresivo de medias moviles de orden (p,q) si es de la forma $X_t = \sum_{j=1}^{p} \theta_j X_{t-j} + a_t + \sum_{j=1}^{q} a_{t-j}\gamma_j + c$ donde $\theta_1, ..., \theta_p, \gamma_1, ..., \gamma_q, c \in \mathbb{R}$ y a_t es ruido balnco para todo t.

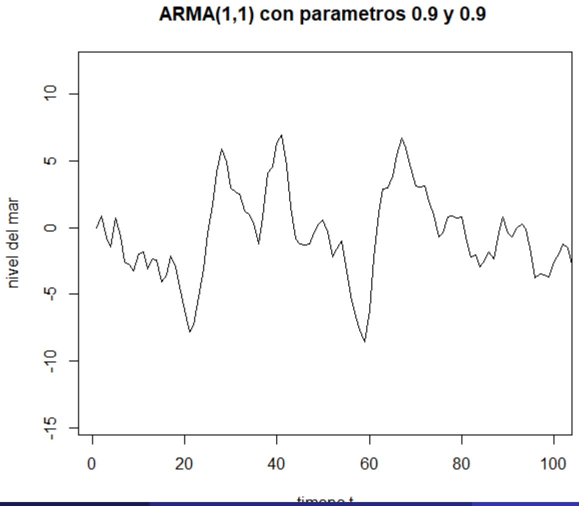
Esperanza del ARMA(p,q) y estacionalidad

Notemos que la esperanza del proceso esta dada por,

$$
\begin{aligned}
E(X_t) &= E(\sum_{j=1}^{p} \theta_j X_{t-j} + a_t + \sum_{j=1}^{q} a_{t-j}\gamma_j + c) \\
&= E(\sum_{j=1}^{p} \theta_j X_{t-j}) + E(a_t) + E(\sum_{j=1}^{q} a_{t-j}\gamma_j) + E(c) \\
&= E(\sum_{j=1}^{p} \theta_j X_{t-j}) + c \\
&= E(AR(p))
\end{aligned}
$$

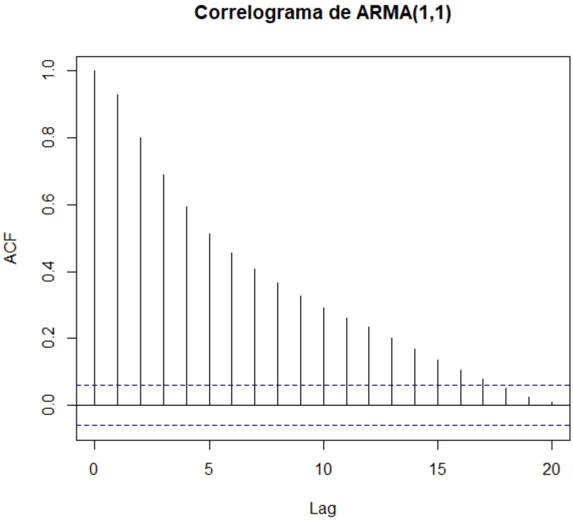
Ejemplo de proceso ARMA(1,1)

Simulando un proceso ARMA(1,1) con parámetros $\theta = 0.9$ y $\gamma = 0.9$, ruido blanco con distribución normal (0,1), tenemos la siguiente imagen



Ejemplo de proceso ARMA(1,1)

El correlograma del proceso anterior estara dado por,



Como usarlo

No todas las series de tiempo se pueden majar usando los proceso aquí presentado por lo que daremos algunas características necesarias para poder usar estos procesos.

Estacionariedad estricta

Una serie de tiempo $\{X_t\}_{t\in T}$, se dice que es estacionaria en el sentido estricto si $(X_{t_1},...,X_{t_n})$ y $(X_{t_1+k},..,X_{t_n+k})$ son iguales en distribución para cualquier $n, t_1,...,t_n \in \mathbb{N}$.

Estacionalidad en covarianza

Una serie de tiempo se dice, $\{X_t\}_{t\in T}$ se dice estacionaria en covarianza si $E(X_t) = \mu$ para algún $\mu \in \mathbb{R}$ y para cualquier $t \in T$, $V(X_t) = \sigma < \infty$ y $cov(X_t, X_s) = E(X_t - \mu)E(X_s + \mu) = \gamma_k < \infty$ para cualquier $k = |t - s|$.

Estacionario de AR(p)

Un proceso autorregresivo de orden p es estacionario si , y solo si, el modulo de las raíz del polinomio $1 - \theta_1 L - \theta_2 L^2 - ... - \theta_p L^p$ esta fuera del circulo unitario (dicho polinomio se conoce como el polinomio auto regresivo).

$$
\begin{aligned}
X_t &= \theta X_{t-1} + a_t \\
&= \theta X_{t-2} + \theta a_{t-1} + a_t \\
&= \theta X_{t-3} + \theta^2 a_{t-2} + \theta a_{t-1} + a_t
\end{aligned}
$$

No anticipante e invertible

Que el proceso sea no anticipante, es decir, que el presente no venga determinado por el futuro

Que el proceso sea invertible, es decir, que el presente dependa de forma convergente de su propio pasado

Ajuste a nuestros datos

Tomemos los datos de la serie de tiempo del nivel medio global del mar a tiempo t.

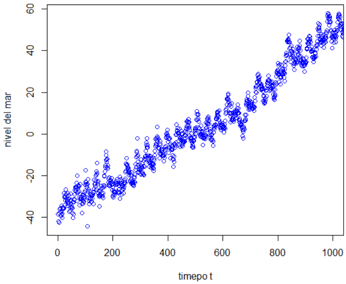


Figure:

Ajuste a nuestros datos

Usando el comando "ARMA = auto.arima(datos)" de R, obtenemos el siguiente resultado

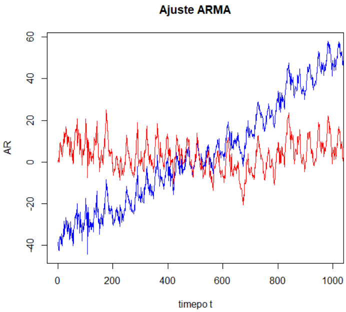


Figure:

Ajuste a nuestros datos

Por otro lado tenemos la serie de tiempo de la diferencias de la serie de tiempo anterior.

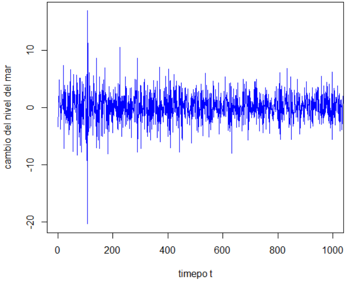


Figure:

Esta serie si cumple con la estacionalidad.

Usando la función "acf(dif)" obtenemos el correlogramas de los datos,

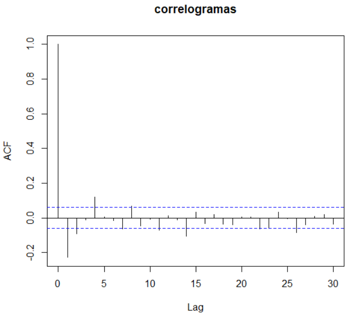


Figure:

por lo que la condición de se invertible se cumple.

Ajuste a nuestros datos

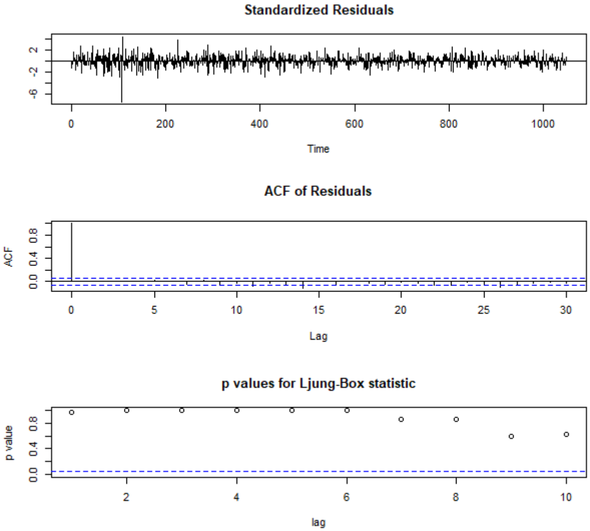
Usando el comando "ajuste = auto.arima(dif)" de R y "summary(ajuste)", obtenemos lo siguiente

```
Series: dif.ts
ARIMA(4,0,1) with non-zero mean

Coefficients:
         ar1      ar2     ar3     ar4      ma1    mean
      0.1195  -0.0478  0.0160  0.1276  -0.3904  0.0856
s.e.  0.1688   0.0540  0.0405  0.0311   0.1685  0.0623

sigma^2 estimated as 6.784:  log likelihood=-2487.34
AIC=4988.69   AICc=4988.79   BIC=5023.37
```

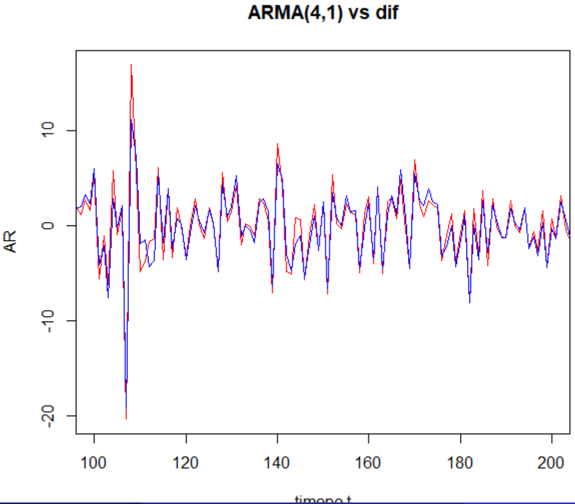
lo que nos indica que el mejor proceso ARMA(p,q) que ajusta los parámetros es ARMA(4,1).

Ajuste a nuestros datos

Usando el comando "tsdiag(ajuste)", obtenemos



Comparando el proceso ARMA(4,1) que generamos con los datos originales, obtenemos.



Para modelar los valores máximos de las series de tiempo, podemos usar teoría de extremos en las series de tiempo aun que tengan dependencia (bajo algunas condiciones)

Condiciones D y D'

Sea $\{X_t\}_{t\in T}$ una succión de variables aleatorias. Se dice que la condición D se cumple si para cualquier enteros positivos $i_1, ..., i_p$ y $j_1, ..., j_q$ tales que $|j_1 - i_p| \geq k$ y cualquier $u \in \mathbb{R}$, se cumple que

$$|F_{i_1,...,i_p,j_1,...,j_q}(u) - F_{i_1,...,i_p}(u)F_{j_1,...,j_q}(u)| \leq g(k)$$

donde g es tal que $g(k) \to 0$ si $k \to \infty$ y
$F_{s_1,...,s_p}(u) = P(X_{s_1} \leq u, ..., X_{s_r} \leq u)$ para cualquier $s_1, ..., s_r \in T$.

Condiciones D y D'

Sea $\{X_t\}_{t\in T}$ una susecion de variables aleatorias y $\{u_n\}_{n\geq 0}$. Se dice que $\{X_t\}_{t\in T}$ cumple la condición $D'(u_n)$ si

$$\lim_{k\to\infty}(\limsup_{n\to\infty} n\sum_{j=2}^{[\frac{n}{k}]} P(X_1 > u_n, X_j > u_n)) = 0$$

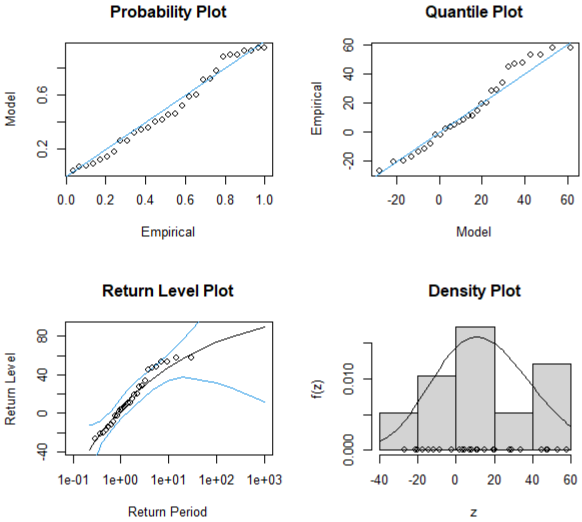
Sea $\{X_t\}_{t\geq 0}$ una succión de variables aleatorias. Si $\{u_n\}_{n\geq 0}$ es una succión tal que $\{X_t\}_{t\geq 0}$ cumple las condiciones $D(u_n)$ y $D'(u_n)$, entonces

$$\lim_{n\to\infty} P(N_n \leq u_n) = e^{-\tau}$$

para $\tau < \infty$.

Supongamos que queremos calcular el tiempo en el cual el valor de datos, estar nuevamente por encima de 50 y la probabilidad de que la diferencia máxima entre dos valores consecutivos de nuestros datos, exceda 16.8. Usando el comando "c = gev.fit(dat)" de R, obtenemos los parámetros "5.1551390, 23.7913723, -0.2217853". Lo que nos indica que una distribución Weibull de extremos es la que mejor ajusta a los máximos de los datos.

Al usar el comando "gev.diag(c)", obtenemos



Usando el comando "pgev" con los anteriores parámetros y evaluando en 50 obtenemos que $0.9166056 = \text{pgev}(50, 5.155, 23.791, -0.221)$ i.e la probabilidad de tener un dato menor a 50 es 0.08339443, lo que implica que la probabilidad de que un dato sea mayor a 50 es 0.08339443 y por ende el valor esperado de retorno es 11.99121. Por otro lado tendriamos que ' 0.01001746 = pgev(16.8, ,5.155, 23.791, -0.221)' optenemos que la probabilidad de que una diferencia maxima exeda 16.8 es 0.01001746.

Bibliografía

" Extremes and Related Properties of Random Sequences and Processes" deM. R. Leadbetter, G. Lindgren, H. Rootzen -

Análisis de series temporales: Modelos ARIMA de María Pilar González Casimiro

The Newsletter of the R Project Volume 2/2, June 2002

"Series de tiempo: Modelos Heterocedásticos. Aplicación a una serie económica usando R" de Yeison López Lizcano

" Notas de extremo " de Dr.Ehyter M. Martin Gonzalez

Extra

"auto.arima(dif.ts, method = "CSS")", "datos = ts(datos)", "library(forecast)" auto.arima, "library(GEVcdn)" y "library(ismev)"