

SIMULACIÓN DE UN PROCESO DE MEDIAS MÓVILES

Nicolás Villegas Vargas, *Estudiante, Universidad EAFIT*, y María Camila Vásquez
Correa, *Estudiante, Universidad EAFIT*,

Resumen—En el marco de la identificación de sistemas, se proponen inicialmente y de manera exploratoria modelos lineales que pueden explicar un conjunto de sucesos sujetos a ciertos parámetros. Los modelos de series de tiempo son algunos de ellos, que permiten modelar el comportamiento de este tipo de datos teniendo en cuenta su historia e innovaciones que suceden en el proceso. En este informe se muestran las generalidades de los procesos de media móvil, su formulación matemática y varias simulaciones de estos en el programa MATLAB.

Palabras clave — MATLAB, Modelación Experimental, Media Móvil, Ruido Blanco, Series de tiempo.

I. INTRODUCCIÓN

La teoría de modelos de series temporales y los métodos se ilustran con muchos ejemplos y varios casos de estudio con problemas y datos que surgen de una variedad de campos aplicados, incluyendo procesamiento de señales, estudios biomédicos y finanzas [1].

La expresión serie de tiempo de datos, o serie de tiempo, generalmente se refiere a un conjunto de observaciones recogidas secuencialmente en el tiempo. Estas observaciones podrían haber sido recogidas en puntos de tiempo igualmente espaciados. En este caso utilizamos la notación Y_t para $t = 0, 1, \dots$

El objetivo de este trabajo es analizar uno de los modelos que ayudan a capturar las dinámicas de una serie de tiempo, llamado proceso de media móvil, que modela las series sin tener en cuenta su historia anterior, sino la historia de las innovaciones ocurridas en las series, o el *ruido*. Para ello debemos:

- Realizar la formulación matemática del modelo, que nos permita identificar sus características teóricas.
- Simular el proceso de media móvil con diferentes parámetros en el programa MATLAB.
- Analizar los resultados de las diferentes simulaciones y compararlos con los resultados teóricos obtenidos.

El trabajo está dividido de la siguiente manera: en la primera sección se introduce la formulación matemática del problema, más tarde se muestran las simulaciones realizadas, su análisis, las conclusiones que se sacaron de el trabajo y por último las referencias.

II. DEFINICIÓN DEL PROBLEMA

Sea $\varepsilon_t \sim N(0, \sigma^2)$ un ruido blanco y considere el proceso:

$$Y_t = \mu + \varepsilon_t + d_1 \varepsilon_{t-1} + d_2 \varepsilon_{t-2} + \dots + d_q \varepsilon_{t-q} \quad (1)$$

Donde μ y d_i para $i = 1, \dots, q$ son constantes. Esta serie de tiempo es el proceso de media móvil de orden q , denotado

$MA(q)$. El término media móvil viene del hecho de que Y_t es construida mediante una suma ponderada, similar a un promedio, de los q valores más recientes de ε_t [2].

II-A. Supuestos

Las variables aleatorias ε_{t-j} y ε_{t-i} son independientes, idénticamente distribuidas y no correlacionadas. Es decir, $E(\varepsilon_i \varepsilon_j) = 0$ para todo i diferente de j . Además, cada ε_t sigue una distribución normal con media cero y varianza constante.

Denotaremos la autocovarianza de un proceso como

$$\gamma_j = \text{Cov}(Y_t, Y_{t-j})$$

Así, $\gamma_0 = V[Y_t]$.

II-B. Momentos

El valor esperado de Y_t está dado por:

$$\begin{aligned} E[Y_t] &= E(\mu + \varepsilon_t + \sum_{i=1}^q d_i \varepsilon_{t-i}) \\ &= \mu + E[\varepsilon_t] + d_1 E[\varepsilon_{t-1}] + \dots + d_q E[\varepsilon_{t-q}] \\ &= \mu \end{aligned}$$

Y la varianza, dado que las innovaciones no son correlacionadas:

$$\begin{aligned} \gamma_0 &= E[Y_t - \mu]^2 \\ &= E[(\varepsilon_t + \sum_{i=1}^q d_i \varepsilon_{t-i})^2] \\ &= (1 + \sum_{i=1}^q d_i^2) \sigma^2 \end{aligned} \quad (1)$$

Ahora, para calcular las autocovarianzas, para $j = 1, 2, \dots, q$, tendremos que:

$$\begin{aligned} \gamma_j &= E[(\varepsilon_t + d_1 \varepsilon_{t-1} + d_2 \varepsilon_{t-2} + \dots + d_q \varepsilon_{t-q}) \\ &\quad \times (\varepsilon_{t-j} + d_1 \varepsilon_{t-j-1} + d_2 \varepsilon_{t-j-2} + \dots + d_q \varepsilon_{t-j-q})] \\ &= E[d_j \varepsilon_{t-j}^2 + d_{j+1} d_1 \varepsilon_{t-j-1}^2 + \dots + d_q d_{q-j} \varepsilon_{t-q}^2] \end{aligned}$$

Los términos que tenían ε 's en diferentes fechas se vuelven cero porque su producto tiene valor esperado cero (son no correlacionadas). Si $j > q$, tendremos que no hay ε 's con fechas comunes en la definición de γ_j , y por tanto ese valor esperado es también cero, así:

$$\gamma_j = \begin{cases} [d_j + d_{j+1} d_1 + \dots + d_q d_{q-j}] \sigma^2 & j = 1, \dots, q \\ 0 & j > q \end{cases} \quad (2)$$

Podemos ver entonces, que para cualquier valor de los parámetros del modelo, el proceso es estacionario en media y autocovarianza, pues ninguna depende del tiempo.

II-C. Invertibilidad del proceso

Vimos que el proceso de media móvil es siempre estacionario. Ahora, consideremos el mismo proceso expresado con el operador de rezagos:

$$(Y_t - \mu) = (1 + d_1 L + d_2 L^2 + \dots + d_q L^q) \varepsilon_t \quad (3)$$

Donde $L(\varepsilon_t) = \varepsilon_{t-1}$. Si las raíces del polinomio de rezagos:

$$(1 + d_1 z + d_2 z^2 + \dots + d_q z^q) = 0 \quad (4)$$

caen fuera del círculo unidad, la Ecuación 3 puede ser escrita como un proceso autorregresivo de orden infinito, es decir, un $AR(\infty)$, invirtiendo el operador del MA:

$$(1 + \eta_1 L + \eta_2 L^2 + \dots)(Y_t - \mu) = \varepsilon_t \quad (5)$$

Donde

$$(1 + \eta_1 L + \eta_2 L^2 + \dots) = (1 + d_1 z + d_2 z^2 + \dots + d_q z^q)^{-1} \quad (6)$$

Si se desea abordar con mayor detalle los aspectos teóricos de los procesos MA se recomienda revisar la sección 3.3 del libro de Hamilton [2].

III. EXPERIMENTACIÓN COMPUTACIONAL

Con el objetivo de analizar el comportamiento del proceso de medias móviles, se ha implementado un código en MATLAB y se han realizado varias simulaciones. En estas simulaciones el proceso de ruido blanco, ε_t tiene media cero y varianza 1. En la figura 1 se han simulado 4 procesos de medias móviles de orden 1 y cada realización del proceso se denota como Y_i para $i = 1, 2, 3, 4$. Del mismo modo en las figuras 2, 3 y 4 se simulan los procesos MA(2), MA(3) y MA(4), respectivamente. Por otro lado, en las tablas I, II, III, IV se puede apreciar los diferentes valores que toman los parámetros en cada uno de los modelos. Es importante resaltar que en todas las simulaciones se utiliza un tiempo $t = 800$.

Cuadro I
PARÁMETROS PARA CADA REALIZACIÓN DEL MODELO MA(1)

	d_1	μ
Y_1	0.5	0
Y_2	-0.5	10
Y_3	2	20
Y_4	-2	30

Cuadro II
PARÁMETROS PARA CADA REALIZACIÓN DEL MODELO MA(2)

	d_1	d_2	μ
Y_1	0.5	0.2	0
Y_2	0.5	-0.2	10
Y_3	-0.5	-0.2	20
Y_4	-0.5	0.2	30

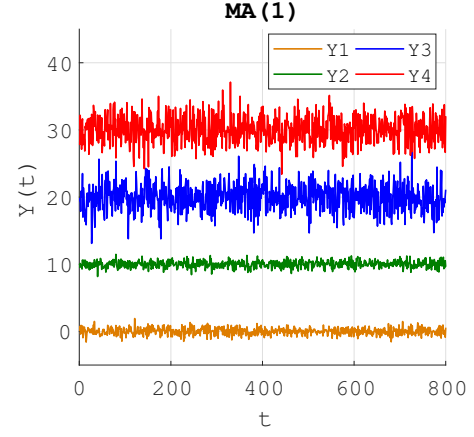


Figura 1. Simulación de 4 diferentes realizaciones del proceso MA(1)

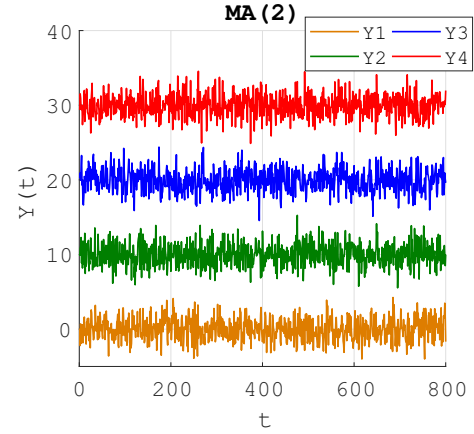


Figura 2. Simulación de 4 diferentes realizaciones del proceso MA(2)

Cuadro III
PARÁMETROS PARA CADA REALIZACIÓN DEL MODELO MA(3)

	d_1	d_2	d_3	μ
Y_1	0.23	0.87	0.92	0
Y_2	-0.23	0.87	0.92	10
Y_3	0.23	1.87	0.92	20
Y_4	0.23	0.87	-0.92	30

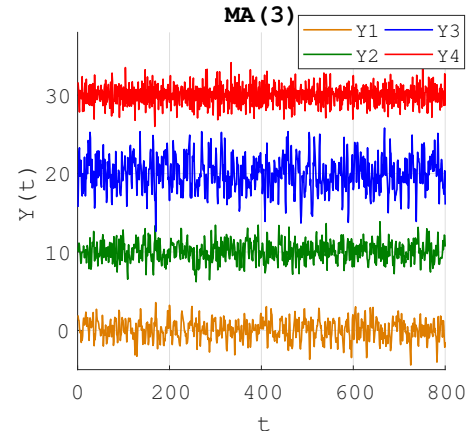


Figura 3. Simulación de 4 diferentes realizaciones del proceso MA(3)

Cuadro IV
PARÁMETROS PARA CADA REALIZACIÓN DEL MODELO MA(4)

	d_1	d_2	d_3	d_4	μ
Y_1	0.23	0.87	0.92	0.00001	0
Y_2	0.23	0.87	0.92	1	10
Y_3	0.23	0.87	0.92	2	20
Y_4	0.23	0.87	0.92	3	35

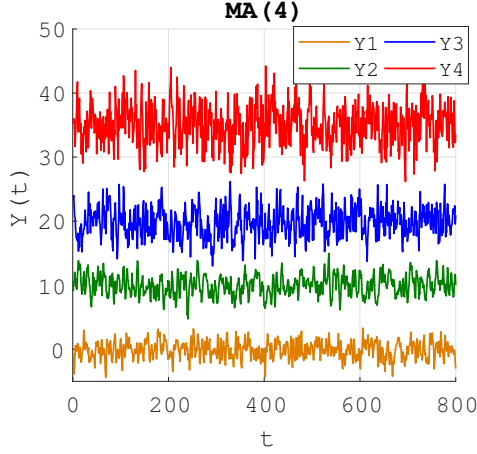


Figura 4. Simulación de 4 diferentes realizaciones del proceso MA(4)

Cuadro V
PARÁMETROS PARA CADA REALIZACIÓN DEL MODELO MA(3) EN DONDE EL RUIDO TIENE UNA DISTRIBUCIÓN DEFINIDA COMO LA COMBINACIÓN LINEAL DE DOS DISTRIBUCIONES NORMALES CON DIFERENTE VARIANZA Y AMBAS CON MEDIA CERO

	d_1	d_2	d_3	μ
Y_1	0.23	0.87	0.92	0
Y_2	-0.23	0.87	0.92	100
Y_3	0.23	1.87	0.92	200
Y_4	0.23	0.87	-0.92	315

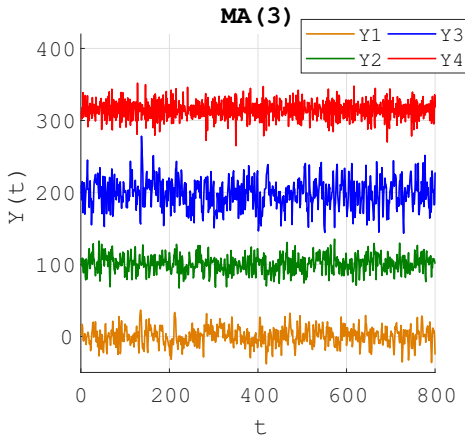


Figura 5. Simulación de 4 diferentes realizaciones del proceso MA(3) en donde el ruido tiene una distribución definida como la combinación lineal de dos distribuciones normales con diferente varianza y ambas con media cero

En la realización de las simulaciones presentadas en la figura 5 se utiliza la ecuación 1 pero en ella el término ε_t tiene una distribución definida como la combinación lineal de dos distribuciones normales con diferentes parámetros.

Específicamente $\varepsilon_t \sim 0.1 * N(0, 1) + 0.9 * N(0, 10), \forall t \in \mathbb{N}$.

En la siguiente sección se realiza un análisis de los resultados obtenidos en cada una de las simulaciones.

IV. ANÁLISIS DE RESULTADOS

El proceso de Medias Móviles de orden q consiste en una ponderación de un ruido blanco con sus q rezagos más recientes, razón por la cual en las figuras desde la 1 hasta la 4 se puede apreciar el movimiento errático del proceso y su cualidad de oscilar siempre en torno a μ , la media del proceso. Por ejemplo, en las simulaciones del proceso MA(3) en la figura 3 se observa que la realización Y_3 presenta picos más altos por lo cual intuitivamente se puede pensar que el proceso presenta mayor variabilidad. Veamos si lo anterior tiene un fundamento matemático. Las ecuaciones que representan a cada realización del proceso MA(3) son:

$$Y_1 = \varepsilon_t + 0.23\varepsilon_{t-1} + 0.87\varepsilon_{t-2} + 0.92\varepsilon_{t-3}$$

$$Y_2 = 10 + \varepsilon_t - 0.23\varepsilon_{t-1} + 0.87\varepsilon_{t-2} + 0.92\varepsilon_{t-3}$$

$$Y_3 = 20 + \varepsilon_t + 0.23\varepsilon_{t-1} + 1.87\varepsilon_{t-2} + 0.92\varepsilon_{t-3}$$

$$Y_4 = 30 + \varepsilon_t + 0.23\varepsilon_{t-1} + 0.87\varepsilon_{t-2} - 0.92\varepsilon_{t-3}$$

Luego, utilizando la ecuación 1 es posible calcular la varianza teórica para cada proceso teniendo en cuenta que $\sigma^2 = 1$. Entonces,

$$VAR(Y_1) = 2.6562$$

$$VAR(Y_2) = 2.6562$$

$$VAR(Y_3) = 5.3962$$

$$VAR(Y_4) = 2.6562$$

Es importante resaltar que como los parámetros de los procesos Y_1, Y_2 y Y_4 son iguales en valor absoluto tal como se puede apreciar en la tabla III, entonces sus varianzas teóricas son iguales.

Ahora vamos a calcular las varianzas de los procesos simulados utilizando el comando **var** de MATLAB. Entonces tenemos,

$$VAR(Y_1) = 1.7449$$

$$VAR(Y_2) = 1.7917$$

$$VAR(Y_3) = 4.5546$$

$$VAR(Y_4) = 1.7773$$

Un aspecto que resalta en el cálculo de las varianzas teóricas y muestrales es que son diferentes, y esto se debe a que en cada simulación se trabaja con un tiempo $t = 800$. Si el tiempo de simulación aumenta, entonces el resultado esperado es que la diferencia entre las varianzas teórica y muestral sea pequeña.

Ahora veamos si uno de los procesos MA(3) es invertible.

$$Y_3 = 20 + \varepsilon_t + 0.23\varepsilon_{t-1} + 1.87\varepsilon_{t-2} + 0.92\varepsilon_{t-3}$$

$$Y_3 - 20 = (1 + 0.23L + 1.87L^2 + 0.92L^3)\varepsilon_t$$

$$Y_3 - 20 = \Theta(L)\varepsilon_t$$

Tal como se vio en la sección II-C, para que el proceso Y_3 sea invertible es necesario que todas las raíces del polinomio

$\Theta(L)$ se encuentren dentro del círculo unitario. Pero al calcular las raíces obtenemos

$$\lambda_1 = -2.1513$$

$$\lambda_2 = 0.0593 + 0.7083i$$

$$\lambda_3 = 0.0593 - 0.7083i$$

pero como $|\lambda_1| = 2.1513 > 1$, entonces el proceso Y_3 no es invertible, por lo cual no puede representarse mediante un proceso $AR(\infty)$.

Al simular un proceso MA con un ruido que no distribuye normal, obtenemos el resultado que se muestra en la Figura V. Aquí, podemos ver que la diferencia en distribución hace que la varianza del proceso sea mucho más grande de lo que lo sería si la distribución fuera normal. Vemos que, teóricamente, la varianza debería estar alrededor de

$$VAR(Y_1) = 26.5620$$

$$VAR(Y_2) = 26.5620$$

$$VAR(Y_3) = 53.9620$$

$$VAR(Y_4) = 26.5620$$

pero en vez de eso, la varianza muestral de cada realización del proceso es

$$VAR(Y_1) = 166.0871$$

$$VAR(Y_2) = 133.9395$$

$$VAR(Y_3) = 373.5614$$

$$VAR(Y_4) = 153.6358$$

La notable diferencia entre la varianza teórica y la varianza muestral se debe a que para el cálculo de la varianza teórica se utilizó la ecuación 1 la cual está fundamentada en unos supuestos los cuales están siendo violados. Por lo tanto se puede apreciar la importancia que tiene el cumplimiento de los supuestos mencionados cuando se tiene la intención de ajustar datos reales mediante un proceso de medias móviles ya que la violación de alguno de estos puede acarrear resultados que se alejan considerablemente de los deseados.

V. CONCLUSIONES

Se realizaron diversas simulaciones que permitieron apreciar diferentes procesos de media móvil, con diferentes parámetros (orden, media y varianza). En estas simulaciones, las diferencias significativas dadas por los parámetros se deben a la varianza del proceso, y son notables cuando los parámetros cambian de manera considerable.

Algunas características teóricas del proceso de media móvil pudieron ser verificadas mediante las simulaciones, tales como el aumento de varianza al aumentar los parámetros, el hecho de que el proceso oscila alrededor de la media y los grandes picos que se presentan dado la aleatoriedad de las innovaciones que constituyen el proceso.

Este modelo podría ser útil para modelar procesos en los que la memoria a largo plazo del proceso no sea relevante, en los que se pueda considerar siempre las innovaciones, o los

agentes externos al proceso como único factor explicativo del fenómeno estudiado.

Es importante además siempre verificar la distribución del ruido asociado a un conjunto de datos que deseen ser explicados con este u otros modelos lineales en los que se considere un término aleatorio, dado que si no son normales, se violan los supuestos y el modelo no estaría bien construido.

REFERENCIAS

- [1] R. Prado and M. West, *Time series: modeling, computation, and inference*. Chapman and Hall/CRC, 2010.
- [2] J. D. Hamilton, "Time series analysis," *Economic Theory. II, Princeton University Press, USA*, pp. 625–630, 1995.