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

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 recomendacionoes. En esencia, resume los punntos 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 $t = 1, 2, ..., n$ una serie de tiempo lineal cualquiera, habitualmente se busca predecir con la mayor precisión posible los valores futuros y_{n+m} con $m = 1, 2, ...h$ esto se puede lograr ajustando los datos a un modelo de predicción ARIMA no estacional [3] siendo procesos que en general crean un buen ajuste a los datos, por esta razón los estadísticos George Box y Gwilym Jenkins en 1970 propusieron el método Box-Jenkins para la predicción con dichos modelos el cual se basa en tres etapas iterativas: selección del modelo, estimación de parámetros y verificación de modelos, así como lo muestra Hyndman en [2].

De manera similar estos modelos también son capaces de describir datos estacionales, esto es agregando una componente estacional en el modelo el cual se denota como SARIMA (Multiplicative seasonal autoregressive integrated moving average) que se muestra en III-B, sin embargo podemos redefinir los modelos antes mencionados a un enfoque bayesiano así como se hace en III-C, por consiguiente es natural preguntarse qué método usar para la predicción a la hora de ajustar los datos a un Modelo SARIMA Bayesiano, lo que conlleva al problema central de este estudio. El objeto de este estudio es proponer una nueva metodología de predicción análoga al método de Box-Jenkins dirigido a estos modelos desde una perspectiva bayesiana.

Cabe destacar que la nueva metodología si bien es cierto es análoga al metodo Box-Jenkins, cada componente usará herramientas detalladas del *Bayesian Workflow* propuesto en [4] 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* [8], 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 lineales como $\{y_t\}$ para $t = 1, ..., n$. Diremos que un proceso $\{x_i\}$ con $i \in \mathbb{Z}$ tiene la propiedad de estacionariedad fuerte si:

$$F_X(x_{t_1}, x_{t_2}, ..., x_{t_n}) = F_X(x_{t_{1+\tau}}, x_{t_{2+\tau}}, ..., x_{t_{n+\tau}})$$

para $t \in \mathbb{Z}$ con $n \in \mathbb{N}$ y cualquier $\tau \in \mathbb{Z}$, esto es si para cualquier colección finita del proceso su distribución conjunta se mantiene constante en el tiempo. Por otro lado diremos que el proceso $\{x_i\}$ tiene la propiedad de estacionariedad débil si:

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

esto es que el proceso tiene una media y varianza constante a través del tiempo.

Diremos que una serie $\{y_t\}$ presenta estacionalidad si muestra alguna tendencia u oscilaciones periódicas constantes sobre la media del proceso, en este sentido para estabilizar la media de la serie y reducir la estacionalidad se aplica 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 ε_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:

$$Z_t = c + \sum_{i=1}^{p} \phi_i Z_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \sum_{k=1}^{P} \Phi_k Z_{t-km} + \sum_{w=1}^{Q} \Theta_w \varepsilon_{t-wm} + \varepsilon_t$$

$$Z_t = \nabla_m^D \nabla^d y_t,$$

donde ε_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:

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

$$\phi_i \sim priori_{\phi i}, \quad i = 1, ..., p$$

$$\theta_j \sim priori_{\theta j}, \quad j = 1, ..., q$$

$$\Phi_k \sim priori_{\Phi k}, \quad k = 1, ..., P$$

$$\Theta_w \sim priori_{\Theta w}, \quad w = 1, ..., Q$$

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

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

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] Rob J. Hyndman (2002). *Box-Jenkins modelling*, In: Informed Student Guide to Management Science, ed., Hans Daellenbach and Robert Flood, Thomson: London. https://robjhyndman.com/papers/BoxJenkins.pdf

[3] Rob J Hyndman (2002). *ARIMA processes*. In: Informed Student Guide to Management Science, ed., Hans Daellenbach and Robert Flood, Thomson: London. https://robjhyndman.com/papers/ARIMA.pdf

[4] Andrew Gelman, Aki Vehtari, Daniel Simpson, Charles C. Margossian, Bob Carpenter, Yuling Yao, Lauren Kennedy, Jonah Gabry, Paul-Christian Bürkner, Martin Modrák (2020). *Bayesian Workflow.*, https://arxiv.org/abs/2011.01808

[5] 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

[6] 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

[7] Robert H. Shumway, David S. Stoffer (2017). *Time Series Analysis and Its Applications* Fourth Edition.

[8] 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