

# Una metodología para la predicción con modelos SARIMA Bayesianos

Daniel Dala  
Departamento de Estadística  
Universidad Nacional Autónoma de Honduras  
e-mail: daniel.dala@unah.hn

## ÍNDICE

|                    |   |   |
|--------------------|---|---|
| <b>I.</b>          | <b>Introducción</b>                     | 1 |
| <b>II.</b>         | <b>Definición del Problema</b>          | 1 |
| <b>III.</b>        | <b>Preliminares y Notación</b>          | 1 |
| III-A.             | Modelos ARIMA no estacionales . . . . . | 1 |
| III-B.             | Modelos ARIMA estacionales . . . . .    | 2 |
| III-C.             | Modelos SARIMA Bayesianos . . . . .     | 2 |
| <b>Referencias</b> |   | 2 |

## ÍNDICE DE FIGURAS

## ÍNDICE DE CUADROS

# Una metodología para la predicción con modelos SARIMA Bayesianos

**Resumen**—El resumen no solo hace referencia al trabajo reportado, también sintetiza el trabajo documentado en aproximadamente 200 palabras. Establece el propósito, reporta la información obtenida, provee conclusiones, y recomendaciones. En esencia, resume los puntos principales del estudio de forma adecuada y precisa. Es importante referirse a los resultados principales obtenidos en tiempo pasado al describir el trabajo realizado.

## I. INTRODUCCIÓN

En la introducción se describe de forma concisa pero con un poco más de detalle el trabajo realizado.

En uno de los párrafos de esta sección se hace especial énfasis en la contribución realizada como parte del trabajo reportado, considerando como punto de partida el problema central que motivó el proyecto de investigación y las soluciones obtenidas por el autor del documento como resultado del trabajo de investigación realizado.

Se presentan además descripciones breves del resto de las secciones del documento.

## II. DEFINICIÓN DEL PROBLEMA

Sea el proceso  $\{y_t\}$  con  $t = 1, 2, \dots, n$  una serie de tiempo cualquiera, generalmente se busca predecir con la mayor precisión posible los valores futuros  $y_{n+m}$  con  $m = 1, 2, \dots, h$  esto se puede lograr ajustando la serie  $\{y_t\}$  a un modelo de predicción SARIMA (Multiplicative seasonal autoregressive integrated moving average), sin embargo es necesario seguir un procedimiento en donde se proponga un modelo SARIMA que se ajuste a la serie luego verificar la precisión del modelo y por ultimo haga la predicción de los valores futuros, de este modo surge la metodología Box-Jenkins la cual muestra cada uno de estos pasos antes mencionados para hacer una predicción precisa.

No obstante si se quiere ajustar la serie a un modelo SARIMA Bayesiano no es factible usar el método Box-Jenkins por tanto el objetivo central de este estudio es proponer una nueva metodología para llevar a cabo el proceso de predicción en modelos SARIMA Bayesianos.

La nueva metodología surge como una iniciativa del *Bayesian Workflow* [2], en donde se expone los pasos a seguir para el análisis bayesiano de datos el cual maneja muy bien la estimación de parámetros de modelos y la incertidumbre en los datos.

Por ultimo, una vez establecido el nuevo método se realizarán tres diferentes pruebas en tres conjuntos de datos que miden el IPC en Honduras de 1980 al 2018, la tasa de cambio de divisas entre Alemania y Reino Unido de 1984 a 1991 y la afluencia de turistas en Australia de 1995 al 2015, cada uno de estos conjuntos se encuentran en el paquete *bayesforecast* [6],

finalmente con los resultados de dichas pruebas se demostrará la funcionalidad del nuevo método propuesto.

## III. PRELIMINARES Y NOTACIÓN

En este estudio denotaremos las series de tiempo como  $\{y_t\}$ . Una serie de tiempo se dice estacionaria si para cualquier colección finita de la serie, su distribución conjunta se mantiene constante en el tiempo o sea:

$$P(y_1, y_2, \dots, y_n) = P(y_{1+h}, y_{2+h}, \dots, y_{n+h})$$

no obstante, esta propiedad generalmente no ocurre por lo que usaremos la estacionariedad débil que implica una media y varianza constante a través del tiempo, esto es:

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

En dado caso que la serie presente alguna tendencia u oscilaciones periódicas constantes sobre la media del proceso (estacionalidad) se puede hacer una transformación en los datos llamada diferenciación denotada por el operador diferencia:

$$\nabla^d y_t = y_t - y_{t-1}$$

Diremos que una serie es diferenciada si al aplicar el operador diferencia se vuelve estacionaria, de manera similar la serie es de diferencias estacionales si al aplicar la diferenciación estacional:

$$\nabla_s^D y_t = y_t - y_{t-s}$$

donde  $s$  es el periodo estacional, se vuelve estacionaria. También diremos que una serie de tiempo es un ruido blanco si es iid y normalmente distribuida con media cero y varianza constante.

### III-A. Modelos ARIMA no estacionales

En el análisis de series de tiempo si combinamos los modelos autorregresivos, modelos de medias móviles y la diferenciación obtenemos los modelos ARIMA (Autoregressive Integrated Moving Average) no estacionales denotado por:

$$\text{ARIMA}(p, d, q)$$

en donde  $p, d$  y  $q$  son los ordenes del modelo autorregresivo, diferenciación y modelo de medias móviles respectivamente, de manera explicita se expresa como:

$$\nabla^d y_t = c + \sum_{i=1}^p \phi_i \nabla^d y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t$$

donde  $\varepsilon_t$  es ruido blanco.

### III-B. Modelos ARIMA estacionales

Los modelos ARIMA pueden modelar datos con estacionalidad esto se puede lograr agregando parámetros estacionales al modelo y lo denotaremos como:

$$\text{SARIMA } (p, d, q) \times (P, D, Q)$$

*SARIMA* (Multiplicative seasonal autoregressive integrated moving average) de manera explicita:

$$\begin{aligned} Z_t = & c + \sum_{i=1}^p \phi_i Z_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \sum_{k=1}^P \Phi_k Z_{t-km} \\ & + \sum_{w=1}^Q \Theta_w \varepsilon_{t-wm} + \varepsilon_t \\ Z_t = & \nabla_m^D \nabla^d y_t, \end{aligned}$$

donde  $\varepsilon_t$  es un ruido blanco, los parámetros  $(p, d, q)$  y  $(P, D, Q)$  representan el orden de la parte no estacional y la parte estacional del modelo respectivamente y  $m$  es el periodo de oscilación de la media en los datos.

### III-C. Modelos SARIMA Bayesianos

En base a la definición previa de un modelo ARIMA estacional podemos definir un Modelo SARIMA Bayesiano como:

$$\text{Modelo} \sim \text{SARIMA}(p, d, q) \times (P, D, Q)_m$$

$$\phi_i \sim \text{priori}_{\phi_i}, \quad i = 1, \dots, p$$

$$\theta_j \sim \text{priori}_{\theta_j}, \quad j = 1, \dots, q$$

$$\Phi_k \sim \text{priori}_{\Phi_k}, \quad k = 1, \dots, P$$

$$\Theta_w \sim \text{priori}_{\Theta_w}, \quad w = 1, \dots, Q$$

$$\mu_0 \sim \text{priori}_{\mu_0}$$

$$\sigma_0 \sim \text{priori}_{\sigma_0}$$

### REFERENCIAS

- [1] Hyndman, R.J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2. Accessed on February 17, 2021.
- [2] Andrew Gelman, Aki Vehtari, Daniel Simpson, Charles C. Margosian, Bob Carpenter, Yuling Yao, Lauren Kennedy, Jonah Gabry, Paul-Christian Bürkner, Martin Modrák (2020). *Bayesian Workflow*, <https://arxiv.org/abs/2011.01808>
- [3] Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). *Stan: A Probabilistic Programming Language*. Journal of Statistical Software, 76(1), 1 - 32. doi: <http://dx.doi.org/10.18637/jss.v076.i01>
- [4] Hyndman, R., & Khandakar, Y. (2008). *Automatic Time Series Forecasting: The forecast Package for R*. Journal of Statistical Software, 27(3), 1 - 22. doi: <http://dx.doi.org/10.18637/jss.v027.i03>
- [5] Robert H. Shumway, David S. Stoffer (2017). *Time Series Analysis and Its Applications* Fourth Edition.
- [6] Matamoros A., Cruz C., Dala A., Hyndman R., O'Hara-Wild M. (2021). **bayesforecast: Bayesian Time Series Modeling with Stan** version 1.0.1, <https://CRAN.R-project.org/package=bayesforecast>