
PRONÓSTICOS NO PARAMÉTRICOS BASADO EN FUNCIONES TIPO KERNEL CON ESTIMACIÓN ROBUSTA DEL ANCHO DE BANDA

Andrés Eugenio Castaño Cárdenas

Estudiante de Maestría en Matemáticas Aplicadas

Universidad EAFIT

Medellín - Colombia

acastan7@eafit.edu.co

Henry Laniado

Departamento de Ciencias Matemáticas

Universidad EAFIT

Medellín - Colombia

hlaniado@eafit.edu.co

5 de octubre de 2020

RESUMEN

Las series de tiempo no paramétricas presentan ventajas sobre aquellas paramétricas de aplicación habitual en la práctica, ya que las primeras no consideran supuestos sobre la información histórica de los datos. Sin embargo, en el tratamiento estadístico no paramétrico de datos temporales aún existen algunas brechas importantes tales como la selección del ancho de banda y la definición del coeficiente de Markov para el kernel seleccionado. En este trabajo, se propone una alternativa robusta para determinar el ancho de banda en series de tiempo no paramétricas que consideran funciones kernel dentro de su procesos de predicción y en este sentido una alternativa sencilla basada en la función de autocorrelación para la selección del coeficiente de Markov es estudiada, quien es el que caracteriza la importancia de los datos históricos. En cuanto al kernel, se presenta un método de tipo heurístico para su selección.

Las nuevas propuestas introducidas en el trabajo son comparadas en una aplicación práctica de predicción del índice de precios al consumidor en Colombia (IPC), donde se presenta evidencia empírica de mejores predicciones respecto a otras técnicas ya introducidas en la literatura que son basadas en redes neuronales. Finalmente una aplicación útil, de las nuevas metodologías introducidas aquí, en epidemiología para la predicción de casos de dengue también es implementada.

Palabras clave: Pronóstico no paramétrico, estimación de kernel, selección ancho de banda, selección coeficientes de Markov

ABSTRACT

The non-parametric time series have advantages over those parametric of usual treatment in practice since the former do not consider assumptions about the historical information of the data. However, in the non-parametric statistical treatment of temporal data there are still some important gaps such as bandwidth selection and the definition of Markov coefficient for the selected kernel. In this work, a robust alternative is proposed to determine the bandwidth in non-parametric time series that consider kernel functions within their prediction processes and in this sense a simple alternative based on the autocorrelation function for the selection of the Markov coefficient is studied, which is what characterizes the importance of historical data. A for the kernel, an heuristic-type method for its selection is also studied.

The new proposals introduced in the work are compared in a practical application of consumer price index prediction in Colombia (IPC), where empirical evidence of better predictions is presented with respect to other techniques already introduced in the literature that are based on neural networks. Finally, a useful application of the new methodologies introduced her, in epidemiology for the prediction of dengue cases is also implemented.

Keywords: Non-parametric forecast, kernel estimation, bandwidth selection, Markov coefficients selection

1. Introducción

El objetivo principal de análisis de las series de tiempo es la predicción de valores futuros. Los métodos estadísticos habituales para modelar las series de tiempo asumen que la historia se repite, no exactamente, pero sí algunos patrones comunes en tendencia y estacionalidad que deben ser capturados en la muestra histórica para estudiarlos, modelarlos y luego utilizar los modelos, para tener alguna sensación precisa del evento en el futuro cercano y así permitan tomar mejores decisiones.

Lograr el objetivo principal del análisis de series de tiempo significa predecir las realizaciones de una variable aleatoria indexada por el tiempo en momentos futuros. Por ejemplo, en ingeniería, predecir la demanda esperada para una línea de producción o servicios con el propósito de asignar los recursos para manufacturar o distribuir. En economía, la predicción de la inflación permitiría hacer ajustes correctos en el IPC y por supuesto en la valoración futura del salario mínimo. En epidemiología, la utilidad de las series de tiempo no se discute, ya que con la estimación futura de nuevos casos, por ejemplo, Dengue, Covid-19 o alguna otra epidemia de contagio masivo se podrían implementar estrategias de anticipación y prevención. En general, son diversos los escenarios de gran utilidad del tratamiento estadístico de datos temporales con fines predictivos.

En la industria, la planificación y la toma de decisiones con base en la experiencia y argumentos empíricos, aunque es habitual y no es del todo incorrecta, toma tiempo para ser implementada. Por lo tanto, las empresas están migrando a una dinámica diferente al entender que la información histórica y/o datos históricos de sus procesos es un activo importante para su análisis y apoyar las decisiones que se toman desde un punto de vista empírico. Ahí es donde

el análisis de series de tiempo toma protagonismo y se convierten en una herramienta esencial para los procesos de planificación tal y como es expuesto en [1].

Hay que tener especial cuidado con la implementación de los métodos estadísticos de predicción, ya que en este aspecto se presentan dos alternativas, aquellas cuyo enfoque es de tipo paramétrico que de acuerdo a [2] consiste básicamente en el logro de dos objetivos, la identificación de un modelo que ajuste adecuadamente los datos con base en la estimación de sus parámetros y luego implementar el modelo identificado para la realización de pronósticos. Siguiendo a [2] se puede afirmar que la metodología Box-Jenkins es una de las más conocidas para este propósito ya que ofrece una alternativa sencilla para la construcción, identificación y predicción, al suponer un proceso autorregresivo de media móvil estacional *SARIMA* pero el supuesto de linealidad considerado en esta metodología es difícil de conseguir en diversas situaciones de tipo práctico [3] [4].

Aun así, la implementación de éstas metodologías paramétricas de pronósticos en aplicaciones de industria y empresa, se ha venido realizando con el fin de confeccionar las cantidades adecuadas de unidades producidas para atender determinada demanda en algún sentido probabilístico. Esto por lo general exige que los analistas en las empresas tengan no solo un amplio dominio de la serie y las características del producto, sino también de los supuestos considerados en las metodologías paramétricas para la definición del modelo predictivo. Sin embargo, es común que los datos históricos cambien de comportamiento en el tiempo y una misma estimación de parámetros estática en un modelo identificado puede no ser la correcta para un horizonte amplio de tiempo, incluso el modelo identificado puede variar. Por esta razón en [5] exponen que muchas empresas han redefinido los procesos de elaboración de pronósticos para llevar a cabo una mejor planeación de sus actividades comerciales.

Por lo tanto, la otra alternativa considerada en la literatura son las metodologías de tipo no paramétrico [6] [7] [8] que presentan notables ventajas sobre aquellas paramétricas de tratamiento habitual en la práctica, ya que éstas no consideran supuestos sobre la información histórica de los datos. El enfoque que se trata en este trabajo para las predicciones es aquel con base en funciones kernel las cuales son funciones no negativas e integrables que caracterizan la importancia de los datos en cada segmento del tiempo.

Sin embargo, en el tratamiento estadístico no paramétrico de datos temporales aún existen algunas brechas importantes para ser investigadas e intentarlas cerrar [9]. Por ejemplo, la selección del ancho de banda para el kernel seleccionado, incluso la elección del kernel aún es un problema que sigue siendo objeto de recientes investigaciones [10] [11] [12]. En este trabajo, se propone una alternativa robusta para determinar el ancho de banda en series de tiempo no paramétricas que consideran funciones kernel dentro de su procesos de predicción. Otra contribución del trabajo es una alternativa sencilla basada en la función de autocorrelación para la selección del coeficiente de Markov quien es el que caracteriza la importancia de los datos históricos. Un método de corte heurístico para la selección del kernel también es considerado y expuesto. Las nuevas propuestas introducidas en el trabajo son comparadas en una aplicación práctica de predicción del índice de precios al consumidor en Colombia, donde se presenta evidencia empírica con la propuesta investigada aquí, de mejores predicciones respecto otras técnicas ya introducidas en la literatura que son basadas en redes neuronales. Además una aplicación útil en epidemiología para predecir los casos de dengue en el municipio de Itagüí también es implementada al final del trabajo.

Este trabajo está organizado de la siguiente forma: en la primera parte se explica el concepto de la regresión no paramétrica destacando su gran utilización en la actualidad junto con las ventajas frente al uso de los modelos clásicos.

Luego se muestran las funciones tipo kernel junto con los dos elementos que definirán el nivel de suavizamiento, el coeficiente de Markov y el ancho de banda, para aplicarlos a diferentes predictores. Finalmente se ilustrará dos aplicaciones de los pronósticos no paramétricos basado en funciones tipo kernel con la nueva propuesta de estimación robusta del ancho de banda, para la predicción de variación porcentual del IPC y la predicción de los casos de dengue en Itagüí.

2. Regresión no paramétrica

En los últimos años el uso de modelos de regresión no paramétricos en el estudio de series de tiempo ha tenido un crecimiento importante debido a la evolución tecnológica que permite cómputos en tiempos razonables y que en el pasado era una de sus principales críticas. Además, este tipo de metodologías permiten ser aplicadas sin imponer supuestos sobre la función de distribución de la serie que se desea estudiar. Entre los métodos estadísticos no paramétricos para predicción están las metodologías basadas en suavizados con funciones de tipo kernel que son funciones no negativas y que permiten caracterizar la importancia de los datos en los tramos del tiempo estudiados, similar a una función de densidad. La regresión no paramétrica, a pesar de sus ventajas, tiene en las redes neuronales artificiales su más directa competidora. Por lo tanto, debido a la amplia literatura reciente y estado modal de las aplicaciones de las redes en diferentes contextos, la regresión y series de tiempo no paramétricas vienen adquiriendo mayor protagonismo en la última década.

El término "no paramétrico" se refiere a la forma funcional de la curva de regresión, ver por ejemplo [13], y es a partir de los datos mismos que se realiza el proceso sin ningún tipo de supuestos sobre ellos. Se deja que los datos hablen por sí solos, y no que algún modelo paramétrico imponga sus leyes de manera forzada. Es así que las nociones de estadística no paramétrica se refieren más que todo al uso de métodos que no dan por conocida la distribución probabilística del objeto de estudio. El enfoque no paramétrico para la estimación de la curva de regresión tiene como propósito, permitir a partir de las observaciones, sin la necesidad del uso de un modelo paramétrico específico, generar predicciones. Es decir, no se considera una forma funcional inicial que describa la estructura de dependencia de la variable respuesta con respecto a los predictores. Aunque la filosofía con que se construyen los estimadores es similar a la del modelo lineal general, y no es más que estimar una media de la variable respuesta condicionada a valores fijados en los predictores. Otra ventaja destacable según [8] es la flexibilidad en la imputación de valores faltantes del conjunto de datos.

En conclusión, la flexibilidad de este tipo de métodos radica en que no se requiere saber de forma preliminar la distribución de los datos.

La regresión no paramétrica considera varias metodologías para su implementación, entre ellas: K-vecinos mas cercanos (knn) [14], suavizamiento de kernel (kde) [15], regresión usando polinomios locales, estimaciones con series ortogonales y suavizamiento de Splines. El presente trabajo está orientado a la estimación de la curva de regresión utilizando suavizamiento con base en funciones kernel, haciendo contribuciones relevantes en tres aspectos fundamentales en este tipo de enfoque. i) Una estimación robusta del ancho de banda al modificar la expresión conocida para tal efecto, la cual se basa en una medida de dispersión para su cálculo. En este trabajo se propone dejar la misma estructura pero cambiando la medida de dispersión habitual por una versión robusta de la misma. ii) El coeficiente de Markov que caracteriza la importancia de los datos históricos, no es analizado con detalle en la literatura, su definición en la mayoría de los casos no tiene una justificación analítica. Sin embargo, en este trabajo se propone estimarlo a través de la

función de autocorrelación. iii) Aunque la elección del kernel es un tema que aún sigue siendo objeto de investigación se ha hecho un exhaustivo estudio de simulación, considerando un conjunto de funciones kernels y sus respectivos parámetros para identificar qué función kernel y qué parámetros son los más adecuados según el conjunto de datos. Las contribuciones de este trabajo de mayor relevancia son las dos primeras, sin embargo la tercera es una idea sencilla y de fácil implementación que resuelve parcialmente la pregunta sobre qué tipo de kernel se debe usar. Las tres contribuciones se implementan y se comparan en datos reales con algunas estrategias implementadas en la literatura específicamente para los datos considerados. Se presenta evidencia empírica de un mejor desempeño con las estrategias propuestas.

Considere una serie observada Z_1, Z_2, \dots, Z_n , el propósito de estudiar estadísticamente estas observaciones será predecir el valor no observado Z_{n+l} para algún entero $l \geq 1$. Una manera sencilla de considerar este problema es suponer que la serie es un proceso autorregresivo de orden q . Es decir,

$$Z_t = m(Z_{t-1}, Z_{t-2}, \dots, Z_{t-q}) + \epsilon_t, \quad (1)$$

donde ϵ_t es el error del proceso, que debe ser independiente del pasado $Z_{t-1}, Z_{t-2}, \dots, Z_{t-q}$. El enfoque supone que para obtener una realización de Z_t se hace necesario identificar la función $m(\cdot)$.

Si se supone que $Z_t : t \in \mathbb{Z}$ es una serie estacionaria estricta univariada, y que sigue un modelo d-Markoviano con valor q , entonces un acercamiento clásico constará de asumir una forma funcional para $m(\cdot)$ y realizar estimaciones para sus parámetros considerados internamente. Mientras que en un acercamiento de tipo no paramétrico no se tendrían que realizar supuestos sobre la forma funcional de $m(\cdot)$, y ésta vendría heredada de las observaciones y desde la estimación no paramétrica del siguiente funcional:

$$m(\mathbf{u}) = E(Z_t | (Z_{t-i_1}, Z_{t-i_2}, \dots, Z_{t-i_q}) = \mathbf{u}), \quad (2)$$

donde $\mathbf{u} \in \mathbb{R}^q$, $0 < i_1 < i_2 < \dots < i_q$. Con el fin de estimar Z_{t+1} , se debe considerar la tupla de vectores $\{(X_j, Y_j) : j \in \mathbb{Z}\}$, con valores en $\mathbb{R}^q \times \mathbb{R} = \mathbb{R}^{q+1}$, donde $X_j = (Z_{j+i_q-i_1}, Z_{j+i_q-i_2}, \dots, Z_j)$ y $Y_j = Z_{(j+i_p)}$, lo cual significa que los datos de la serie de tiempo observada se pasan a una adaptación de la forma:

$$\{(x_j, y_j) : 1 \leq j \leq n - i_q\}, \quad (3)$$

para ser luego implementado un modelo de regresión. Observe que la ecuación (2), es una esperanza condicional cuya obtención teórica, supone el conocimiento de la densidad condicional, y la densidad asociada a las variables aleatorias que caracterizan el evento condición. Estas densidades, en la práctica son desconocidas, por lo tanto para estimar el funcional definido en (2), se necesita estimaciones de aquellas densidades, y es ahí donde las funciones de tipo kernel son consideradas para ese propósito.

3. Suavizamiento de kernel

La estimación de densidades sigue siendo un tema de recientes investigaciones en el área de estadística no paramétrica. Una de las alternativas para ello, es realizar las estimaciones a través de funciones tipo kernel, que permiten caracterizar la importancia de los datos en cada tramo del rango al imponer pesos. Es conocido que un primer acercamiento simple a una densidad es el histograma de frecuencias relativas, que no es más que una estimación de la densidad a través de un kernel uniforme (constante). Algunas gráficas de kernel de usos común son expuestas a continuación:

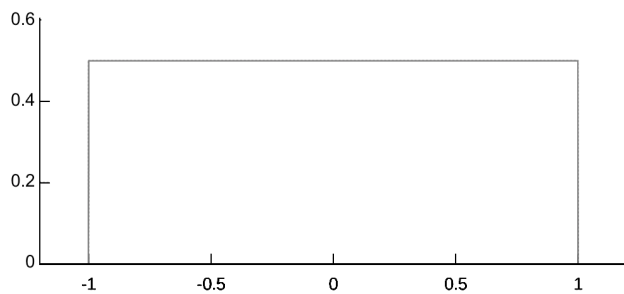


Figura 1: Kernel Uniforme $K(u) = \frac{1}{2}$, soporte: $|u| \leq 1$

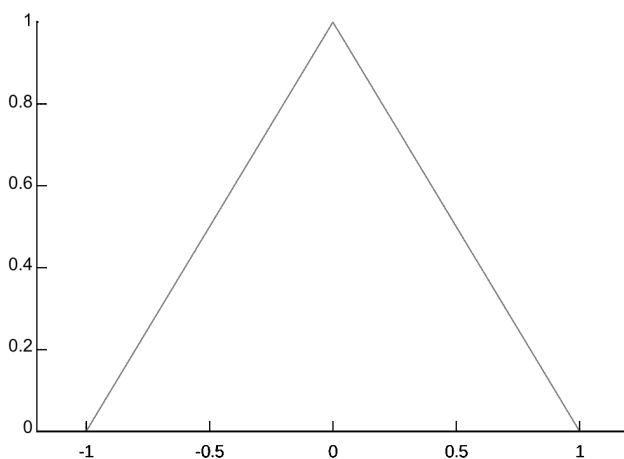


Figura 2: Kernel Triangular $K(u) = (1 - |u|)$, soporte: $|u| \leq 1$

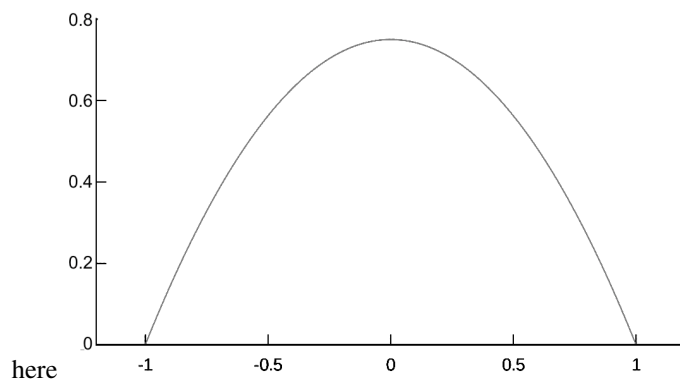


Figura 3: Kernel Epanechnikov $K(u) = \frac{3}{4}(1 - u^2)$, soporte: $|u| \leq 1$

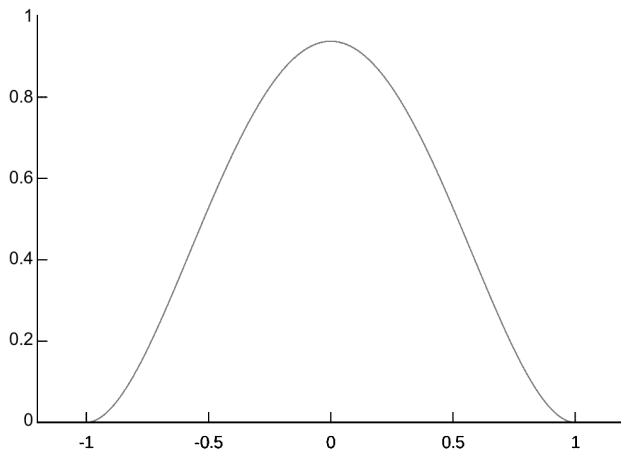


Figura 4: Kernel Quartic $K(u) = \frac{15}{16}(1 - u^2)^2$, soporte: $|u| \leq 1$

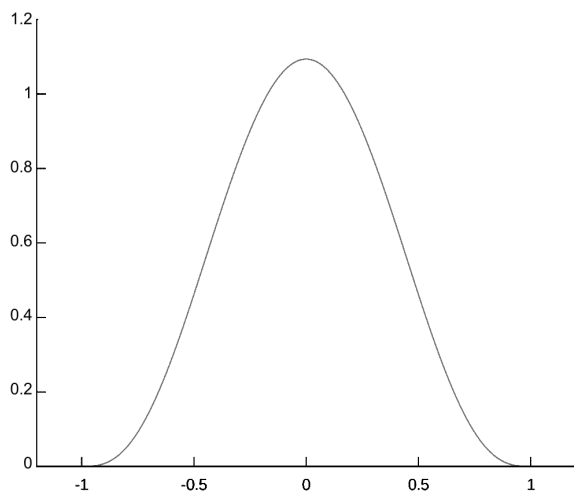


Figura 5: Kernel Triweight $K(u) = \frac{35}{32}(1 - u^2)^3$, soporte: $|u| \leq 1$

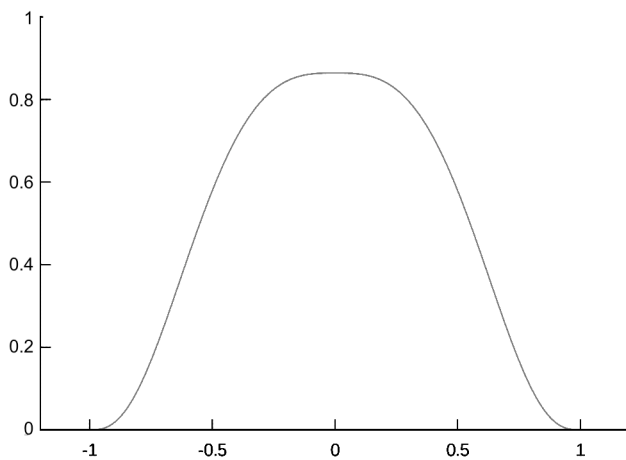


Figura 6: Kernel Tricube $K(u) = \frac{70}{81}(1 - |u|^3)^3$, soporte: $|u| \leq 1$

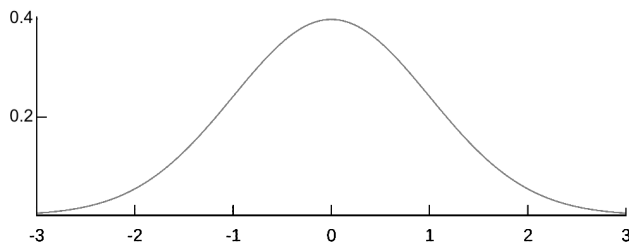


Figura 7: Kernel Gaussiano $K(u) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}$

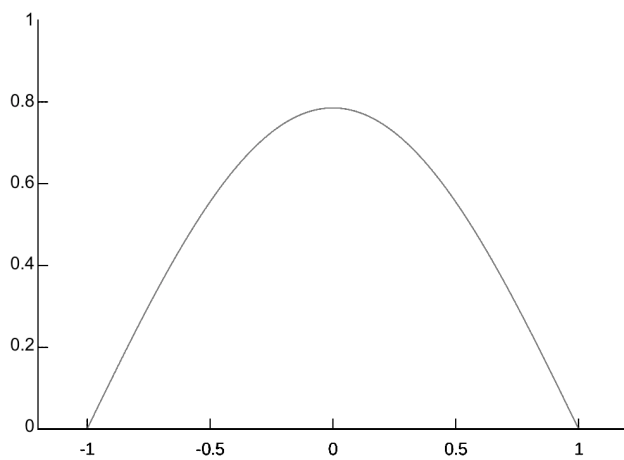


Figura 8: Kernel Coseno $K(u) = \frac{\pi}{4}\cos(\frac{\pi}{2}u)$, soporte: $|u| \leq 1$

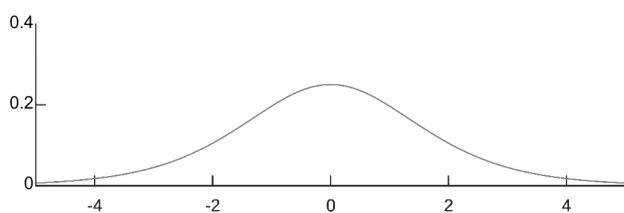


Figura 9: Kernel Logístico $K(u) = \frac{1}{e^u + 2 + e^{-u}}$

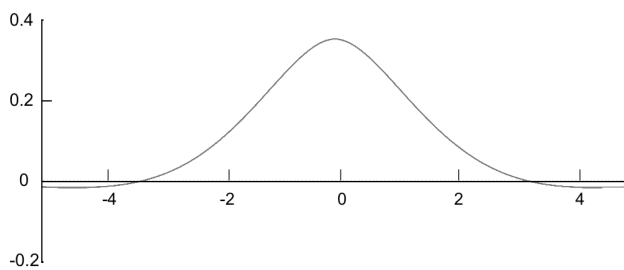


Figura 10: Kernel Silverman $K(u) = \frac{1}{2}e^{-\frac{|u|}{\sqrt{2}}} * \text{sen}(\frac{|u|}{\sqrt{2}} + \frac{\pi}{4})$

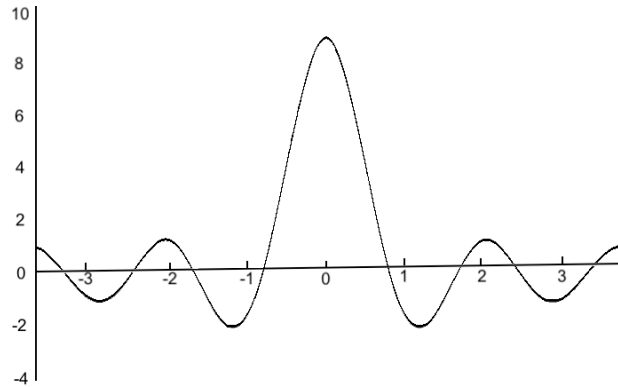


Figura 11: Kernel Dirichlet[2] $K(u) = \frac{\text{sen}((M+0.5)\mu)}{2\pi \text{sen}(0.5\mu)}$

Una función tipo kernel debe satisfacer algunas propiedades de regularidad deseables, por ejemplo, debe ser simétrica respecto al origen y unimodal para garantizar que los pesos que ella asigna se repartan de manera simétrica y con mayor concentración respecto a la observación que hace las veces de centro. No negativa e integrable para que permita estimar una densidad con las propiedades que la caracterizan. Debe tener segundo momento finito que garantiza que el estimador construido tenga un sesgo asintótico acotado. Tener j derivadas continuas para la obtención de una estimación suave, y soporte compacto para asuntos de costo computacional. Las anteriores son propiedades deseables para la función kernel, pero son prescindibles.

Con lo anterior, dada una muestra, X_1, \dots, X_n , un estimador para una función de densidad, a través de un kernel $K(\cdot)$ podría ser definido como

$$\hat{f}_K(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right), \quad (4)$$

donde h denota el ancho de banda o *bandwidth* en inglés. Este parámetro h es el radio del intervalo donde la función kernel está realizando su función. Está asociado con el radio de cada clase en una tabla tradicional descriptiva de frecuencias. Este ancho de banda es considerado también como un parámetro de suavizado que controla la concentración del peso alrededor de cada observación X_i . Si h es pequeño las observaciones más cercanas a x serán las más relevantes en una estimación de la densidad que tendría mucha varianza pero poco sesgo. Mientras que si h es grande se permite que en la estimación, las observaciones más lejanas de x tengan contribución, generando estimaciones con poca varianza y mucho sesgo. Por lo tanto, la estimación de h es de notable relevancia por generar amplia sensibilidad en la estimación final de una función de densidad. De acuerdo a [16] en la página 132, una buena estimación del ancho de banda es más relevante que la propia selección del kernel. En esta dirección está una de las principales contribuciones de este trabajo al proponer una versión robusta para la estimación del ancho de banda.

Las estimaciones de densidades son útiles en el análisis de regresión. Por ejemplo, suponga una muestra aleatoria biva-riada $(X_1, Y_1), \dots, (X_n, Y_n)$ y se está interesado en encontrar una función r tal que $r(X)$ sea una buena aproximación de Y , es decir, $Y_i = r(X_i) + \epsilon_i$ tal que $E(\epsilon_i|X_i) = 0$, ver por ejemplo [17]. Se puede mostrar que la aproximación óptima Y_i , en el sentido del menor error cuadrático medio, esta dada por la siguiente esperanza condicional [2]:

$$r(x) = E(Y|X = x), \quad x \in \mathbb{R}. \quad (5)$$

Si se impone un supuesto de linealidad sobre r , entonces el problema se resuelve a través del método de mínimos cuadrados.

Dado a que no se hace ningún supuesto sobre la linealidad de dicha relación, entonces la obtención de la función r se debe hacer a través de estimaciones no paramétricas con funciones kernel.

Observe que para obtener el valor de r en (5), se debe conocer la densidad conjunta $f(x, y)$, para luego calcular la densidad condicional y sobre ésta última calcular el valor esperado. Por lo tanto,

$$r(x) = E(Y|X = x) = \int y \frac{f(x, y)}{f(x)} dy, \quad (6)$$

donde $f(x)$ es la densidad marginal de X y dada por; $f(x) = \int f(x, y) dy$. Denote como,

$$\gamma(x) = \int y f(x, y) dy, \quad (7)$$

entonces la esperanza condicional definida en (6) puede ser escrita como,

$$r(x) = \frac{\gamma(x)}{f(x)}. \quad (8)$$

Un estimador para $\gamma(x)$ según [13] puede ser expresado de la forma;

$$\hat{\gamma}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) Y_i. \quad (9)$$

Por lo tanto, combinando (4) y (9), se puede definir un estimador para la expresión (5) con base en funciones kernel dado por;

$$\hat{r}(x) = \frac{\hat{\gamma}(x)}{\hat{f}(x)} = \frac{\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) Y_i}{\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)}. \quad (10)$$

En general el procedimiento asigna mayor peso a los puntos cercanos a x , y menor o ningún peso a puntos que estén más lejanos a x , es decir, es como un promedio ponderado de los Y , donde el peso depende de la distancia entre X_i y x , la cual es cuantificada por la función $K(\cdot)$. La expresión para $\hat{r}(x)$ dada en (10), se conoce como el estimador de Nadaraya-Watson donde los pesos terminan definiendo una vecindad de puntos alrededor de x . Para mayor detalle sobre este estimador y sus propiedades veáse [18] y referencias internas.

Observe que la forma funcional r para la esperanza condicional puede ser calculada a través de estimaciones de funciones de densidad que consideran algún kernel en su proceso. Dentro de este proceso el ancho de banda h debe ser estimado así como una forma de seleccionar el tipo de kernel. En este trabajo se propone alternativas simples pero novedosas para tratar ambos aspectos y luego utilizar la metodología investigada para aplicarla en datos de tipo temporal haciendo estimaciones adecuadas de la expresión (2).

Cuando un kernel es aplicado a datos dependientes, como sería para el caso de las series de tiempo, el efecto de dependencia que se tiene sobre las observaciones, al ser aplicado en pequeños bloques y no sobre todos los datos, se reduce entre las estimaciones; y es por esta razón que muchas de las técnicas desarrolladas para datos independientes pueden ser aplicados a las series de tiempo, este principio se llama "the whitening by windowing principle", como es expuesto en [19].

La estimación de densidades usando kernel también es conocido como el método de ventanas de Parzen-Rosenblatt, estudiado en [20, 21]. Las funciones kernel terminan definiendo el grado de importancia, según su forma, a los pesos asociados con cada observación para la estimación.

4. Estimación de parámetros

Una vez se ha expuesto una metodología no paramétrica para estimar densidades y su aplicación en regresión, en esta sección se expone las principales contribuciones de la investigación al proponer alternativas sencillas, pero con mejor desempeño, en cuanto a la precisión del pronóstico para la selección del ancho de banda, el coeficiente de Markov y la selección del kernel.

4.1. Ancho de banda

El ancho de banda es el parámetro que determina el nivel de suavizamiento del predictor, ver por ejemplo [22]. Un ancho de banda muy pequeño terminará replicando la serie original, ya que este parámetro en la función kernel hace las veces de filtro de selección de los candidatos para asignarles un peso determinado, como es explicado en [23] y en la Figura 12. Si el ancho de banda se incrementa, la varianza del predictor decrece mientras que el sesgo aumenta, situación que ya se ha expuesto anteriormente pero que está explicado con detalle en [24]. Un conjunto de alternativas estadísticas para la selección del ancho de banda es discutida en [2], entre ellas la validación cruzada que ha sido de la más comunes, pero supone un esfuerzo y costo computacional importante. Una alternativa para la selección del parámetro, con menos esfuerzo de cómputo, es estudiado en [25] donde se propone la estimación del ancho de banda utilizando

$$h = \hat{\sigma} * (n^{\frac{-1}{d+4}}), \quad (11)$$

donde $\hat{\sigma}$ es la estimación natural de la desviación estándar de la serie de tiempo y d es el coeficiente de Markov, definido en 4.2. Este estimador para h es de implementación frecuente, sobre todo cuando el kernel es gaussiano y es considerado en [2] para la predicción no paramétrica de series de tiempo pero en el mismo trabajo no se logra mejorar el comportamiento de la estimación no paramétrica respecto al modelo tradicional paramétrico ARIMA. Una de las posibles causas es que el estimador natural para la desviación estándar, aunque es insesgado para σ , su forma, la cual depende de una sumatoria, es bastante sensible a la presencia de datos atípicos y cuando esto ocurre, las estimaciones puntuales no son correctas afectando notablemente la estimación del parámetro de suavizado h .

Dada la relevancia del parámetro h en la estimación de densidades, en este trabajo se propone una nueva versión para su estimación, considerando las ideas propuestas en [25], pero confeccionando una versión robusta. La idea principal es modificar el estimador de la variabilidad $\hat{\sigma}$ por una alternativa más robusta como la desviación mediana absoluta MAD la cual es introducida en [26].

La MAD , es considerada en [26], como una versión robusta para la desviación estándar σ . Para una variable aleatoria X , MAD es dada por:

$$MAD = med(|X - med(X)|), \quad (12)$$

donde $med(\cdot)$ significa la mediana poblacional. Dada una muestra aleatoria X_1, \dots, X_n y denote por $\hat{med}(X)$ a la mediana empírica de X_1, \dots, X_n , entonces una estimación natural para la MAD , es

$$\hat{MAD}(X) = med(|X_i - \hat{med}(X)|). \quad (13)$$

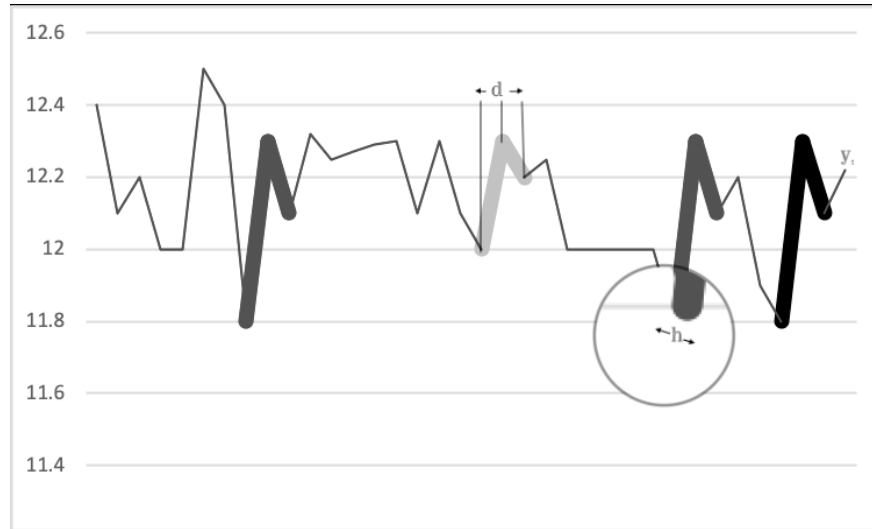


Figura 12: Tomando como referente [2] A) En la lupa h representa el ancho de banda en el kernel. B) El tamaño de la ventana o el coeficiente de Markov está representado por la cantidad de puntos que serán considerados para construir las trazas; en este ejemplo $d = 3$. C) La traza de color más oscuro representa el referente de comparación cuando $m = 1$. D) La influencia para determinar Y_{t+1} que tiene la subtraza de gris más claro es menor que las subtrazas de color más oscuro.

Existe una relación directa bajo normalidad entre la desviación estándar y la MAD . Considere una variable aleatoria X , normal de parámetros μ y σ , siguiendo a [26], esta relación es

$$MAD(X) = \sigma \text{med} \left(\left| \frac{X - \mu}{\sigma} \right| \right) = \sigma \Phi^{-1}(0.75),$$

siendo Φ^{-1} la función cuantílica de la distribución normal estándar. Esta misma relación pero sin asumir el supuesto de normalidad ha sido investigada recientemente en [27] donde se estudia la equivalencia para otros tipos de distribuciones y se calcula el factor de corrección, en algunas de una forma exacta y en otras se introduce una expresión de tipo numérico para su cálculo.

La consistencia fuerte del estimador definido en (13) está garantizada por el Corolario 3.1 en [26], donde se afirma que si F y H son continuas y estrictamente crecientes en $F^{-1}(-1/2)$ y $H^{-1}(1/2)$, respectivamente, entonces

$$M\hat{A}D(X) \longrightarrow MAD(X) \quad c.s.^1$$

Donde F y H se refiere a las funciones de distribución de X , y $(X - \text{med}(X))^2$, respectivamente. La normalidad asintótica del estimador en cuestión está establecida en el Teorema 4.1 en [26]. Por lo tanto, la versión muestral de MAD tiene propiedades asintóticas deseables y en este trabajo es considerado para la determinación de una forma robusta de obtener el ancho de banda h a través de la expresión

$$h = M\hat{A}D(X)[n]^{\frac{-1}{d+4}}. \quad (14)$$

¹convergencia casi segura

4.2. Coeficiente de Markov

Para el uso de métodos no paramétricos sobre series de tiempo, se debe definir la serie $\{Z_t\}$ como un proceso d -Markoviano, donde d es un coeficiente de Markov, quien es el que relaja la memoria en los procesos Markovianos naturales. Es decir, la distribución condicional de Z_t dado cualquier evento pasado Z_{t-s} , es igual a la distribución condicional de Z_t dado los d primeros eventos pasados. Por lo tanto, el coeficiente de Markov es el atributo dentro de la estimación no paramétrica, que define la profundidad que se desea tenga influencia el pasado en un horizonte específico de la serie. Dada una realización z_1, z_2, \dots, z_n se debe intentar predecir una variable aleatoria no observada Z_{n+m} con $m \geq 1$ en donde m es el horizonte de predicción. Se crea un proceso asociado $\{X_t, Y_t\}$, donde

$$X_t = (Z_t, \dots, Z_{t-d+1}) \quad y \quad Y_t = Z_{t+m}, \quad t \in d, \dots, n. \quad (15)$$

Observe que para hacer la predicción se debe calcular $E(Y_t|X_t)$. Si se conociera la distribución condicional sería abordable el problema, sin embargo en la práctica esta distribución es desconocida. Por lo tanto, se debe implementar un estimador como aquel definido en la expresión (10)

$$\hat{r}(x) = \hat{E}(Y_t|X_t) = \sum_{t=d}^{n-m} W_t^m(x) Y_t, \quad (16)$$

donde

$$W_t^m(x) = \frac{K_d\left(\frac{x-X_t}{h_n}\right)}{\sum_{t=d}^{n-m} K_d\left(\frac{x-X_t}{h_n}\right)}, \quad x \in \mathbb{R}^d. \quad (17)$$

La función K_d es una función kernel pero multivariada de dimensión d , quien controla la forma de los pesos y su importancia, en donde x es un trozo de la serie original en una ventana de tamaño d que sirve de referencia para comparar con diferentes trozos de la serie de un mismo tamaño d , luego una secuencia de observaciones tendrá mayor peso en el pronóstico que se desee realizar si es más parecido en términos de distancia al bloque de referencia x .

Escoger el coeficiente de Markov d lo más grande posible sería lo adecuado para capturar la mayor información del pasado para la comparación entre los bloques. Sin embargo, la elección del coeficiente d está limitada por la cantidad de datos disponible, toda vez que al crecer el coeficiente de Markov la cantidad disponible de datos para la predicción disminuye.

Dentro del proceso de implementación para encontrar un d óptimo autores como [2] y [28] proponen el uso de un método empírico orientado a encontrar el mejor de los tres tipos de predictores condicionales: media, mediana y moda. La idea discutida en [2] y [28], para la selección d es que al considerar c como la predicción en un tiempo t , dados el coeficiente de Markov $d \in \{1, \dots, d_{max}\}$ y el ancho de banda h para cada uno de los tres predictores, se deben calcular las siguientes tres funciones:

$$f_1(d) = p^{-1} \sum_t |Z_t - \hat{Z}_t(d, h_n)|, \quad (18)$$

$$f_2(d) = p^{-1} \sum_t \{Z_t - \hat{Z}_t(d, h_n)\}^2, \quad (19)$$

$$f_3(d) = \sup |Z_t - \hat{Z}_t(d, h_n)|. \quad (20)$$

Si las funciones $f_j(d)$, $j \in \{1, 2, 3\}$ no se incrementan después de cierto valor, se selecciona el valor $f_j(d)$ con el cual ocurre la estabilización. Este método puede llegar a ser computacionalmente intensivo y adicional, cuenta que d_{max} deberá ser lo suficientemente grande para capturar cualquier efecto estacional. Aquí es donde se introduce otra de las contribuciones de este trabajo y es la selección del coeficiente de Markov a través de la función de autocorrelación parcial (PACF) de la siguiente manera:

- Se define el valor del intervalo de confianza para la función de autocorrelación parcial.
- Los palotes que hayan superado el intervalo de confianza serán seleccionados para ser candidatos como coeficientes de Markov.

Este proceso es una selección natural para d dada la definición de la PACF la cual es una medida de asociación en la serie entre valores en un instante y aquellos pasados. La PACF de una serie de tiempo define los momentos separados por k unidades de tiempo (x_t y x_{t-k}). Observe que es un procedimiento simple, intuitivo. Además el menor esfuerzo computacional respecto a la estrategia considerada en [2] y [28] es de notable ventaja. También con este proceso se evidencia mejoría en las predicciones respecto a las medidas de desempeño que serán expuestas más adelante.

4.3. Selección de función de kernel

Una contribución sencilla, pero útil, en este trabajo es darle importancia a la selección de todos los parámetros que intervienen en la aplicación, entre ellos el kernel, la selección del kernel en la propuesta implementada en [2] es hecha sin mayores explicaciones y de una manera ad-hoc. En este artículo la selección se hace a través de una búsqueda exhaustiva considerando todas las funciones tipo kernel, expuestas anteriormente con un rango de valores definidos por el coeficiente de Markov d y el ancho de banda h , para luego seleccionar el kernel con el menor error según algunas de las métricas que serán expuestas más adelante. La comparación de las metodologías introducidas aquí, serán comparadas con otros trabajos donde se hacen predicciones del IPC. En general, el procedimiento implementado puede ser resumido en los siguientes pasos:

1. Seleccionar una quinta parte de los datos para realizar el entrenamiento de las diferentes combinaciones de parámetros, kernel, coeficiente de Markov y ancho de banda.
2. Calcular la función de autocorrelación parcial (PACF) para seleccionar los palotes que salgan del intervalo de confianza definido que serán los coeficientes de Markov o tamaños de las ventanas.
3. Preparar la serie según los diferentes coeficientes de Markov.
4. Calcular el ancho de banda de acuerdo a la expresión (14).
5. Iterar sobre la lista de los diferentes kernels, para el caso en particular serían sobre kernels Gaussiano, logístico, sigmoide, Silverman, triangular, coseno, Epanechnikov, triweight, quartic, tricube, Dirichlet en sus diferentes variaciones de potencias desde la 2 hasta la 13. Definiendo en cada proceso el coeficiente de Markov.
6. Seleccionar el kernel que tenga mejor comportamiento sobre los datos de entrenamiento, ajustando los mejores parámetros de coeficientes de Markov y anchos de banda. Para identificar la combinación de mejor ajuste se utilizan algunas de las medidas de error que serán introducidas en la Sección 6.

7. Se repite el numeral anterior ya únicamente con el kernel ganador y luego se correrán las combinaciones (d y h) con los datos de prueba.
8. Generar el pronóstico.

Hay que tener en cuenta que las funciones kernel que se consideran en este trabajo son de dominio acotado salvo el kernel Gaussiano, por lo tanto los datos se deben escalar al dominio de la respectiva función kernel utilizando una normalización adecuada, cuando la serie de estudio así lo exija.

5. Predictores

Para la estimación del bloque de predicción según el horizonte m dado por la variable Z_{t+m} o Y_t , según la expresión (15) con $m \geq 1$, se debe encontrar la variable más cercana, en términos de alguna función de pérdida, dado todo el pasado de la serie. Si aquella función de pérdida es convexa $l(\cdot)$ con mínimo único en cero, entonces se define $r(x)$ como la función que minimiza $E(l(Y - a)|X = x)$:

$$r(x) = \arg \min_{a \in \mathbb{R}} \{E(l(Y - a)|X = x)\}. \quad (21)$$

Dependiendo del tipo de función convexa l existen varios candidatos para asumir el rol de predictor. Entre los más comunes son las tres medidas de tendencia central, donde a continuación se explica cómo se selecciona cada una de ellas.

5.1. Media condicional

La media condicional es el predictor de uso más frecuente, ver por ejemplo [29]. Viene cómo resultado de considerar en (21) una función de la forma $l(u) = u^2$ y cuyo resultado es la media condicional $r(x) = E(Y|X = x)$, cuya estimación dado un proceso d -Markoviano, viene dada por la expresión ya discutida y definida en la ecuación (16)

5.2. Mediana condicional

El regresor de la mediana condicional sirve como una alternativa al autorregresor de la media condicional, es usual encontrar en la literatura su uso en áreas financieras, véase [30]. Para este caso, el uso de la mediana se comportará mejor cuando la distribución sea más pesada en un extremo, asimétrica o haya sospechas de datos atípicos. Este predictor es obtenido cuando la función de pérdida es la forma $l(u) = |u|$. También puede ser definido como $r(x) = \inf\{y : F(y|X = x) \geq \frac{1}{2}\}$. Ahora, dado un proceso d -Markoviano una estimación no paramétrica de r puede ser obtenida como:

$$\hat{r}(x) = \inf \left\{ y : \sum_{t=d}^{n-m} W_t^m(x) I_{\{Y_t \leq y\}} \geq \frac{1}{2} \right\}, \quad (22)$$

y el estimador no paramétrico para un horizonte m sería:

$$Z_{n+m} = \inf \left\{ y : \sum_{t=d}^{n-m} W_t^m(X_n) I_{\{Y_t \leq y\}} \geq \frac{1}{2} \right\}, \quad (23)$$

inspirados en la mediana condicional, en este trabajo también se propone usar como predictor otro percentil que será discutido en la sección 5.4

5.3. Moda condicional

En la literatura también se encuentran estimadores no paramétricos basados en la función moda, véase [31] para mayor detalle. La moda como predictor viene definida como $r(x) = \operatorname{argmax}_{y \in \mathbb{R}} f(Y|X = x)$. Dada la función de pérdida con mínimo único $l(u) = 0$ cuando $u = 0$ y $l(u) = 1$ en otro caso y un proceso d -Markoviano, entonces una estimación no paramétrica para r sería:

$$\hat{r}(x) = \operatorname{arg max}_{y \in \mathbb{R}} h^{-1} \sum_{t=d}^{n-m} W_t^m(x) K\left(\frac{y - Y_t}{h_n}\right), \quad (24)$$

luego el estimador no paramétrico para un horizonte m sería:

$$Z_{n+m} = \operatorname{arg max}_{y \in \mathbb{R}} h^{-1} \sum_{t=d}^{n-m} W_t^m(X_n) K\left(\frac{y - Y_t}{h_n}\right). \quad (25)$$

5.4. Percentil condicional

El predictor basado en el percentil condicional sirve como una alternativa al predictor de la mediana condicional, este es uno de los aportes propuestos por nuestra metodología. El regresor del percentil condicional puede verse como una generalización de aquel asociado con la mediana y puede ser definido como $r(x) = \operatorname{inf}\{y : F(y|X = x) \geq P_i\}$ para el i -ésimo percentil. Por lo tanto, dado un proceso d -Markoviano se obtiene:

$$\hat{r}(x) = \operatorname{inf} \left\{ y : \sum_{t=d}^{n-m} W_t^m(x) I_{\{Y_t \leq y\}} \geq P_i \right\}, \quad (26)$$

luego el estimador no paramétrico para un horizonte m sería:

$$Z_{n+m} = \operatorname{inf} \left\{ y : \sum_{t=d}^{n-m} W_t^m(X_n) I_{\{Y_t \leq y\}} \geq P_i \right\}. \quad (27)$$

Observe que estos cuatro estimadores consideran en sus ecuaciones de predicción los pesos $W_t^m(X_n)$, que garantizan que las observaciones más importantes en los estimadores serán aquellas que tengan pesos mayores y los bloques que estén afectados por pesos cercanos a cero tendrán poca influencia en el estimador, lo que hace a la metodología de estimación de kernel poco sensible a valores atípicos.

6. Medidas de error

Evaluar un modelo hace parte esencial del proceso de generación de pronósticos y para este trabajo se han escogido tres métricas para la comparación de pronósticos: La raíz del error cuadrático medio (RMSE), el error cuadrático medio (MSE) y el error absoluto medio (MAE) y definidas como:

$$RMSE = \sqrt{\frac{\sum_{t=1}^m (Z_t - \hat{Z}_t)^2}{m}}, \quad (28)$$

$$MSE = \frac{1}{m} \sum_{t=1}^m (Z_t - \hat{Z}_t)^2, \quad (29)$$

$$MAE = \frac{1}{m} \sum_{t=1}^m |Z_t - \hat{Z}_t|. \quad (30)$$

7. Aplicación

El objetivo de esta sección es determinar si la estimación no paramétrica propuesta aquí, con una nueva metodología de selección de kernel, coeficiente de Markov y ancho de banda, captura la dinámica de la serie en el grupo de entrenamiento y genera mejores predicciones que anteriores aplicaciones de metodologías no paramétricas y de modelos de redes neuronales artificiales expuestas en [32, 33]. Adicionalmente, se aplicó la técnica introducida en este trabajo para pronosticar los casos de dengue para el municipio de Itagüí en Antioquia mostrando resultados prometedores.

Los resultados aquí presentados fueron obtenidos en una maquina Intel Core i7-4850HQ @ 8x 3.5GHz, utilizando python 3.7.6.

7.1. Predicción del IPC en Colombia

Esta propuesta se comparará la metodología, seleccionando el kernel de Dirichlet en potencia 4 con un predictor de media, con la implementación y los resultados presentados en [32, 33], donde se toma la serie del IPC colombiano con respecto a la variación mensual según datos del DANE, desde enero 1998 hasta junio de 2005 y pronosticando 6 meses adelante. El comparativo se ilustra en la Figura 13, donde se observa desde un punto de vista visual mejores predicciones que aquellas realizadas en [32, 33]. Además, en la Tabla 2 se observan las comparaciones respecto a las tres medidas de error nombradas anteriormente, donde se nota de manera clara un mejor desempeño de la propuesta introducida en este artículo.

Tabla 1: Resultados pronósticos [33] vs DirichletM4-Suavizamiento de kernel

	Real	Santana2006	DirichletM4
1	0.0487	0.008	-0.0142
2	0.0015	-0.116	0.0387
3	0.4278	0.379	0.3330
4	0.2301	0.43	0.2292
5	0.1143	0.107	-0.0178
6	0.0682	0.236	0.01088

Modelo	MSE	MAE
Santana[33]	0.014	0.097
Velásquez-Franco[32]	0.007	0.081
DirichletM4	0.003	0.0642

El caso de la comparación de este trabajo con los pronósticos punto a punto hecho por [32] no se realiza debido a que allí no fueron reportados en el artículo. Sin embargo, en el trabajo se realizó una descripción de los resultados de las diferentes topologías de la red neuronal artificial que en todo caso fueron mejorados por esta propuesta.

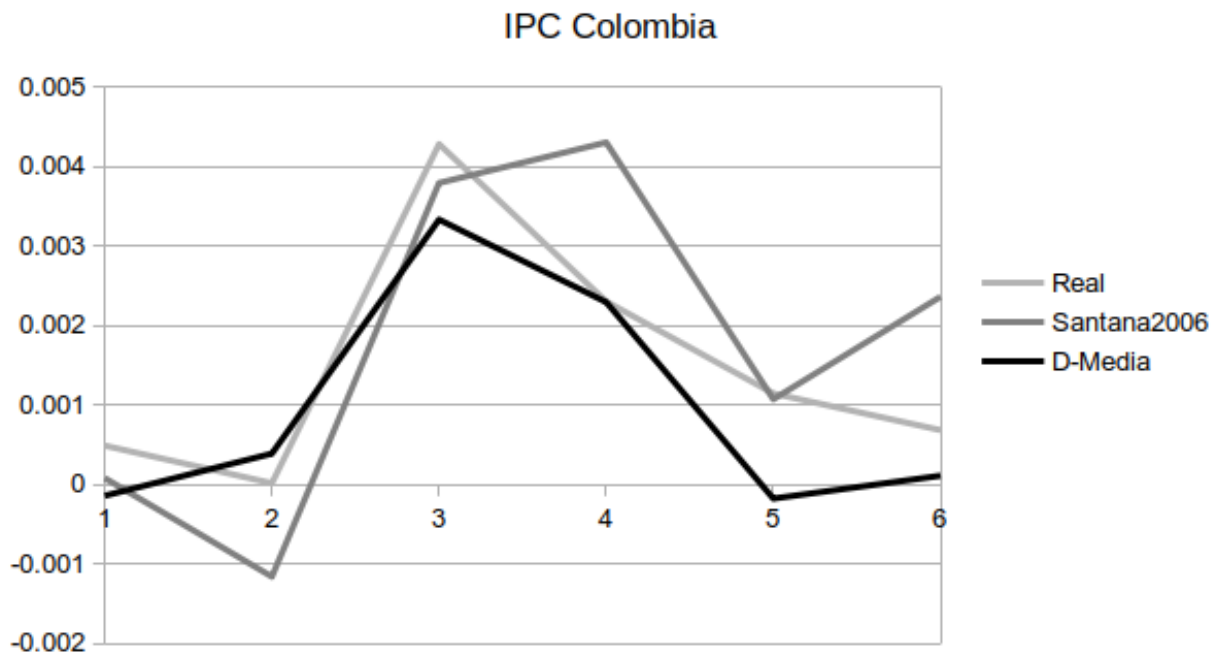


Figura 13: Comparación de pronósticos no paramétricos versus IPC real

7.2. Aplicación casos de dengue Itagüí-Antioquia

Con base en la metodología introducida y seleccionado el kernel sigmoide con el predictor de media, se hace otra aplicación útil para las unidades de salud en el tema del control epidemiológico. Es sabido que el dengue es una enfermedad que está teniendo un impacto negativo importante en diferentes centros urbanos de países tropicales. Colombia no es la excepción y la realización de predicciones a mediano plazo, permitirá a las autoridades de salud competentes realizar programas de prevención y anticipación. Por lo tanto, la metodología investigada aquí, es una alternativa novedosa y útil para realizar esas predicciones de una manera precisa. En la aplicación a casos reportados de dengue en el municipio de Itagüí, en el Departamento de Antioquia, se tomaron los datos de contagios en la zona y para la elección de los coeficientes se tomó a partir de la primera semana de 2013 hasta la décima semana del año 2016 que contiene 164 datos. Para encontrar los coeficientes se tomaron datos desde la primera semana de 2013 hasta la cuarta semana de 2015, es decir, iniciando con 132 observaciones; luego se aumenta el tamaño de muestra hasta la décima semana de 2016. Los resultados se comparan con un modelo $ARIMA(3,1,0)(2,1,0)$ y pueden verse en la Figura 14.

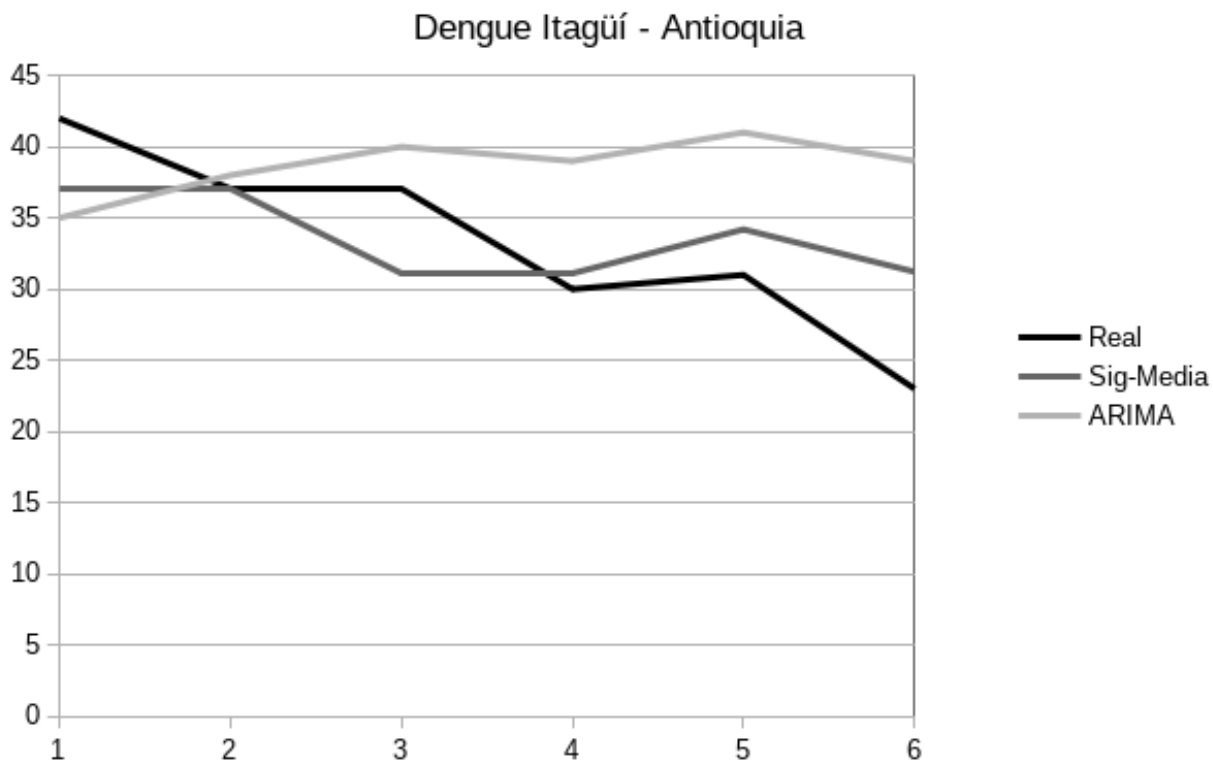


Figura 14: Estimación casos de dengue en el municipio de Itagüí - Antioquia

Tabla 2: Resultados pronósticos ARIMA vs Sigmoide-Suavizamiento de kernel

	Real	ARIMA	Sigmoide
1	42	35	37
2	37	38	37
3	37	40	31
4	30	39	31
5	31	41	34
6	23	39	31

Modelo	MSE	MAE
ARIMA	82.667	7.667
Sigmoide	22.5	3.833

8. Conclusiones

El resultado de este trabajo es la creación de una metodología para seleccionar los parámetros en la ejecución de pronósticos usando regresión no paramétrica con suavizamiento de kernel. Este modelo permite obtener:

- Pronósticos más acertados, sin requerimientos de recursos altamente especializados (humanos, tecnológicos, infraestructura, entre otros).
- Permite el acercamiento de las pymes a otro tipo de metodología para apoyar el proceso de planificación.
- Definición de los coeficientes de Markov de forma automática explotando una de las consideraciones básicas para su uso, como es el concepto autoregresivo.

El éxito de la estimación no paramétrica con suavizamiento de kernel radica principalmente en el correcto ajuste de los parámetros como lo son el ancho de banda, donde se obtuvo un nuevo método robusto para su estimación y el coeficiente de Markov el cual es obtenido teniendo en cuenta la función de autocorrelación parcial. Los nuevos resultados son comparados con otros trabajos ya introducidos en la literatura donde se obtiene un mejor desempeño respecto a las predicciones y a las medidas de precisión. Para la elección del kernel, se introduce una técnica que, aunque requiere un esfuerzo computacional alto, es útil ya que identifica el tipo de función kernel y sus parámetros internos de mejor desempeño. Como trabajo de investigación futuro, es necesario explorar con mayor detalle el tema de la selección del kernel o considerar combinaciones lineales de funciones kernel ya introducidas, con el fin de identificar pesos y combinaciones óptimas que mejoren aún más las predicciones sin tanto esfuerzo computacional.

Referencias

- [1] Stevenson W. *Operations Management*. McGraw Hill, New York, 2009.
- [2] Rodríguez Norberto; Siado Patricia. Un pronóstico no paramétrico de la inflación colombiana. *Revista Colombiana de Estadística*, 26(2):89 – 128, 2003.
- [3] Marcella Corduas. Nonlinearity tests in time series analysis. *Statistical Methods and Applications*, 3:291–313, 10 1994.
- [4] Ahmed Radhwan, Mahmod Kamel, Mohamed Dahab, and Aboul Ella Hassanien. *Forecasting exchange rates: A chaos-based regression approach*, pages 1864–1883. 01 2015.
- [5] Carlos Alberto Zuluaga Castro. *Planeación de la producción*. Fondo Editorial Universidad EAFIT, Medellín.
- [6] Santiago Gallón and Karoll Gomez. Nonparametric time series analysis of the conditional mean and volatility functions for the cop/usd exchange rate returns. *Revista Colombiana de Estadística*, 33:25–41, 06 2010.
- [7] Lázió Györfi, W. Härdle, P. Sarda, and P. Vieu. Nonparametric curve estimation from time series. 1989.
- [8] Wolfgang Härdle. *Applied Nonparametric Regression*. Econometric Society Monographs. Cambridge University Press, 1990.
- [9] Siegfried Heiler and Nonparametric Regression. A survey on nonparametric time series analysis. 05 1999.
- [10] Inés Barbeito and Ricardo Cao. Nonparametric curve estimation and bootstrap bandwidth selection. *WIREs Computational Statistics*, 12(3):e1488, 2020.
- [11] Wanli Qiao. Asymptotics and optimal bandwidth for nonparametric estimation of density level sets. *Electron. J. Statist.*, 14(1):302–344, 2020.
- [12] S. Rahnamay Kordasiabi and S. Khazaei. Bayesian nonparametric estimation of bandwidth using mixtures of kernel estimators for length-biased data. *Journal of Statistical Computation and Simulation*, 90(10):1849–1874, 2020.
- [13] Aman Pagan, Adrian.; Ullah. *Nonparametric Econometrics*. Themes in Modern Econometrics. Cambridge University Press, 1999.
- [14] N. S. Altman. An introduction to kernel y nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185, 1992.
- [15] M. P Wand; M. C. Jones. *Kernel Smoothing*. Number 60 in Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Chapman & Hall, Boca Raton, FL, U.S., December 1994.
- [16] Larry Wasserman. *All of Nonparametric Statistics (Springer Texts in Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006.
- [17] Ronaldo Zambom, Adriano; Dias. A review of kernel density estimation with applications to econometrics. 5, 12 2012.
- [18] D. Bosq. *Nonparametric Statics for Stockhastic Processes*. Number 110 in Lecture Notes in Statistics. Springer, Boca Raton, FL, U.S., 2012.

- [19] Jeffrey D. Hart. Some automated methods of smoothing time-dependent data. *Journal of Nonparametric Statistics*, 6(2-3):115–142, 1996.
- [20] E Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33:1065–1076, 1962.
- [21] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27:832–837, 1956.
- [22] A. G. Bors; N. Nasios. Kernel bandwidth estimation for nonparametric modeling. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(6):1543–1555, Dec 2009.
- [23] Philippe Härdle, Wolfgang Karl; Vieu. Kernel regression of time series. *Journal of Time Series Analysis*, 13:209 – 232, 06 2008.
- [24] Ali; De Gooijer Jan G. Matzner-Løfber, Eric; Gannoun. Nonparametric forecasting: A comparison of three kernel-based methods. *Communications in Statistics - Theory and Methods*, 27:1593–1617, 06 2007.
- [25] Paul Deheuvels. Estimation non paramétrique de la densité par histogrammes généralisés. *Revue de Statistique Appliquée*, 25(3):5–42, 1977.
- [26] Michael Falk. On mad and comedians. *Annals of the Institute of Statistical Mathematics*, 49(4):615 – 644, 1997.
- [27] Velasco H.; Laniado H.; Toro M.; Leiva V.; Lio Y. Robust three-step regression based on comedian and its performance in cell-wise and case-wise outliers. *Mathematics*, 8(8):1–18, 2020.
- [28] Eric Matzner-Løfber ; Ali Gannoun ; Jan G. De Gooijer. Nonparametric forecasting: a comparison of three kernel-based methods. *Communications in Statistics - Theory and Methods*, 27(7):1593–1617, 1998.
- [29] L. Györfi y W. Härdle y P. Sarda y P. Vieu. *Nonparametric Curve Estimation from Time Series*. Springer-Verlag, New York, NY, 1990.
- [30] Elisabeth Rached, Imen; Larsson. *Tail Distribution and Extreme Quantile Estimation Using Non-parametric Approaches*, pages 69–87. Springer International Publishing, Cham, 2019.
- [31] A. Quintela-Del-Río ; Ph. Vieu. A nonparametric conditional mode estimate. *Journal of Nonparametric Statistics*, 8(3):253–266, 1997.
- [32] Carlos Jaime Velásquez H., Juan David; Franco C. Nota sobre la predicción del Índice de precios al consumidor usando redes neuronales artificiales. *Cuadernos de Administración*, 2010.
- [33] Juan Camilo Santana. Predicción de series temporales con redes neuronales: una aplicación a la inflación colombiana. *Revista Colombiana de Estadística*, 29:77 – 92, 06 2006.