



UNIVERSIDAD NACIONAL DE COLOMBIA

Análisis comparativo de los puentes estocásticos: simulación, estimación y bondad de ajuste

Jhonier Sebastian Rangel Gutiérrez

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Estadística.
Bogotá, Colombia
2024

Análisis comparativo de los puentes estocásticos: simulación, estimación y bondad de ajuste

Jhonier Sebastian Rangel Gutiérrez

Trabajo final presentado como requisito para optar al título de:
Magister en Ciencias Estadística

Director:
Viswanathan Arunachalam Ph.D.

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Departamento de Estadística.
Bogotá, Colombia
2024

"The only way to do great work is to love what you do"

- Steve Jobs

Agradecimientos

Querido Dios, padres Alejo y Jeimy, hermana Mariana, abuela Lety, amigos Sebas Ochoa, Cindy Echeverry, Diego Salazar, Valeria Ramírez, Felipe Basto y a todos los que han sido parte fundamental en mi camino para culminar mis estudios de maestría: quiero dedicar estas palabras de agradecimiento. Vuestro apoyo, motivación y ánimo han sido la fuerza que me ha impulsado a seguir adelante, incluso en los momentos más difíciles. Agradezco a mis maestros y en especial al director de mi trabajo, profesor Viswanathan Arunachalam Ph.D., Profesor Ramón Giraldo Henao Ph.D., Liliana Blanco Castañeda Ph.D., Rubén Darío Guevara Ph.D., Oscar Orlando Melo Ph.D. y todos aquellos que han compartido su sabiduría y experiencia conmigo.

También quiero agradecer a la Universidad Nacional de Colombia y al Departamento de Estadística de la Universidad Nacional de Colombia, por brindarme su capacitación y apoyo para lograr mis objetivos. Quiero extender mi agradecimiento al glorioso Instituto Técnico Central de La Salle, institución donde realicé mi educación básica y media y donde conocí a la profesora Amalia Cristina Torres Montiel Ms.C., quién fue mi primera guía en el mundo de la matemática y a quien debo mi amor por esta ciencia tan hermosa, a ella le envió un caluroso abrazo al cielo. A mis compañeros de pregrado y maestría, gracias por compartir sus conocimientos y por ser parte de esta gran aventura. Este logro no es solo mío, es de todos ustedes que han sido parte importante de mi camino.

Resumen

Título: Análisis comparativo de los puentes estocásticos: simulación, estimación y bondad de ajuste

En este trabajo se presenta un análisis comparativo entre el puente browniano clásico y puentes brownianos fraccionarios con diferentes parámetros de Hurst. También se aborda un enfoque para la inferencia clásica sobre los parámetros de estos modelos. Se explica además cómo simular estos procesos cuando los parámetros ya son conocidos o estimados. Asimismo, se describe una metodología para llevar a cabo pruebas de bondad de ajuste mediante el uso de técnicas de envolvimiento. Por último, se presentan tres aplicaciones en datos funcionales, estadística espacial y procesamiento del lenguaje natural.

Palabras clave: Puente Browniano; Puente Browniano Fraccional; Puente Gaussiano; Estimación; Simulación mediante métodos numéricos; Método de Euler-Maruyama; Método de Cholesky.

Abstract

Title: Comparative analysis of stochastic bridges: simulation, estimation and goodness of fit

This work presents a comparative analysis between the classic Brownian bridge and fractional Brownian bridges with different Hurst parameters. It also introduces an approach to classical inference on the parameters of these models. The simulation of these processes is explained when the parameters are either known or estimated. Additionally, a method for conducting goodness-of-fit tests using the envelopment technique is discussed. Finally, three applications in functional data, spatial statistics, and natural language processing are presented.

Key Words: Brownian Bridge; Fractional Brownian Bridge; Gaussian Bridge; Estimation; Simulation by numerical methods; Euler-Maruyama method; Cholesky method.

Contenidos

Agradecimientos	vi
Resumen	vii
Lista de Tablas	x
Lista de Figuras	xi
1 Introducción	1
2 Preliminares	4
2.1 Concepto básico	4
2.2 Convergencia de sucesiones de variables aleatorias	6
2.3 Estimación	7
2.4 Pruebas de hipótesis	10
2.5 Pruebas de bondad de ajuste	12
2.6 Conceptos básicos de procesos estocásticos	15
2.7 Métodos de simulación	21
3 Inferencia y bondad de ajuste sobre puentes estocásticos	25
3.1 Introducción	25
3.2 Puente browniano clásico	27
3.3 Inferencia y bondad de ajuste sobre el puente browniano clásico	28
3.4 Puente browniano fraccionario	31
3.5 Inferencia y bondad de ajuste sobre el puente browniano fraccionario	32
3.6 Estimación del parámetro H	35
4 Simulación y aplicaciones	37
4.1 Datos funcionales con ruidos basados en puentes	37
4.2 Puente browniano fraccionario para modelar trayectorias de migración	46
4.3 Aplicación en lenguaje natural	51
5 Conclusiones	58
6 Referencias	61

Lista de Tablas

- 4-1 Organización de las realizaciones parciales de datos funcionales 42
- 4-2 Organización de datos combinando las realizaciones parciales de las dos muestras
funcionales 42
- 4-3 Cuantiles del test Kolmogorov Smirnov para $m = 10$ 44
- 4-4 Cuantiles del test Kolmogorov Smirnov para $m = 50$ 45
- 4-5 Simulaciones de embeddings de la frase La universidad nacional es una de las
mejores universidades en Colombia con movimientos brownianos con $\sigma = 0.1$. 54
- 4-6 Simulaciones de embeddings de la frase La universidad nacional es una de las
mejores universidades en Colombia con movimientos brownianos con $\sigma = 0.05$ 55
- 4-7 Simulaciones de embeddings de la frase La universidad nacional es una de las
mejores universidades en Colombia con puentes brownianos con $\sigma = 0.1$ 56

Lista de Figuras

3-1	Resultados de la maximización de la verosimilitud.	35
4-1	Realizaciones del modelo funcional con ruidos fraccionarios, considerando diferentes parámetros de Hurst y de volatilidad, junto con la función media y bandas de confianza al 95% y 99%. https://github.com/jrangelg/Trabajo_final_maestr-a/tree/main	40
4-2	Cuantiles del test Kolmogorov Smirnov para $m = 10$ https://github.com/jrangelg/Trabajo_final_maestr-a/tree/main	44
4-3	Cuantiles del test Kolmogorov Smirnov para $m = 50$ https://github.com/jrangelg/Trabajo_final_maestr-a/tree/main	45
4-4	Modelo de migración con puente browniano clásico https://github.com/jrangelg/Trabajo_final_maestr-a/tree/main	46
4-5	Región de confianza del 95% bajo un $H = 1/2$ https://github.com/jrangelg/Trabajo_final_maestr-a/tree/main	47
4-6	Realizaciones del puente browniano fraccionario para modelar migraciones, considerando diferentes parámetros de Hurst y de volatilidad, junto con las bandas de confianza al 95%. https://github.com/jrangelg/Trabajo_final_maestr-a/tree/main	48

1 Introducción

Los puentes estocásticos son modelos estadísticos que se adaptan a situaciones en las cuales se tiene sobre una frontera valores fijos y a medida que se aleja de la frontera se aumenta la variabilidad, estos procesos tienen una gran variedad de aplicaciones, tales como: el movimiento migratorio de aves [18], ya que hay un punto inicial fijo y un punto final fijo. Las aves al partir del punto inicial se van separando lo cual en estadística se puede considerar como un aumento en la varianza y al acercarse al punto final esta distancia se va reduciendo hasta que llegan a la meta.

En ese sentido dicha varianza se vuelve cero. En las últimas investigaciones sobre las trayectorias de animales se ha expuesto la necesidad de flexibilizar los supuestos. En el año 2017 Kranstauber soluciona el problema de la isotropía haciendo uso de una distribución normal bi-variada, mientras que en 2019 [19] propone el modelo del puente browniano con covariables. En el año 2017 el mismo autor soluciona el problema de la isotropía haciendo uso de una distribución normal bi-variada, sin embargo no se han flexibilizado supuestos relacionados a la normalidad, incrementos independientes, entre otros. Otra aplicación relevante del puente browniano se encuentra en el estudio de las propiedades de los polímeros y su comportamiento en soluciones [35].

Es así que, en este contexto, el puente browniano forma estructuras entre segmentos poliméricos debido a la difusión térmica, siendo esencial para comprender las propiedades mecánicas y físicas de los polímeros, así como para la síntesis y procesamiento de materiales, en situaciones que involucran la fluidez y viscosidad de polímeros en solución, como la fabricación de recubrimientos, el puente browniano impacta la formación de películas uniformes y suaves, siendo crucial para lograr recubrimientos de alta calidad.

Estas condiciones desempeñan un papel significativo en la viscosidad y elasticidad de los materiales, influyendo en su capacidad para fluir, mezclarse y responder a esfuerzos mecánicos. El puente browniano también se presenta como la solución de la ecuación de calor bajo una frontera con una temperatura fija. Este planteamiento constituye un problema de valor en la frontera que satisface las condiciones de Dirichlet, y su solución fuerte se materializa en el puente estocástico.

Hasta ahora, los trabajos existentes han considerado el puente browniano clásico en este contexto. Sin embargo, es importante señalar que existe una generalización de esta definición que incorpora ruidos fraccionarios, ampliando así la comprensión y aplicabilidad de estos modelos. En el segundo capítulo, se aborda mediante una metodología clásica el proceso inferencial sobre el

parámetro de volatilidad y el parámetro fraccional.

Este enfoque proporciona un marco que permite analizar la variabilidad y estructura de los datos, resultando esencial para comprender los procesos estocásticos subyacentes. Dentro de este proceso de estimación del parámetro fraccionario de Hurst, se adopta la metodología de máxima verosimilitud, que implica la construcción de una grilla y la aproximación gradual al valor óptimo de H , esta técnica ofrece una aproximación efectiva. Esto muestra la necesidad de seguir investigando en la optimización analítica de la función de verosimilitud.

De esta manera la grilla proporciona una estrategia práctica para evaluar diferentes valores de H y determinar aquel que maximiza la verosimilitud del modelo en cuestión de la decisión de dejar la formalización analítica como un problema futuro sugiere un reconocimiento de la complejidad inherente al proceso y la necesidad de desarrollos adicionales en el marco teórico, teniendo en cuenta la formalización analítica de la estimación del parámetro de Hurst representa un desafío que podría requerir técnicas estadísticas avanzadas y una comprensión más profunda de las propiedades de los procesos estocásticos fraccionarios.

Así mismo en el tercer capítulo se presenta un análisis comparativo de puentes estocásticos, explorando metodologías inferenciales clásicas para estimar la volatilidad asociada a estos modelos, el cual se aborda desde la simulación de datos funcionales con ruido de puentes, proponiendo regiones de confianza y validando la bondad de ajuste, de esta manera la investigación incluye la simulación de trayectorias de animales con puentes brownianos fraccionarios y la aplicación del puente browniano fraccionario para mejorar la calibración del modelo COLOR.

De ahí que al abordar un par de ideas sobre la investigación que estás llevando a cabo por medio de la simulación de trayectorias de animales con puentes brownianos Fraccionarios bajo la exploración de la movilidad animal que utiliza la simulación de trayectorias de animales con puentes brownianos fraccionarios para comprender mejor los patrones de movimiento de diversas especies en entornos naturales. Esto podría incluir la modelización de la búsqueda de alimentos, migraciones estacionales o comportamientos territoriales bajo el impacto del entorno en el movimiento animal, investiga cómo factores ambientales como la topografía del terreno, la disponibilidad de recursos y la presencia humana influyen en las trayectorias de los animales simulados. Esto podría ayudar a prever cómo el cambio climático u otros cambios en el entorno pueden afectar los patrones de movimiento de las especies.

De esta manera en la medida que hay una optimización de la calibración del modelo relacionado con utilizar el puente browniano fraccionario para mejorar la precisión de la calibración del modelo COLOR (Coherent dialOgue pLanning approach via BrOwnian bRidge). Esto puede implicar desarrollar algoritmos que utilicen información para ajustar más eficientemente los parámetros del modelo.

Estas contribuciones resaltan la versatilidad de los puentes estocásticos en diversos contextos. En términos metodológicos, se propone una solución numérica para estimar el coeficiente fraccionario de manera eficiente, reconociendo su complejidad analítica. Además, se sugiere la aplicación de puentes estocásticos como ruido en datos funcionales, simplificando la estimación de parámetros a dos: el coeficiente de volatilidad y el coeficiente fraccionario.

Es por ello por lo que se explora la técnica de embedding para evaluar la bondad de ajuste, mostrando mayor potencia que enfoques tradicionales, la comparación de diversos puentes en trayectorias destaca la flexibilidad del puente browniano fraccionario para capturar aumentos en la volatilidad, superando limitaciones de modelos clásicos. Se sugiere evaluar este enfoque en escenarios realistas y extensos. Además, se plantea el desafío futuro de desarrollar una metodología numérica para abordar ecuaciones diferenciales estocásticas multivariadas asociadas al puente browniano multivariado, ampliando su aplicabilidad en sistemas complejos. También se motiva a la aplicación rigurosa de la estadística clásica sobre estos modelos y exploración de estas técnicas en diversos campos de la estadística como la estadística no paramétrica y la estadística bayesiana.

En este documento, el objetivo general es realizar un análisis comparativo de los puentes estocásticos con respecto al número de parámetros, simulación, estimación y bondad de ajuste, con el fin de proporcionar una metodología estadística de dichos procesos, para hacer métodos inferenciales sobre los parámetros. Para esto se tienen los siguientes objetivos específicos.

- Identificar las propiedades del puente browniano clásico y el puente browniano fraccionario para presentar un análisis de las características de cada tipo de puente estocástico.
- Describir las diferencias fundamentales entre los puentes estocásticos con respecto a la simulación, estimación de parámetros y bondad de ajuste.
- Clasificar los diferentes puentes estocásticos con respecto a los supuestos y a los usos.
- Proponer una metodología estadística para poder realizar inferencia sobre los parámetros de los diferentes puentes estocásticos

2 Preliminares

2.1 Concepto básico

Se exponen los principales resultados de la teoría de probabilidad y estadística matemática que serán usados en este trabajo. Las definiciones y teoremas que irán preparando el camino para exponer el gran teorema de Neyman-Pearson, que será utilizado posteriormente para hacer inferencia sobre los puentes estocásticos. La teoría de la probabilidad juega un papel muy importante a la hora de trabajar con procesos estocásticos, por eso se hace un recuento breve sobre los fundamentos básicos de esta área, si se desea profundizar, se sugiere el siguiente libro [26]

Definición. 1 Sea \mathcal{F} una familia de conjuntos contenidos en Ω , se dice que \mathcal{F} es σ -álgebra sobre Ω si y solo si

- $\Omega \in \mathcal{F}$
- Si $A \in \mathcal{F}$ entonces $A^c \in \mathcal{F}$
- Si $A_1, A_2, \dots \in \mathcal{F}$ entonces

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}.$$

Esta familia contiene a los conjuntos a los cuales se les puede asociar un grado de incertidumbre; por lo tanto, si una σ -álgebra contiene a otra, la σ -álgebra más grande contiene más información del experimento, este razonamiento es crucial a la hora de hablar de filtraciones y martingalas. A la dupla (Ω, \mathcal{F}) se le llama espacio medible. El espacio medible usual sobre los números reales es el espacio de Borel $(\mathbb{R}, B(\mathbb{R}))$, el cual está conformado por el conjunto de los números reales y la menor σ -álgebra que contiene a los intervalos abiertos.

Definición. 2 Sea P una función cuyo dominio es \mathcal{F} y cuyo co-dominio es \mathbb{R} . P es una medida de probabilidad si y solo si

- $P(\Omega) = 1$
- $P(A) \geq 0$

- Si $A_1, A_2, \dots \in \mathcal{F}$ disyuntos dos a dos entonces

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

(Ω, \mathcal{F}, P) se le llama espacio de probabilidad. Note que la información de la incertidumbre de un experimento aleatorio se condensa en el espacio muestral y en la medida de probabilidad. En muchos casos se hace estudio de una característica determinada sobre cada elemento del espacio muestral, a dichas características se les denomina variables aleatorias y se les denota con letras mayúsculas.

Definición. 3 Sea X una función cuyo dominio es el espacio muestral y cuyo co-dominio es el conjunto de los números reales, X es $B(\mathbb{R})$ - \mathcal{F} variable aleatoria si y solo si para todo $B \in B(\mathbb{R})$ se tiene que $X^{-1}(B) \in \mathcal{F}$

En este trabajo se dirá simplemente que X es una variable aleatoria real y se entenderá como una función de valor real medible con respecto a $B(\mathbb{R})$ - \mathcal{F} . No en todos los casos se considerarán como valores posibles a todo el conjunto de los números reales. Por eso, se procede a trabajar con el soporte de la variable definido como $\aleph = \{x \in \mathbb{R} | f_X(x) > 0\}$. Por otro lado se usará notación de Lebesgue para la definición de momentos de una variable, si desea profundizar más sobre dicha integral puede estudiarla en [26].

Definición. 4 Sea (Ω, \mathcal{F}, P) un espacio de probabilidad y X una variable aleatoria real sobre él. Se le llama esperanza o valor esperado de X , al primer momento, es decir:

$$\mu_X = E(X) = \int_{\Omega} X dP = \int_{\aleph} x dF(x) = \begin{cases} \sum_{x \in \aleph} x P(X = x) & \text{discreta} \\ \int_{\aleph} x f_X(x) dx & \text{continua} \end{cases}$$

Definición. 5 Sea (Ω, \mathcal{F}, P) un espacio de probabilidad y X una variable aleatoria real sobre él. Se le llama varianza o segundo momento centrado a:

$$Var(X) = E((X - \mu_X)^2) = E(X^2) - \mu_X^2.$$

La esperanza y la varianza de una variable aleatoria no siempre existen; por ejemplo, en la distribución Cauchy. Por eso es relevante definir el espacio $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ como el conjunto de variables aleatorias sobre (Ω, \mathcal{F}, P) que tienen esperanza y varianza finitas. La esperanza y la varianza son las características más usadas de una variable aleatoria, sin embargo para realizar inferencia sobre parámetros es necesario distinguir cuando un momento está centrado en cero y cuando está centrado por la media.

Hay espacios más generales que el espacio $\mathcal{L}^2(\Omega, \mathcal{F}, P)$, por ejemplo al espacio de las variables reales sobre (Ω, \mathcal{F}, P) tales que $E|X|^r < \infty$ se le llama $\mathcal{L}^r(\Omega, \mathcal{F}, P)$, usualmente $r \in \mathbb{N}$.

Definición. 6 Sea $X \in \mathcal{L}^r(\Omega, \mathcal{F}, P)$ una variable aleatoria real. Se define el r -ésimo momento centrado en cero a lo siguiente:

$$\mu'_r = E(X^r).$$

Note que $\mu_X = \mu'_1$

Definición. 7 Sea $X \in \mathcal{L}^r(\Omega, \mathcal{F}, P)$ una variable aleatoria real. Se define el r -ésimo momento centrado por la media a lo siguiente:

$$\mu_r = E((X - \mu_X)^r).$$

Note que $\sigma_X^2 = \mu_2$, esta notación de los momentos es la más usada y dichas definiciones serán cruciales a la hora de realizar estimadores.

Otro tema importante para considerar las características de un buen estimador es la convergencia estocástica, para hacer profundidad en el tema se sigue el siguiente texto [8]

2.2 Convergencia de sucesiones de variables aleatorias

Definición. 8 Sea X, X_1, X_2, \dots una sucesión de variables aleatorias sobre el mismo espacio de probabilidad, se dice que X_n converge casi seguramente si y solo si:

$$P(\{\omega \in \Omega \mid \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}) = 1$$

la notación que se usará en este trabajo es

$$X_n \xrightarrow{\text{c.s.}} X.$$

Definición. 9 Sea X, X_1, X_2, \dots una sucesión de variables aleatorias sobre el mismo espacio de probabilidad, se dice que X_n converge estocásticamente si y solo si para todo $\epsilon > 0$

$$\lim_{n \rightarrow \infty} P(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \epsilon\}) = 0$$

la notación que se usará en este trabajo es

$$X_n \xrightarrow{P} X.$$

Definición. 10 Sea X, X_1, X_2, \dots una sucesión de variables aleatorias sobre el espacio $\mathcal{L}^r(\Omega, \mathcal{F}, P)$. Se dice que X_n converge en r -media si y solo si

$$\lim_{n \rightarrow \infty} E(|X_n - X|^r) = 0$$

la notación a ser usada es

$$X_n \xrightarrow{L^r} X.$$

Definición. 11 Sea X, X_1, X_2, \dots una sucesión de variables aleatorias, se dice que X_n converge en distribución si y solo si

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$$

la notación que se usará en este trabajo es

$$X_n \xrightarrow{D} X.$$

Luis Rincón en su libro [29] demuestra la relación entre los diferentes tipos de convergencia, es decir, se verifica que si una familia de variables aleatorias convergen casi seguramente, entonces dicha familia converge en probabilidad. Si una familia de variables aleatorias convergen en L_r , entonces dicha familia converge en probabilidad. Por último si una familia de variables aleatorias convergen en probabilidad, entonces dicha familia converge en distribución.

2.3 Estimación

En la teoría clásica de inferencia estadística, por ese motivo se incluyen algunas definiciones y resultados de la estadística matemática, con énfasis en el teorema de Neyman Pearson que ayuda a construir las pruebas de hipótesis uniformemente más potentes, si desea profundizar en esta área de la estadística se recomienda el texto [4].

Definición. 12 Dada una distribución $f_X(x|\theta)$, se define espacio paramétrico al conjunto de los posibles parámetros θ que caracteriza completamente a la distribución.

Por ejemplo el espacio paramétrico de la varianza es $\Theta = (0, \infty)$

Definición. 13 Una muestra aleatoria es una familia finita de variables aleatorias independientes e idénticamente distribuidas. La notación que se usará es: $X_1, \dots, X_n \stackrel{iid}{\sim} f_X(x|\theta)$.

En la vida real solo se tiene la realización de la muestra aleatoria, cuya notación será x_1, \dots, x_n

Definición. 14 A una función de la muestra aleatoria se le llama estadístico y si dicho estadístico tiene características apropiadas para aproximarse al parámetro recibe el nombre de estimador.

Por ejemplo, $T(X_1, \dots, X_n) = \sum_{i=1}^n X_i$ es un estadístico y $\hat{\theta}(X_1, \dots, X_n) = \frac{\sum_{i=1}^n X_i}{n}$ es un estimador para la media. La pregunta que sigue y que es de gran importancia en el área de la estadística matemática es: ¿Cómo encontrar estimadores?, para responder a esta pregunta se recuerdan lo métodos más usados para encontrar estimadores

Definición. 15 Sean $X_1, \dots, X_n \stackrel{iid}{\sim} f_X(x|\theta)$, la función de verosimilitud se define como la distribución conjunta de la muestra vista como función del parámetro.

$$L(\theta|X_1 = x_1, \dots, X_n = x_n) := \prod_{i=1}^n f_{X_i}(x_i|\theta).$$

Definición. 16 Sean $X_1, \dots, X_n \stackrel{iid}{\sim} f_X(x|\theta)$. Se define el estimador de máximo verosimilitud, cuando existe, como sigue:

$$\hat{\theta}_{MV}(X_1, \dots, X_n) := \underset{\theta \in \Theta}{\operatorname{argmax}} \{L(\theta|X_1 = x_1, \dots, X_n = x_n) := \prod_{i=1}^n f_{X_i}(x_i|\theta)\}.$$

Definición. 17 Sean $X_1, \dots, X_n \stackrel{iid}{\sim} f_X(x|\theta)$, se define el estimador por el método de momentos, si existe, como $\hat{\theta}_{MOM}(X_1, \dots, X_n)$ solución del siguiente sistema, de ecuaciones:

$$\begin{cases} \mu'_1(\hat{\theta}) = \sum_{i=1}^n \frac{X_i^1}{n} \\ \vdots \\ \mu'_p(\hat{\theta}) = \sum_{i=1}^n \frac{X_i^p}{n} \end{cases}.$$

La cantidad de momentos a considerar generalmente dependerá de la dimensión del espacio paramétrico.

Ahora se considera la siguiente pregunta: ¿Es $\hat{\theta}(X_1, \dots, X_n)$ un buen estimador?, esta pregunta lleva a definir las características que debe cumplir un buen estimador.

Definición. 18 Sean $X_1, \dots, X_n \stackrel{iid}{\sim} f_X(x|\theta)$ y $\hat{\theta}(X_1, \dots, X_n)$ un estimador para θ , $\hat{\theta}(X_1, \dots, X_n)$ es insesgado para θ , si y solo si

$$E(\hat{\theta}(X_1, \dots, X_n)) = \theta.$$

Definición. 19 Sean $X_1, \dots, X_n \stackrel{iid}{\sim} f_X(x|\theta)$, sean $\hat{\theta}_1(X_1, \dots, X_n)$ y $\hat{\theta}_2(X_1, \dots, X_n)$ estimadores insesgados para θ , $\hat{\theta}_1(X_1, \dots, X_n)$ es más eficiente en varianza $\hat{\theta}_2(X_1, \dots, X_n)$, si y solo si

$$\operatorname{Var}(\hat{\theta}_1(X_1, \dots, X_n)) < \operatorname{Var}(\hat{\theta}_2(X_1, \dots, X_n)).$$

Definición. 20 Al estimador insesgado, que considera toda la información muestral y que tiene mínima varianza se le conoce como $UMVU(\theta)$ (Uniformly minimum variance unbiased).

En muchos casos, con las definiciones directas, no es sencillo construir el $UMVU$ de un parámetro, sin embargo, existen resultados que ayudan a encontrar fácilmente el $UMVU$, para ello es necesario recordar las siguientes definiciones

Definición. 21 Sean $X_1, \dots, X_n \stackrel{iid}{\sim} f_X(x|\theta)$ y $T(X_1, \dots, X_n)$ una estadística, $T(X_1, \dots, X_n)$ es suficiente para θ , si y solo si $f_{(X_1, \dots, X_n)|T=t}(x_1, \dots, x_n|t)$ no depende de θ .

No siempre es sencillo demostrar de manera directa la suficiencia, el teorema de la factorización de Fisher-Neyman es una equivalencia que caracteriza a los estadísticos suficientes

Teorema. 1 (Fisher-Neyman) Sean $X_1, \dots, X_n \stackrel{iid}{\sim} f_X(x|\theta)$. Una estadística $T(X_1, \dots, X_n)$ es suficiente, si y solo si, existe funciones g y h , tales que

$$L(\theta|X_1, \dots, X_n) = g(T(X_1, \dots, X_n), \theta)h(X_1, \dots, X_n)$$

Definición. 22 Sean $X_1, \dots, X_n \stackrel{iid}{\sim} f_X(x|\theta)$ y $T(X_1, \dots, X_n)$ una estadística, $T(X_1, \dots, X_n)$ es completa para θ si y solo si $E(U(T(X_1, \dots, X_n))) = 0$ implica $U \stackrel{c.s}{=} 0$.

Al igual que la suficiencia, no siempre es sencillo demostrar de manera directa la completitud de una estadística, pero la familia exponencial de distribuciones ayuda a encontrar estadísticas suficientes y completas

Definición. 23 Sean $X_1, \dots, X_n \stackrel{iid}{\sim} f_X(x|\theta)$, f_X pertenece a la familia exponencial de distribuciones si y solo si existen funciones T, c, d, S

$$L(\theta|X_1, \dots, X_n) = \exp\{T(X_1, \dots, X_n)^t c(\theta) + d(\theta) + S(X_1, \dots, X_n)\} I_{\mathbb{R}^n}(X_1, \dots, X_n)$$

Si desea profundizar en las propiedades de la familia exponencial puede ver [37]

Teorema. 2 Sean $X_1, \dots, X_n \stackrel{iid}{\sim} f_X(x|\theta)$, si f_X pertenece a la familia exponencial de distribuciones entonces $T(X_1, \dots, X_n)$ es suficiente y completa para θ , con $\theta \in \text{int}(c)$

Este teorema sobre la familia exponencial facilita la verificación de la suficiencia y la completitud sobre una estadística, dichas propiedades son importantes para el teorema de Lehmann Scheffé

Teorema. 3 (Lehmann Scheffé) Sean $X_1, \dots, X_n \stackrel{iid}{\sim} f_X(x|\theta)$, sea $T(X_1, \dots, X_n)$ una estadística suficiente y completa y sea $\hat{\theta}_1(X_1, \dots, X_n)$ un estimador insesgado para θ , el estimador

$$\hat{\theta}_2(X_1, \dots, X_n) := E(\hat{\theta}_1(X_1, \dots, X_n)|T(X_1, \dots, X_n))$$

es el *UMVU* de θ .

En la practica es más usado el siguiente corolario

Corolario. 1 Si $\hat{\theta}(X_1, \dots, X_n)$ es insesgado y es función de una estadística suficiente y completa ,entonces el *UMVU* $(\theta) = \hat{\theta}(X_1, \dots, X_n)$.

Luego de tener buenos estimadores se plantea el requerimiento de dos cotas para el parámetro, es decir se desea estimar por medio de un intervalo

Definición. 24 Sean $X_1, \dots, X_n \stackrel{iid}{\sim} f_X(x|\theta)$, un intervalo de probabilidad $(1 - \alpha)100\%$ es un par de estadísticas $L(X_1, \dots, X_n)$ y $U(X_1, \dots, X_n)$, tales que

$$P(L(X_1, \dots, X_n) \leq \theta \leq U(X_1, \dots, X_n)) = 1 - \alpha.$$

En la vida real solo se tiene la realización de un intervalo de probabilidad llamado intervalo de confianza

$$IC(\theta) := (l(x_1, \dots, x_n), u(x_1, \dots, x_n)).$$

Ahora la pregunta que nos surge es: ¿Dicho intervalo de confianza es el mejor?, existen muchos caminos para verificar si un intervalo de confianza es el mejor, el más usado, y que en algunos casos tiene relación con la prueba de hipótesis uniformemente más potente, es el intervalo de mínima longitud.

Definición. 25 Sean $X_1, \dots, X_n \stackrel{iid}{\sim} f_X(x|\theta)$ y sea $IC(\theta) = (L(X_1, \dots, X_n), U(X_1, \dots, X_n))$ un intervalo de confianza para θ , $IC(\theta)$ es el intervalo de confianza $(1 - \alpha)100\%$ de menor longitud si y solo si para todo intervalo de confianza $(1 - \alpha)100\%$ el $IC(\theta)$ tiene menor longitud

Bajo la distribución normal existe un intervalo de confianza para la media, sin embargo, en la vida real no se conoce la varianza, esto llevo a que en 1908 William Sealy Gosset, quien trabajaba en una fábrica de cerveza, definiera la distribución $t - student$. Se puede verificar que el intervalo de confianza de menor longitud para la media es $[\bar{X} - t_{\alpha/2, n-1} \frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2, n-1} \frac{S}{\sqrt{n}}]$ al nivel de confianza $(1 - \alpha)100\%$. Si desea profundizar sobre la historia de esta distribución puede ver [27] y [14].

2.4 Pruebas de hipótesis

Las pruebas de hipótesis ayudan a contrastar la plausibilidad de afirmaciones sobre parámetros de una población. En esta parte se verán algunos fundamentos de las pruebas estadísticas, encaminados al teorema de Neyman Pearson que se usará para realizar inferencia sobre los puentes estocásticos. Si desea profundizar estos temas puede ver [4]

Definición. 26 Sean $X_1, \dots, X_n \stackrel{iid}{\sim} f_X(x|\theta)$, una hipótesis sobre el parámetro θ es una afirmación matemática que se realiza sobre él.

Existen dos tipos de hipótesis sobre parámetros. Las hipótesis simples caracterizan completamente la distribución, por ejemplo: $H : \theta = \theta_0$, $H : \mu \leq 5$. Si una hipótesis no es simple recibe el nombre de compuesta.

Definición. 27 Sean $X_1, \dots, X_n \stackrel{iid}{\sim} f_X(x|\theta)$, y sean θ_0 y θ_1 dos elementos distintos del espacio paramétrico, se define un contraste de hipótesis simples al par de hipótesis:

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{cases}$$

a H_0 se le llama hipótesis nula y a H_1 se le llama hipótesis alterna.

Las conclusiones sobre un contraste de hipótesis se indican sobre la hipótesis nula, y las reglas de decisión se definen bajo la hipótesis nula, esto nos lleva a definir formalmente una región de rechazo

Definición. 28 Sean $X_1, \dots, X_n \stackrel{iid}{\sim} f_X(x|\theta)$, y sea

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{cases}$$

un contraste de hipótesis simple. Se dice que $\mathcal{C} \subseteq \mathfrak{N}^n$ es una región de rechazo de significancia α si y solo si

$$P((X_1, \dots, X_n) \in \mathcal{C} | \theta = \theta_0) = \alpha.$$

Cuando un contraste de hipótesis es complementario, es decir que solo puede ocurrir H_0 o H_1 , se pueden cometer dos tipos de error, el error tipo uno corresponde a rechazar H_0 cuando H_0 es cierto, es decir $P((X_1, \dots, X_n) \in \mathcal{C} | \theta = \theta_0) = \alpha$ y el error tipo dos corresponde a no rechazar H_0 cuando H_1 es cierto, es decir $P((X_1, \dots, X_n) \notin \mathcal{C} | \theta = \theta_1) = \beta(\theta_1)$. Lo ideal es que las probabilidades de cometer estos errores sean pequeñas, sin embargo en la vida real se define el nivel de significancia previo al estudio, por ese motivo el error tipo dos no se puede controlar, sin embargo, se pueden buscar regiones óptimas que permitan tener menor error tipo dos dado un nivel de significancia, dicha región recibe el nombre de más potente, esto se debe a la siguiente definición

Definición. 29 Sean $X_1, \dots, X_n \stackrel{iid}{\sim} f_X(x|\theta)$, considere el siguiente sistema de hipótesis

$$\begin{cases} H_0 : \theta \in \Theta_0 \\ H_1 : \theta \in \theta_1 \end{cases}$$

se define la función potencia sobre el espacio alterno θ_1 como sigue :

$$\begin{aligned} \pi : \theta_1 &\rightarrow [0, 1] \\ \theta_1 &\mapsto \pi(\theta_1) := 1 - \beta(\theta_1). \end{aligned}$$

Definición. 30 Sean $X_1, \dots, X_n \stackrel{iid}{\sim} f_X(x|\theta)$, considere el siguiente sistema de hipótesis

$$\begin{cases} H_0 : \theta \in \Theta_0 \\ H_1 : \theta \in \theta_1 \end{cases}$$

Sea \mathcal{C} una región crítica de nivel α , \mathcal{C} es la región crítica uniformemente más potente, si y solo si, para toda $\mathcal{C}' \subseteq \mathfrak{N}^n$ región crítica de significancia α y para todo $\theta_1 \in \theta_1$, se cumple que $\pi(\theta_1) \leq \pi'(\theta_1)$, con π y π' las funciones de potencia de \mathcal{C} y \mathcal{C}' respectivamente.

De manera directa es difícil encontrar regiones de rechazo uniformemente más potente, sin embargo el siguiente teorema va a demostrar que el test de razón de verosimilitud define la región de rechazo uniformemente más potente.

Teorema. 4 (Neyman-Pearson) Sean $X_1, \dots, X_n \stackrel{iid}{\sim} f_X(x|\theta)$, y sea

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{cases}$$

la región

$$\mathcal{C} := \{(x_1, \dots, x_n) \in \mathbb{N}^m \mid \lambda(x_1, \dots, x_n) := \frac{L(x_1, \dots, x_n|\theta_1)}{L(x_1, \dots, x_n|\theta_0)} > c\},$$

con $c > 0$ es la región de rechazo uniformemente más potente.

Al estadístico $\lambda(X_1, \dots, X_n) = \frac{L(X_1, \dots, X_n|\theta_1)}{L(X_1, \dots, X_n|\theta_0)}$ se le conoce con el nombre de razón de verosimilitud y para evaluar un contraste de hipótesis con dicho estadístico se puede utilizar la siguiente convergencia $-2\ln(\lambda(X_1, \dots, X_n)) \xrightarrow{D} \chi_{dim(\theta_1) - dim(\theta_0)}^2$, donde $dim(\theta_1)$ y $dim(\theta_0)$ corresponden a las dimensiones de θ_1 y θ_0 . Puede ocurrir que $dim(\theta_1) = dim(\theta_0)$, en estos casos no se puede hacer uso de la convergencia, sin embargo se pueden usar métodos de simulación para la distribución del estadístico de razón de verosimilitud bajo la hipótesis nula.

Bajo la distribución normal se puede demostrar que el complemento del intervalo de confianza para la media, con varianza desconocida y usando los cuantiles simétricos de la t - *student* es la región de rechazo uniformemente más potente sobre $H_0 : \mu = \mu_0$.

2.5 Pruebas de bondad de ajuste

En la vida real se tienen datos observados cuya notación es x_1, \dots, x_n . Nuestro objetivo es realizar inferencia sobre los parámetros del modelo de probabilidad del cual provienen los datos, una pregunta natural que surge antes de realizar inferencia es: ¿De qué distribución provienen estos datos? Un camino para dar respuesta a esta pregunta es hacer uso de las llamadas pruebas de bondad de ajuste. Aquí se dará un breve recuento sobre ellas, si desea profundizar más sobre dichas pruebas se sugiere el siguiente libro [36]

- **Prueba de bondad χ^2**

La prueba de bondad de ajuste χ^2 es un método estadístico utilizado para determinar si existe una diferencia significativa entre la distribución observada de los datos y una distribución teórica esperada. Esta prueba se basa en el cálculo de una estadística χ^2 que sigue una distribución chi-cuadrado.

El proceso para hacer uso de esta prueba se basa en los siguientes pasos:

- Hipótesis nula (H_0): Se plantea una distribución teórica esperada y se asume que los datos observados se ajustan a esta distribución. La hipótesis nula establece que no

hay diferencia significativa entre la distribución observada y la distribución teórica esperada.

- Hipótesis alternativa (H_1): Se plantea una distribución alternativa que difiere de la distribución teórica esperada. La hipótesis alternativa sugiere que existe una diferencia significativa entre la distribución observada y la distribución teórica esperada.
- Cálculo de la estadística χ^2 : Se calcula la estadística χ^2 sumando las diferencias al cuadrado entre las frecuencias observadas y las frecuencias esperadas, ponderadas por las frecuencias esperadas. La fórmula para calcular la estadística χ^2 es la siguiente:

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

Donde: χ^2 es la estadística χ^2 , O_i son las frecuencias observadas en cada categoría, E_i son las frecuencias esperadas en cada categoría.

Un problema de esta prueba es la distribución de la estadística, ya que usa dos convergencias, la primera es la convergencia de la binomial a la poisson, esto para que la varianza de la frecuencia observada sea igual a la frecuencia esperada, y después de esto considera el teorema de límite central, para hacer uso de esta prueba se sugiere que el número de grados de libertad sea grande.

• Prueba de bondad de Kolmogorov Smirnov

La prueba de bondad de ajuste de Kolmogorov-Smirnov es un método estadístico utilizado para determinar si una muestra de datos sigue una distribución continua específica. Esta prueba se basa en la comparación de la Función de Distribución acumulativa Empírica (FDE) de los datos observados con la Función de Distribución Acumulativa (FDA) teórica esperada.

El proceso para llevar a cabo dicho contraste de hipótesis se basa en los siguientes pasos:

- Hipótesis nula ($H_0: X \sim F$): Se plantea una distribución teórica esperada y se asume que los datos observados se ajustan a esta distribución. La hipótesis nula establece que no hay diferencia significativa entre la distribución observada y la distribución teórica esperada.
- Hipótesis alternativa ($H_1: X \not\sim F$): Se plantea una distribución alternativa que difiere de la distribución teórica esperada. La hipótesis alternativa sugiere que existe una diferencia significativa entre la distribución observada y la distribución teórica esperada.
- Cálculo de la estadística de prueba (D): La estadística de prueba (D) se define como la máxima diferencia vertical entre la FDE de los datos observados y la FDA teórica esperada. La fórmula para calcular D es la siguiente:

$$D = \max_i (|F_i - F(x_i)|)$$

Donde: D es la estadística de prueba, F_i es la FDE en el i-ésimo punto de datos, $F(x_i)$ es la FDA teórica evaluada en el i-ésimo punto de datos.

La gran ventaja de esta prueba es que se puede utilizar para cualquier distribución continua. Esta prueba se sugiere usar con un tamaño de muestra grande cuando la distribución teórica es la distribución normal y los estimadores son los corregidos por Lilliefors.

• Prueba de bondad de Shapiro Wilks

La prueba de bondad de ajuste de Shapiro-Wilk es un método estadístico utilizado para determinar si una muestra de datos sigue una distribución normal. Esta prueba se basa en la comparación de los valores observados con los valores esperados bajo la hipótesis de normalidad.

El proceso para llevar a cabo dicho contraste de hipótesis se basa en los siguientes pasos:

- Hipótesis nula ($H_0: X \sim F$): Se plantea que los datos observados siguen una distribución normal. La hipótesis nula establece que no hay diferencia significativa entre la distribución observada y la distribución normal esperada.

- Hipótesis alternativa ($H_1: X \not\sim F$): Se plantea que los datos observados no siguen una distribución normal. La hipótesis alternativa sugiere que existe una diferencia significativa entre la distribución observada y la distribución normal esperada.
- Cálculo de la estadística de prueba (W): La estadística de prueba (W) se calcula utilizando una combinación de los coeficientes de regresión y la varianza residual obtenidos de un modelo de regresión ajustado a los datos. La fórmula para calcular W es la siguiente:

$$W = \frac{\left(\sum_{i=1}^n a_i x_{(i)} \right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Donde: W es la estadística de prueba, a_i son los coeficientes de regresión, $x_{(i)}$ son los valores observados ordenados de menor a mayor, x_i son los valores observados, \bar{x} es la media de los valores observados.

La prueba de bondad de ajuste de Shapiro-Wilk es utilizada para evaluar si una muestra de datos sigue una distribución normal. Esta prueba se sugiere usar cuando el tamaño de muestra es pequeño. Existen muchas más pruebas de normalidad, sin embargo en este trabajo se van a considerar principalmente la prueba de Lilliefors y Shapiro-Wilks.

2.6 Conceptos básicos de procesos estocásticos

Los puentes estocásticos son procesos estocásticos, por este motivo se hace necesario realizar una recopilación de los resultados importantes de dicha teoría. Se dará énfasis a la definición del movimiento browniano fraccionario. Este ha sido de interés en los últimos años ya que debilita el supuesto de incrementos independientes sobre los ruidos. Se sugieren tres libros para profundizar en el tema [21, 25]

Definición. 31 (Proceso estocástico a tiempo continuo.) Un proceso estocástico en tiempo continuo es una familia de variables aleatorias $\{X_t\}_{t \geq 0}$ sobre un espacio de probabilidad (Ω, \mathcal{F}, P)

Ahora se explicarán algunas generalizaciones de medidas de tendencia central, de dispersión y de correlación entre dos variables pertenecientes a un proceso estocástico compuesto por variables cuadrado integrables.

Definición. 32 (Función de media de un proceso estocástico) Sea $\{X_t\}_{t \geq 0}$ un proceso estocástico en tiempo continuo se define la función de media como sigue:

$$\mu_t := E(X_t).$$

Definición. 33 (Función de autocovarianza de un proceso estocástico) Sea $\{X_t\}_{t \geq 0}$ un proceso estocástico en tiempo continuo se define la función de autocovarianza como sigue:

$$\gamma_{s,t} := Cov(X_s, X_t).$$

Definición. 34 (Función de autocorrelación de un proceso estocástico) Sea $\{X_t\}_{t \geq 0}$ un proceso estocástico en tiempo continuo se define la función de autocovarianza como sigue:

$$\rho_{s,t} := \frac{Cov(X_s, X_t)}{\sqrt{Var(X_s)}\sqrt{Var(X_t)}}.$$

Definición. 35 (Proceso estocástico con incrementos independientes) Sea $\{X_t\}_{t \geq 0}$ un proceso estocástico en tiempo continuo se dice que $\{X_t\}_{t \geq 0}$ es un proceso estocástico con incrementos independientes si y solo si:

$$X_u - X_t \text{ ind } X_t - X_s$$

Para todo $u \geq t \geq s \geq 0$.

Definición. 36 (Proceso estocástico con incrementos estacionarios) Sea $\{X_t\}_{t \geq 0}$ un proceso estocástico en tiempo continuo se dice que $\{X_t\}_{t \geq 0}$ es un proceso estocástico con incrementos estacionarios si y solo si:

$$X_{t+h} - X_t \stackrel{d}{=} X_h - X_0$$

Para todo $t, h \geq 0$.

Definición. 37 (Proceso estocástico estacionario en sentido fuerte) Sea $\{X_t\}_{t \geq 0}$ un proceso estocástico en tiempo continuo se dice que $\{X_t\}_{t \geq 0}$ es un proceso estocástico estacionario en sentido fuerte si y solo si:

$$(X_{t_1}, X_{t_2}, \dots, X_{t_n}) \stackrel{d}{=} (X_{t_1+h}, X_{t_2+h}, \dots, X_{t_n+h})$$

Para todo $t_1, \dots, t_n, h \geq 0$ y para todo $n \in \mathbb{N}$.

Definición. 38 (Proceso estocástico estacionario en sentido débil) Sea $\{X_t\}_{t \geq 0}$ un proceso estocástico en tiempo continuo se dice que $\{X_t\}_{t \geq 0}$ es un proceso estocástico estacionario en sentido débil si y solo si:

$$\mu_t = \mu$$

$$\gamma_{t,t+h} = c(h)$$

Para todo $t, h \geq 0$.

Proposición. 1 Todo proceso estocástico estacionario en sentido fuerte es estacionario en sentido débil.

Demostración.

Se verifican las dos propiedades.

- $\mu_t = E(X_t) = E(X_s)$, para todo $s, t \geq 0$, entonces, $\mu_t = \mu$ para todo $t \geq 0$.
- Como el proceso es estacionario en sentido fuerte $(X_t, X_{t+h}) \stackrel{d}{=} (X_{t'}, X_{t'+h})$, entonces, $\gamma_{t,t+h} = \gamma_{t',t'+h}$, para todo $s, t, t' \geq 0$, entonces, $\gamma_{t,t+h} = c(h)$ para todo $t, h \geq 0$.

□

Definición. 39 (Proceso gaussiano) Sea $\{X_t\}_{t \geq 0}$ un proceso estocástico en tiempo continuo se dice que $\{X_t\}_{t \geq 0}$ es un proceso estocástico gaussiano, si y solo si, para todo $t_1, \dots, t_n \geq 0$ y para todo $n \in \mathbb{N}$, se cumple:

$$(X_1, \dots, X_n) \sim N_n(\mu, \Sigma).$$

Proposición. 2 Un proceso estocástico gaussiano estacionario en sentido débil es estacionario en sentido fuerte.

Demostración.

Sean $t_1, \dots, t_n, h \geq 0$, considerando la función generadora de momentos, se tiene que:

$$M_{(X_{t_1+h}, \dots, X_{t_n+h})}(k_1, \dots, k_n) = \exp \left\{ (k_1 \ \dots \ k_n) \begin{pmatrix} \mu_{t_1+h} \\ \vdots \\ \mu_{t_n+h} \end{pmatrix} + \frac{(k_1 \ \dots \ k_n) \begin{pmatrix} \gamma_{t_1+h, t_1+h} & \gamma_{t_n+h, t_1+h} \\ \gamma_{t_n+h, t_1} & \gamma_{t_n+h, t_n+h} \end{pmatrix} \begin{pmatrix} k_1 \\ \vdots \\ k_n \end{pmatrix}}{2} \right\}$$

Como el proceso es estacionario en sentido débil entonces, $\mu_{t_i+h} = \mu$ y $\gamma_{t_i+h, t_j+h} = c(|t_i - t_j|)$ para todo $i, j = 1, \dots, n$.

$$= \exp \left\{ (k_1 \ \dots \ k_n) \begin{pmatrix} \mu_{t_1} \\ \vdots \\ \mu_{t_n} \end{pmatrix} + \frac{(k_1 \ \dots \ k_n) \begin{pmatrix} \gamma_{t_1, t_1} & \gamma_{t_n, t_1} \\ \gamma_{t_n, t_1} & \gamma_{t_n, t_n} \end{pmatrix} \begin{pmatrix} k_1 \\ \vdots \\ k_n \end{pmatrix}}{2} \right\}$$

$$= M_{(X_{t_1}, \dots, X_{t_n})}(k_1, \dots, k_n)$$

y como la función generadora de momentos caracteriza completamente la distribución, se concluye que:

$$(X_{t_1}, X_{t_2}, \dots, X_{t_n}) \stackrel{d}{=} (X_{t_1+h}, X_{t_2+h}, \dots, X_{t_n+h})$$

y por definición el proceso gaussiano es fuertemente estacionario.

□

Los modelos estadísticos contienen una parte determinista y una parte aleatoria o de ruido. En los modelos de regresión y de series de tiempo, se considera un ruido blanco o una caminata aleatoria ponderada. Un problema al trabajar con procesos estocásticos en tiempo continuo es la definición de este ruido.

La solución a este problema es construir un ruido parametrizado sobre un espacio continuo de tal manera que los subconjuntos finito-dimensionales cumplan con la característica de ruido blanco o una caminata aleatoria usual, el proceso que cumple con estas características es el movimiento browniano, con el cual se modela el ruido en las ecuaciones diferenciales estocásticas clásicas.

Existen dos perspectivas sobre la construcción del movimiento browniano. La primera es la perspectiva estadística que considera un modelo ruidoso en un dominio continuo, y la segunda perspectiva está relacionada con la función de difusión del calor. En esta sección nos centraremos en la versión estadística y utilizaremos algunos resultados de la física para modelar el ruido en un modelo determinista. Se pueden encontrar más detalles sobre esta construcción en [5], [10], [28] y [28].

Definición. 40 El movimiento browniano unidimensional estándar B_t es un proceso estocástico $(B_t)_{t \geq 0}$ con trayectorias continuas, que comienza en cero, y que cumple lo siguiente:

$$i \ B_0 \stackrel{c.s.}{=} 0.$$

- ii El proceso tiene incrementos independientes, es decir, para $0 \leq s < t$, el incremento $B_t - B_s$ es independiente de los valores de $(B_u)_{0 \leq u < s}$.
- iii El proceso tiene incrementos estacionarios, es decir, para $0 \leq r < s < t$, el incremento $B_t - B_s$ tiene la misma distribución que el incremento $B_s - B_r$.
- iv El incremento $B_t - B_s$ tiene una distribución $\mathcal{N}(0, t - s)$.
- v Las trayectorias de dicho proceso son continuas con probabilidad 1.

Definición. 41 Un movimiento browniano con parámetro σ es un proceso estocástico en tiempo continuo $\{\sigma B_t\}_{t \geq 0}$ donde $\{B_t\}_{t \geq 0}$ es un movimiento browniano estándar.

Proposición. 3 (Caracterización de Paul Lévy) Sea $\{B_t\}_{t \geq 0}$ un proceso estocástico en tiempo continuo. Si $\{B_t\}_{t \geq 0}$ satisface:

- i $\{B_t\}_{t \geq 0}$ es un proceso gaussiano con media cero.
- ii $\{B_t\}_{t \geq 0}$ es una martingala.
- iii $\{B_t^2 - t\}_{t \geq 0}$ es una martingala.

entonces, $\{B_t\}_{t \geq 0}$ es un movimiento browniano estándar.

Si desea revisar la demostración de esta proposición, puede encontrarla en [5]. Ahora se expondrá un resultado importante para realizar la simulación del movimiento browniano.

Definición. 42 Sean Z_1, \dots, Z_n variables aleatorias independientes e idénticamente distribuidas $N(0, \sigma^2)$, se define paseo aleatorio gaussiano como:

$$X_n = \sum_{i=1}^n Z_i$$

Proposición. 4 Sea $\{B_t\}_{t \geq 0}$ un movimiento browniano estándar, este es el límite del paseo aleatorio gaussiano X_n .

Ahora se va a definir un proceso muy importante para generalizar el movimiento browniano clásico, esto se llama puente browniano fraccionario. Se utilizará el trabajo sobre procesos con autosimilitud de Embrechts [10]

Definición. 43 Un proceso estocástico $\{X(t), t \geq 0\}$ se dice tener autosimilitud si, para cualquier $a > 0$, existe $b > 0$ tal que

$$\{X(at)\} \stackrel{d}{=} \{bX(t)\}.$$

Decimos que $\{X(t), t \geq 0\}$ es estocásticamente continuo en t si, para cualquier $\epsilon > 0$,

$$\lim_{h \rightarrow 0} P\{|X(t+h) - X(t)| > \epsilon\} = 0.$$

También decimos que $\{X(t), t \geq 0\}$ es trivial si $L(X(t))$ es una medida delta para cada $t > 0$.

Embrechts [10] resalta que el movimiento browniano es un ejemplo muy importante de un proceso estocástico. Es un proceso gaussiano, un proceso de difusión, un proceso de Lévy, un proceso de Markov, un martingala y un proceso con autosimilitud. Cada propiedad mencionada anteriormente del movimiento browniano fue el punto de partida de un nuevo campo en la teoría de procesos estocásticos. Actualmente, los procesos gaussianos, procesos de difusión, procesos de Lévy, procesos de Markov y martingalas constituyen áreas importantes de investigación en la teoría moderna de procesos estocásticos.

La noción de autosimilitud no alcanzó inmediatamente el mismo nivel fundamental; sin embargo, muchas aplicaciones más recientes han exigido una comprensión más profunda. Los procesos con autosimilitud son procesos estocásticos que son invariantes en distribución bajo una adecuada escala de tiempo y espacio. Estos procesos también se utilizan naturalmente en el análisis de fenómenos aleatorios (en el tiempo) que exhiben ciertas formas de dependencia a largo plazo.

El movimiento browniano fraccionario, que es un proceso gaussiano con autosimilitud con incrementos estacionarios, fue discutido por primera vez por Kolmogorov [17]. El primer artículo que proporciona un tratamiento probabilístico riguroso de procesos con autosimilitud generales es de Lamperti [20]. Más tarde, el estudio de procesos con autosimilitud no gaussianos con incrementos estacionarios fue iniciado por Taqqu [33], quien desarrolló aún más un teorema límite no gaussiano de Rosenblatt.

La noción de fractal se caracteriza por la formalización de la con autosimilitud, donde al escalar la estructura, sigue exhibiendo el mismo comportamiento, incluso cuando el escalamiento es muy pequeño. En el artículo de Embrechts [10] también se halla la estructura de covarianza de un proceso con autosimilitud con varianza finita.

Proposición. 5 Dado $\{X(t)\}$ no trivial y con autosimilitud con parámetro H , y $E[|X(1)|^2] < \infty$. Entonces,

$$E[X(t)X(s)] = \frac{1}{2}\{t^{2H} + s^{2H} - |t - s|^{2H}\}E[|X(1)|^2].$$

De esta construcción, se define el movimiento browniano fraccionario

Definición. 44 El Movimiento browniano Fraccionario con parámetro de Hurst $H \in (0, 1)$ es un proceso estocástico gaussiano $B_H = \{B_H(t), t \geq 0\}$ con media cero, incrementos estacionarios y función de covarianza dada por $\mathbb{E}[B_H(t)B_H(s)] = \frac{1}{2}[|t|^{2H} + |s|^{2H} - |t - s|^{2H}]$ para todo $s, t \geq 0$.

El parámetro de Hurst puede tomar valores en el rango de 0 a 1, donde un valor de 0.5 indica un comportamiento aleatorio similar al movimiento browniano clásico. Un valor mayor a 0.5 sugiere una tendencia a la persistencia, donde las observaciones tienden

a persistir en el tiempo, mientras que un valor menor a 0.5 indica una tendencia a la antipersistencia, donde las tendencias observadas tienden a revertirse.

Este concepto es relevante en diversas áreas, como hidrología, climatología, finanzas y biología. En finanzas [30], por ejemplo, el análisis del parámetro de Hurst puede ofrecer información valiosa sobre la naturaleza de las tendencias en los precios de los activos financieros, ayudando a los analistas a comprender mejor la dinámica subyacente.

Además, en campos como la climatología y la hidrología, el análisis de series temporales climáticas, el parámetro de Hurst proporciona información valiosa sobre patrones meteorológicos y la variabilidad a largo plazo. Este enfoque es esencial para la planificación de recursos hídricos, la gestión de eventos extremos como sequías o inundaciones, y la comprensión de la dinámica climática en diferentes regiones. En biología, el parámetro de Hurst también se aplica para analizar series temporales biológicas, como la expresión genética, lo que ayuda a identificar patrones persistentes en la actividad celular y la dinámica de poblaciones.

En el ámbito industrial, el parámetro de Hurst se utiliza en el control de procesos para prever fallos, optimizar operaciones y mejorar la eficiencia en la producción. Además, en el diseño y gestión de redes de comunicación, el análisis de la persistencia a largo plazo a través de este parámetro es crucial para garantizar una transmisión de datos eficiente y confiable. En resumen, el parámetro de Hurst y este proceso son herramientas versátiles que desempeñan un papel fundamental en el análisis y comprensión de fenómenos con memoria a largo plazo en diversos campos de estudio.

2.7 Métodos de simulación

Un método clásico para simular una ecuación diferencial ordinaria es el método de Euler. Sin embargo, una generalización para simular una ecuación diferencial estocástica clásica es el Método de Euler - Maruyama.

Según [41], el método de Euler-Maruyama se basa en el siguiente esquema iterativo:

$$X_{n+1} = X_n + f(X_n, t_n)\Delta t + g(X_n, t_n)\Delta B_n,$$

donde X_n es la aproximación de la solución en el tiempo t_n , $f(X_n, t_n)$ es el término de deriva, $g(X_n, t_n)$ es el término de difusión, $\Delta t = t_{n+1} - t_n$ es el paso de tiempo, y ΔB_n es una variable aleatoria que representa el incremento de un proceso de Wiener durante el paso de tiempo. El movimiento browniano es un proceso estocástico de tiempo continuo con incrementos independientes y distribuidos normalmente.

El método de Euler-Maruyama tiene una tasa de convergencia fuerte de orden $1/2$ cuando se aplica a EDEs con coeficientes continuos de Lipschitz. Sin embargo, cuando los coeficientes de las EDEs dependen de un proceso de Markov, la tasa de convergencia puede verse afectada. En particular, la tasa de convergencia puede disminuir cuando el proceso de Markov tiene un gran número de estados o cuando las tasas de transición entre los estados son pequeñas.

Para abordar estos problemas de convergencia, se han propuesto modificaciones al método de Euler-Maruyama. Una de estas modificaciones consiste en utilizar una aproximación más precisa del término de difusión, como el método de Milstein. Otra modificación consiste en utilizar un paso de tiempo más pequeño cuando el proceso de Markov se encuentra en ciertos estados, como los estados lentos.

En general, el método de Euler-Maruyama es una herramienta útil para simular EDEs con conmutación markoviana, pero se debe tener cuidado para asegurarse de que el método se utilice correctamente y se comprendan las propiedades de convergencia.

Existen varias formas para simular el movimiento browniano, se puede hacer uso de la caminata aleatoria simétrica, la caminata aleatoria normal, caminata aleatoria trinomial. Si desea profundizar puede ver el siguiente artículo [2]

1. Caminata aleatoria trinomial

La simulación del puente browniano mediante una caminata trinomial se puede realizar siguiendo los siguientes pasos:

- i. Define los parámetros necesarios: el número de pasos en el puente (número de incrementos), el tamaño de cada paso y el parámetro de volatilidad.
- ii. Inicializa el puente browniano con valores iniciales, por ejemplo, $B_0 = 0$ y $B_n = 0$, donde n es el número de pasos.
- iii. Para cada paso i desde 1 hasta $n - 1$, realiza los siguientes pasos:

Genera tres números aleatorios siguiendo una distribución trinomial, que representarán los movimientos hacia arriba, hacia abajo y sin cambio. Calcula el incremento del puente como la diferencia entre los movimientos hacia arriba y hacia abajo, multiplicado por el tamaño del paso y el parámetro de volatilidad. Actualiza el valor del puente sumando el incremento al valor anterior. Almacena el valor del puente en el vector de valores del puente.

- iv. Al finalizar el bucle, tendrá un vector de valores que representa el puente browniano simulado mediante una caminata trinomial.

2. Caminata aleatoria normal

La simulación del puente browniano mediante una caminata aleatoria normal se puede realizar siguiendo los siguientes pasos:

- i. Define los parámetros necesarios: el número de pasos en el puente (número de incrementos), el tamaño de cada paso y el parámetro de volatilidad.
- ii. Inicializa el puente browniano con valores iniciales, por ejemplo, $B_0 = 0$ y $B_n = 0$, donde n es el número de pasos.
- iii. Para cada paso i desde 1 hasta $n - 1$, realiza los siguientes pasos:
 Genera un número aleatorio Z_i siguiendo una distribución normal estándar. Calcula el incremento del puente como el producto del número aleatorio por el tamaño del paso y el parámetro de volatilidad. Actualiza el valor del puente sumando el incremento al valor anterior. Almacena el valor del puente en el vector de valores del puente.
- iv. Al finalizar el bucle, se tendrá un vector de valores que representa el puente browniano simulado mediante una caminata aleatoria normal.

El método que se sugiere es el gaussiano ya que converge más rápido al movimiento browniano. Esto al tener incrementos normales.

Para simular el movimiento browniano fraccionario se expone el siguiente método.

Carlson (2020) [3] propone el uso de la matriz raíz para simular procesos de difusión fraccionaria. En [1], puede encontrar el método de Cholesky para simular un movimiento browniano fraccionario. Para aplicar este método, se debe seguir el siguiente proceso.

Suponga que se quiere simular un movimiento browniano fraccionario $B_H(t)$ en un conjunto de tiempos t_1, t_2, \dots, t_n . La matriz de covarianza de los incrementos del proceso puede expresarse como:

$$C_{i,j} = \frac{1}{2} [(t_i^{2H} + t_j^{2H}) - |t_i - t_j|^{2H}]$$

donde H es el índice de Hurst.

La descomposición de Cholesky de esta matriz se puede expresar como:

$$C = LL^T,$$

donde L es una matriz triangular inferior con entradas diagonales positivas. Se puede utilizar esta descomposición para generar un conjunto de variables aleatorias normales correlacionadas Z_1, Z_2, \dots, Z_n utilizando la transformación $LZ = Y$, donde Y es un vector de variables aleatorias normales no correlacionadas.

El movimiento browniano fraccionario simulado se puede obtener tomando la suma acumu-

lativa de las variables aleatorias generadas multiplicadas por los factores de escala apropiados. Específicamente, tenemos:

$$B_H(t_i) = \sum_{j=1}^i a_j Y_j,$$

donde a_j es el factor de escala dado por:

$$a_j = \sqrt{\frac{t_j^{2H} - t_{j-1}^{2H}}{2}}.$$

Repitiendo este proceso de simulación para diferentes valores del índice de Hurst H , se puede generar trayectorias de muestra del movimiento browniano fraccionario.

Este método será fundamental para simular soluciones de una ecuación diferencial estocástica sin el movimiento browniano clásico, pero con un ruido browniano fraccionario. Se puede utilizar el cálculo de Malliavin, pero en este trabajo final, se considera el método de simulación para estimar una trayectoria de la solución. Si desea explorar más sobre el cálculo de Malliavin, puede encontrar más información puede ver [24].

El uso de procesos con autosimilitud, como el movimiento browniano fraccionario, se presenta como una alternativa valiosa en situaciones donde los procesos exhiben memoria a largo plazo o no cumplen el supuesto de incrementos independientes. En contraste con los procesos estocásticos tradicionales, los cuales asumen la independencia de los incrementos en el tiempo, los procesos con autosimilitud permiten modelar fenómenos con estructuras de correlación más complejas y dependencias a largo plazo.

En el caso específico del movimiento browniano fraccionario, su capacidad para capturar la con autosimilitud y la dependencia a largo plazo lo hace especialmente útil en la modelización de sistemas que exhiben comportamientos persistentes en el tiempo. La propiedad de con autosimilitud implica que el proceso tiene la misma estructura estadística a diferentes escalas temporales, lo cual es consistente con la presencia de memoria a largo plazo.

Así, al emplear el movimiento browniano fraccionario u otros procesos con autosimilitud, se ofrece una herramienta flexible y poderosa para abordar fenómenos estocásticos con complejas interdependencias temporales, proporcionando modelos más realistas y adaptados a las características específicas de los datos observados.

Cabe destacar que la simulación de puentes estocásticos depende tanto de la simulación de movimientos brownianos clásicos como fraccionarios, ya que los puentes se definen a partir de estos movimientos, tanto en la ecuación diferencial estocástica como en la transformación directa del movimiento browniano.

3 Inferencia y bondad de ajuste sobre puentes estocásticos

3.1 Introducción

El proceso estocástico conocido como Puente browniano, detallando muchas de sus propiedades fundamentales. Este proceso se deriva del movimiento browniano al restringir el parámetro temporal a un intervalo finito y limitar el tiempo final, esto es introducido por Chow y Jones [6]. El Puente browniano, se revela como una herramienta valiosa en el análisis estadístico y modelado matemático de eventos del mundo real. La exposición de las propiedades del Puente browniano por parte se destaca la interacción de diversos conceptos matemáticos y estadísticos en su derivación, evidenciando la convergencia de varias ramas de las matemáticas y la estadística. Además, ilustran aplicaciones prácticas del Puente browniano, mostrando su relevancia en la derivación de pruebas estadísticas no paramétricas esenciales y herramientas de estimación de intervalos. Un ejemplo adicional presentado en su trabajo destaca la aplicación del Puente browniano en una ecuación diferencial estocástica, demostrando su utilidad en el modelado matemático y simulación de fenómenos complejos.

Una aproximación analítica en la inferencia sobre los puentes estocásticos se encuentra en Gorgens [13]. Para ello, utiliza el teorema de Karhunen-Loève, que es la generalización de la descomposición espectral en espacios estocásticos. Gorgens aborda un problema de decisión simple sobre el parámetro de escala en el Puente browniano α , denotado como $X(\alpha)$ en el intervalo $[0, 1]$.

Dados dos valores $\alpha_0, \alpha_1 \geq 0$ con $\alpha_0 + \alpha_1 \geq 1$ y algún tiempo $0 \leq T \leq 1$, el objetivo es contrastar $H_0 : \alpha = \alpha_0$ frente a $H_1 : \alpha = \alpha_1$ basándose en la observación de $X(\alpha)$ hasta el tiempo T . La razón de verosimilitud se expresa como un funcional de una forma cuadrática $\psi(X(\alpha))$ de $X(\alpha)$. Para calcular la distribución de $\psi(X(\alpha))$ bajo la hipótesis nula, Gorgens generaliza el Teorema de Karhunen-Loève a medidas finitas positivas en $[0, 1]$ y calcula la expansión de Karhunen-Loève de $X(\alpha)$ bajo dicha medida. Basándose en esta expansión, la distribución de $\psi(X(\alpha))$ se obtiene mediante la fórmula de Smirnov.

Se propone una metodología para llevar a cabo inferencia clásica sobre los parámetros tanto del puente browniano clásico como del puente browniano fraccionario. Este enfoque se basa en las propiedades de la distribución normal multivariada, el teorema central del límite multivariado, y la raíz de una matriz definida positiva, entre otros. Además, se presenta una propuesta para evaluar la bondad de ajuste. Sin embargo, como mejora adicional, se sugiere la implementación de técnicas de envolvente propuestas por Cook [7], las cuales resuelven problemas asociados a la normal multivariada bajo la propiedad de Bonferroni, especialmente en entornos de alta dimensionalidad.

Por último, se aborda el desafío de encontrar un estimador para el coeficiente fraccionario. En este trabajo, se propone realizar una exploración en una grilla en el intervalo $(0,1)$ y seleccionar el modelo que mejor se ajuste a los datos simulados. No obstante, como trabajo futuro, se plantea la necesidad de desarrollar métodos inferenciales analíticos para abordar este problema de manera más precisa y eficiente.

Algunas propiedades del puente se pueden encontrar en [22], y estas son:

Si $B(t)$ es un proceso Wiener estándar entonces

$$X(t) = B(t) - \frac{t}{T}B(T)$$

es un puente browniano para $t \in [0, T]$.

Recíprocamente, si $X(t)$ es un puente browniano y Z es una variable aleatoria normal estándar independiente de X , entonces el proceso

$$B(t) = X(t) + tZ$$

es un movimiento browniano para $t \in [0, 1]$. Más generalmente, un proceso Wiener $B(t)$ para $t \in [0, T]$ puede descomponerse en

$$B(t) = \sqrt{T}X\left(\frac{t}{T}\right) + \frac{t}{\sqrt{T}}Z.$$

La covarianza entre $B(s)$ y $B(t)$ está dada por

$$\text{Cov}(B(s), B(t)) = \min(s, t) - \frac{st}{T},$$

si $s < t$.

Otra representación del puente browniano basada en el movimiento browniano es, para $t \in [0, T]$:

$$X(t) = \frac{T-t}{\sqrt{T}}B\left(\frac{t}{T-t}\right).$$

Recíprocamente, para $t \in [0, \infty)$:

$$B(t) = \frac{T+t}{T} X\left(\frac{Tt}{T+t}\right).$$

El puente browniano también puede representarse como una serie de Fourier con coeficientes estocásticos, como

$$X_t = \sum_{k=1}^{\infty} Z_k \frac{\sqrt{2T} \sin(k\pi t/T)}{k\pi},$$

donde Z_1, Z_2, \dots son variables aleatorias normales estándar independientes e idénticamente distribuidas. Esta descomposición se debe a la generalización estocástica del teorema de descomposición espectral, dicho teorema se llama el teorema de Karhunen-Loève.

Proposición. 6 Sea $X(t)$ un proceso estocástico centrado en $t \in [a, b]$ y con función de covarianza definida como

$$R_X(t, s) = \text{Cov}(X(t), X(s)).$$

Si $X(t)$ es cuadrado integrable en el sentido de Lebesgue en $[a, b]$ y si existe una secuencia de funciones ortogonales $\{\phi_n(t)\}$ y una secuencia de variables aleatorias independientes Z_n .

Las funciones $\{\phi_n(t)\}$ forman un conjunto ortonormal completo en el espacio de Hilbert $L^2([a, b])$ de funciones cuadrado integrables, es decir,

$$\int_a^b \phi_m(t) \phi_n(t) dt = \delta_{mn},$$

donde δ_{mn} es la delta de Kronecker, entonces, $X(t)$ puede representarse como una serie convergente en media cuadrática,

$$X(t) = \sum_{n=1}^{\infty} Z_n \phi_n(t),$$

donde la convergencia es en el sentido de la norma L^2 y las variables aleatorias Z_n son no correlacionadas ($\text{Cov}(Z_m, Z_n) = 0$ para $m \neq n$).

Entonces, el proceso $X(t)$ puede expresarse como una serie de Karhunen-Loève.

3.2 Puente browniano clásico

Para definir el puente browniano clásico existen muchas caracterizaciones, en este trabajo de grado se considerará la siguiente:

Definición. 45 (Puente browniano estándar) : Sea $\{B_t\}_{t \leq 0}$ un movimiento browniano estándar, se define $\{X_t = B_t - tB_1 : t \in [0, 1]\}$ como un puente browniano estándar entre $[0, 1]$.

Se puede verificar que el puente browniano satisface la siguiente ecuación diferencial estocástica

$$dX_t := \frac{X_t}{1-t} dt + dB_t, t \in [0, 1].$$

Si $\{X_t\}_{t \in [0,1]}$ es un Puente browniano estándar, entonces, $\{Y_t : \sigma X_t\}_{t \in [0,1]}$ se le conoce como puente browniano de parámetro σ

Nota. En aplicaciones de la vida real el modelo considera un punto inicial arbitrario y un punto final arbitrario, sin embargo, se puede usar el hecho de que cualquier segmento de recta es homeomorfo al intervalo $[0, 1]$, por lo tanto, se puede transformar el soporte de los datos a un subconjunto del intervalo $[0, 1]$

3.3 Inferencia y bondad de ajuste sobre el puente browniano clásico

Nuestro aporte consistirá en considerar el método estadístico clásico, ya que las integrales utilizadas en la metodología de Karhunen-Loève, en la práctica, no son conocidas. En este caso, se contempla la realización parcial de un puente browniano con parámetro σ . La falta de conocimiento de σ indica la necesidad de buscar un método para estimarlo. Para ello, consideramos lo siguiente: Sea $Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}$ una muestra aleatoria parcial de $\{Y_t\}$, en realidad, se tiene $y_{t_1}, y_{t_2}, \dots, y_{t_n}$ como la realización de la muestra. El problema que se discutirá y resolverá es el de estimar σ . Cuando se desea obtener la función de verosimilitud de manera directa, se requiere el conocimiento de W_1 (Punto final del movimiento browniano que genera al puente browniano); sin embargo, esto se puede superar aplicando la siguiente proposición.

Proposición. 7 Sea $y_0 = 0, Y_{t_1}, Y_{t_2}, \dots, Y_{t_{n-1}}, y_{t_n} = 0$ una muestra aleatoria parcial de un puente browniano de parámetro σ . Se define

$$\mathbf{U} = \begin{bmatrix} U_{t_2} \\ U_{t_3} \\ \vdots \\ U_{t_{n-1}} \end{bmatrix} = \begin{bmatrix} Y_{t_2} - \frac{t_2}{t_1} Y_{t_1} \\ Y_{t_3} - \frac{t_3}{t_2} Y_{t_2} \\ \vdots \\ Y_{t_{n-1}} - \frac{t_{n-1}}{t_{n-2}} Y_{t_{n-2}} \end{bmatrix}$$

Con \mathbf{U} un vector aleatorio centrado normal $n-2$ variante que no depende de W_1 .

Demostración. Note que $U_{t_{i+1}} = Y_{t_{i+1}} - \frac{t_{i+1}}{t_i} Y_{t_i} = \sigma W_{t_{i+1}} - t_{i+1} \sigma W_1 - \frac{t_{i+1}}{t_i} \sigma W_{t_i} +$

$t_{i+1}\sigma W_1 = \sigma(W_{t_{i+1}} - \frac{t_{i+1}}{t_i}W_{t_i})$, entonces,

$$\mathbf{U} = \begin{bmatrix} Y_{t_2} - \frac{t_2}{t_1}Y_{t_1} \\ Y_{t_3} - \frac{t_3}{t_2}Y_{t_2} \\ \vdots \\ Y_{t_{n-1}} - \frac{t_{n-1}}{t_{n-2}}Y_{t_{n-2}} \end{bmatrix} = \sigma \begin{bmatrix} W_{t_2} - \frac{t_2}{t_1}W_{t_1} \\ W_{t_3} - \frac{t_3}{t_2}W_{t_2} \\ \vdots \\ W_{t_{n-1}} - \frac{t_{n-1}}{t_{n-2}}W_{t_{n-2}} \end{bmatrix}$$

Note que el vector \mathbf{U} se pudo escribir independiente del valor de W_1

$$= \sigma \begin{bmatrix} -\frac{t_2}{t_1} & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\frac{t_3}{t_2} & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\frac{t_{n-1}}{t_{n-2}} & 1 \end{bmatrix}_{(n-2) \times (n-1)} \begin{bmatrix} W_{t_1} \\ W_{t_2} \\ \vdots \\ W_{t_{n-2}} \\ W_{t_{n-1}} \end{bmatrix}_{(n-1) \times 1} = \sigma \mathbf{R}\mathbf{W}$$

Como el proceso de Wiener es gaussiano y la covarianza es el tiempo mínimo entre las dos variables, entonces,

$$\mathbf{W} = \begin{bmatrix} W_{t_1} \\ W_{t_2} \\ \vdots \\ W_{t_{n-2}} \\ W_{t_{n-1}} \end{bmatrix} \sim N_{n-1} \left(\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} t_1 & t_1 & \cdots & t_1 & t_1 \\ t_1 & t_2 & \cdots & t_2 & t_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ t_1 & t_2 & \cdots & t_{n-2} & t_{n-2} \\ t_1 & t_2 & \cdots & t_{n-2} & t_{n-1} \end{bmatrix} \right) \stackrel{D}{=} N_{n-1}(\mathbf{0}, \mathbf{S})$$

Como $\mathbf{U} = \sigma \mathbf{R}\mathbf{W}$, entonces, por propiedades de las distribución normal multivariada

$$\mathbf{U} = \sigma \mathbf{R}\mathbf{W} \sim N_{n-2}(\mathbf{0}, \sigma^2 \mathbf{R}\mathbf{S}\mathbf{R}')$$

por el teorema de la estandarización multivariada

$$\mathbf{Z} = \frac{(\mathbf{R}\mathbf{S}\mathbf{R}')^{-1/2}}{\sigma} \mathbf{U} \sim N_{n-2}(\mathbf{0}, \mathbb{I}) \quad \square$$

Note que $\mathbf{U} = \mathbf{R}\mathbf{Y}$, con $\mathbf{Y} = (Y_i)_i, i = 1, 2, \dots, n-1$, entonces

$$\mathbf{Z} = \frac{(\mathbf{R}\mathbf{S}\mathbf{R}')^{-1/2}}{\sigma} \mathbf{U} = \frac{(\mathbf{R}\mathbf{S}\mathbf{R}')^{-1/2}}{\sigma} \mathbf{R}\mathbf{Y} \sim N_{n-2}(\mathbf{0}, \mathbb{I})$$

En esta última expresión tenemos una forma lineal que depende de la muestra parcial del puente browniano, además, dicha forma lineal distribuye normal multivariada estándar. Ahora, se va a encontrar un estimador insesgado para σ^2 que solo dependa de la muestra aleatoria parcial del puente browniano. Para validar la bondad de ajuste del puente browniano sobre \mathbf{y} , se puede realizar la transformación $\mathbf{z} = \frac{(\mathbf{R}\mathbf{S}\mathbf{R}')^{-1/2}}{\sigma} \mathbf{R}\mathbf{y}$, consecuentemente, realizar una prueba de normalidad.

Proposición. 8 Sea $y_0 = 0, Y_{t_1}, Y_{t_2}, \dots, Y_{t_{n-1}}, y_{t_n} = 0$ una muestra aleatoria parcial de un puente browniano de parámetro σ . Se cumple que

$$\hat{\sigma}^2 := \frac{\mathbf{Y}'\mathbf{R}'(\mathbf{RSR}')^{-1}\mathbf{R}\mathbf{Y}}{n-2}$$

es un estimador insesgado para σ^2 .

Demostración.

Haciendo uso del teorema de formas cuadráticas, se tiene:

$$\mathbf{Z}'\mathbf{Z} = \frac{\mathbf{Y}'\mathbf{R}'(\mathbf{RSR}')^{-1/2}}{\sigma} \frac{(\mathbf{RSR}')^{-1/2}\mathbf{R}\mathbf{Y}}{\sigma} \sim \chi_{n-2}^2$$

$$\mathbf{Z}'\mathbf{Z} = \frac{\mathbf{Y}'\mathbf{R}'(\mathbf{RSR}')^{-1}\mathbf{R}\mathbf{Y}}{\sigma^2} \sim \chi_{n-2}^2$$

Entonces $E(\mathbf{Z}'\mathbf{Z}) = n - 2$, y por linealidad de la esperanza se cumple

$$E\left(\frac{\mathbf{Y}'\mathbf{R}'(\mathbf{RSR}')^{-1}\mathbf{R}\mathbf{Y}}{n-2}\right) = \sigma^2 \quad \square$$

Nota. Teniendo la estimación de la varianza, se puede evaluar la verosimilitud y tener criterios de evaluación mediante las medidas de información tales como:

$$AIC := -2\log\left(N_{n-2}(\mathbf{0}, \mathbb{I})\left[\frac{(\mathbf{RSR}')^{-1/2}}{\sigma}\mathbf{R}\mathbf{Y}\right]\right) + 2k$$

$$BIC := -2\log\left(N_{n-2}(\mathbf{0}, \mathbb{I})\left[\frac{(\mathbf{RSR}')^{-1/2}}{\sigma}\mathbf{R}\mathbf{Y}\right]\right) + k\log(n)$$

Con $N_p(\mathbf{0}, \mathbb{I})[\cdot]$ la distribución normal multivariada, con vector de medias cero y matriz de varianzas y covarianzas la identidad, evaluada en un vector dado.

Un problema al que podemos enfrentarnos es someter la hipótesis de igualdad de volatilidad en dos puentes brownianos independientes, para esto se puede considerar el siguiente test, este test se construye basados en el test clásico de la distribución F .

Sea $y_0^{(1)} = 0, Y_{t_1^{(1)}}, Y_{t_2^{(1)}}, \dots, Y_{t_{n-1}^{(1)}}, y_{t_n^{(1)}} = 0$ una muestra aleatoria parcial de un puente browniano entre $[0, 1]$ de parámetro σ_1 y sea $y_0^{(2)} = 0, Y_{t_1^{(2)}}, Y_{t_2^{(2)}}, \dots, Y_{t_{m-1}^{(2)}}, y_{t_m^{(2)}} = 0$ una muestra aleatoria parcial de un puente browniano entre $[0, 1]$ de parámetro σ_2 , se desea someter el siguiente contraste de hipótesis:

$$\begin{cases} H_0 : \sigma_1 = \sigma_2 \\ H_1 : \sigma_1 \neq \sigma_2 \end{cases}$$

Como

$$\frac{(n-2)\hat{\sigma}_1^2}{\sigma_1^2} \sim \chi_{n-2}^2 \quad y \quad \frac{(m-2)\hat{\sigma}_2^2}{\sigma_2^2} \sim \chi_{m-2}^2$$

y dichos puentes son independientes, entonces, el estadístico para someter el contraste es:

$$F \stackrel{H_0}{=} \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \sim F_{m-2}^{n-2}$$

Con esta estadística a dos colas se puede someter el contraste de interés.

3.4 Puente browniano fraccionario

El puente browniano clásico es la solución a una ecuación diferencial que asume como ruido un movimiento browniano, dicho browniano asume incrementos independientes, algo que no siempre se cumple en situaciones reales, sin embargo, el movimiento browniano fraccionario modela la dependencia de los incrementos, en esta sección se propondrán métodos inferenciales para la ecuación de puente con ruido fraccionario, si desea profundizar más sobre el puente fraccionario puede ver [11].

Definición. 46 (Puente browniano fraccionario) : Sea $\{B_t^H\}_{t \leq 0}$ un movimiento browniano fraccionario estándar con parámetro de Hurst $H \in [0, 1]$, se define $\{X_t = B_t^H - tB_1^H : t \in [0, 1]\}$ como un puente browniano fraccionario estándar entre $[0, 1]$ con parámetro de Hurst H .

Se puede verificar que el puente browniano fraccionario satisface la siguiente ecuación diferencial estocástica fraccionaria

$$dX_t := \frac{X_t}{1-t} dt + dB_t^H, t \in [0, 1]$$

Si $\{X_t\}_{t \in [0,1]}$ es un Puente browniano fraccionario estándar, entonces, $\{Y_t : \sigma X_t\}_{t \in [0,1]}$ se le conoce como puente browniano fraccionario de parámetro σ

Nota. Note que a diferencia del puente browniano clásico, se deben estimar dos parámetros H y σ . Al igual que el puente browniano clásico el modelo considera un punto inicial fijo y un punto final fijo, sin embargo, se puede usar el hecho de que cualquier segmento de recta es homeomorfo al intervalo $[0, 1]$, por lo tanto, se puede transformar el soporte de los datos a un subconjunto del intervalo $[0, 1]$

Si $\{Y_t^{(1)}\}_{t \in [0,1]}$ y $\{Y_t^{(2)}\}_{t \in [0,1]}$ son dos puentes brownianos fraccionarios independientes con coeficiente de Hurst H y parámetro de volatilidad σ_1^2 y σ_2^2 respectivamente. El proceso estocástico

$$Y_t = \alpha Y_t^{(1)} + \beta Y_t^{(2)}$$

Es un puente estocástico fraccionario con parámetro H y coeficiente de volatilidad $\alpha^2\sigma_1^2 + \beta^2\sigma_2^2$

La combinación lineal de puentes brownianos con diferentes parámetros de Hurst no siempre resulta en un puente browniano. Sin embargo, estos procesos exhiben propiedades interesantes. Para un análisis más detallado, se puede consultar el trabajo sobre movimientos brownianos fraccionarios mixtos [32].

3.5 Inferencia y bondad de ajuste sobre el puente browniano fraccionario

El método estadístico considera la realización parcial de un puente browniano fraccionario de parámetros H y σ , el desconocimiento de H y σ nos indica que se debe buscar un método para estimarlos, para ello se considera lo siguiente: Sea $Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}$ una muestra aleatoria parcial de $\{Y_t\}$, en realidad, se tiene $y_{t_1}, y_{t_2}, \dots, y_{t_n}$ la realización de la muestra. El problema que se discutirá es el de estimar H y σ . Cuando se quiere hallar la función de verosimilitud, de manera directa, se requiere el conocimiento de W_1^H (Punto final del movimiento browniano fraccionario que genera al puente browniano fraccionario), sin embargo, esto también se puede solucionar aplicando transformaciones hasta llegar a una muestra aleatoria gaussiana.

Recuerde que bajo el movimiento browniano fraccionario la covarianza es $\frac{1}{2} [|t|^{2H} + |s|^{2H} - |t-s|^{2H}]$. Considere la matriz de varianzas y covarianzas fraccionaria $\mathbf{C} := (C_{i,j})$ definida en [9]

Proposición. 9 Sea $y_0 = 0, Y_{t_1}, Y_{t_2}, \dots, Y_{t_{n-1}}, y_{t_n} = 0$ una muestra aleatoria parcial de un puente browniano fraccionario de parámetro σ , suponga que se conoce $H \in (0, 1)$. Se define

$$\mathbf{U} = \begin{bmatrix} U_{t_2} \\ U_{t_3} \\ \vdots \\ U_{t_{n-1}} \end{bmatrix} = \begin{bmatrix} Y_{t_2} - \frac{t_2}{t_1} Y_{t_1} \\ Y_{t_3} - \frac{t_3}{t_2} Y_{t_2} \\ \vdots \\ Y_{t_{n-1}} - \frac{t_{n-1}}{t_{n-2}} Y_{t_{n-2}} \end{bmatrix}$$

Con \mathbf{U} es un vector aleatorio centrado normal $n-2$ variante que no depende de W_1 .

Demostración.

Note que $U_{t_{i+1}} = Y_{t_{i+1}} - \frac{t_{i+1}}{t_i} Y_{t_i} = \sigma W_{t_{i+1}}^H - \frac{t_{i+1}}{t_i} \sigma W_{t_i}^H = \sigma (W_{t_{i+1}}^H - \frac{t_{i+1}}{t_i} W_{t_i}^H)$

$\frac{t_{i+1}}{t_i} W_{t_i}^H$), entonces,

$$\mathbf{U} = \begin{bmatrix} Y_{t_2} - \frac{t_2}{t_1} Y_{t_1} \\ Y_{t_3} - \frac{t_3}{t_2} Y_{t_2} \\ \vdots \\ Y_{t_{n-1}} - \frac{t_{n-1}}{t_{n-2}} Y_{t_{n-2}} \end{bmatrix} = \sigma \begin{bmatrix} W_{t_2}^H - \frac{t_2}{t_1} W_{t_1}^H \\ W_{t_3}^H - \frac{t_3}{t_2} W_{t_2}^H \\ \vdots \\ W_{t_{n-1}}^H - \frac{t_{n-1}}{t_{n-2}} W_{t_{n-2}}^H \end{bmatrix}$$

Note que el vector \mathbf{U} se pudo escribir independiente del valor de W_1^H

$$= \sigma \begin{bmatrix} -\frac{t_2}{t_1} & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\frac{t_3}{t_2} & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\frac{t_{n-1}}{t_{n-2}} & 1 \end{bmatrix}_{(n-2) \times (n-1)} \begin{bmatrix} W_{t_1}^H \\ W_{t_2}^H \\ \vdots \\ W_{t_{n-2}}^H \\ W_{t_{n-1}}^H \end{bmatrix}_{(n-1) \times 1} = \sigma \mathbf{R} \mathbf{W}^H$$

Ahora, como $\mathbf{W}^H \sim N_{n-1}(\mathbf{0}, \mathbf{C})$, se hace la descomposición de Cholesky [9] para hallar la raíz de la matriz \mathbf{C} , entonces, $\mathbf{C}^{-\frac{1}{2}} \mathbf{W}^H \sim N_{n-1}(\mathbf{0}, \mathbf{I})$, por otro lado haciendo uso de los teoremas de formas lineales bajo la distribución normal multivariada, se tiene que $\sigma \mathbf{R} \mathbf{W}^H \sim N_{n-2}(\mathbf{0}, \sigma^2 \mathbf{R} \mathbf{C} \mathbf{R}')$, por lo tanto \mathbf{U} es un vector centrado y gaussiano \square

También sobre el vector \mathbf{U} se puede aplicar el método de estandarización multivariada

$$\begin{aligned} \mathbf{Z} &= (\mathbf{R} \mathbf{C} \mathbf{R}')^{-\frac{1}{2}} \mathbf{R} \mathbf{W}^H \sim N_{n-2}(\mathbf{0}, \mathbf{I}) \\ &= \frac{(\mathbf{R} \mathbf{C} \mathbf{R}')^{-\frac{1}{2}} \mathbf{R} \mathbf{Y}}{\sigma} \sim N_{n-2}(\mathbf{0}, \mathbf{I}) \end{aligned}$$

Note que cuando $H = \frac{1}{2}$ se cumple que $\mathbf{S} = \mathbf{C}$, esto hace que se concluya que el método de estimación sobre un puente browniano fraccionario es la generalización de la estimación del puente browniano clásico

Proposición. 10 Sea $y_0 = 0, Y_{t_1}, Y_{t_2}, \dots, Y_{t_{n-1}}, y_{t_n} = 0$ una muestra aleatoria parcial de un puente browniano fraccionario de parámetros H conocido y σ desconocido. Se cumple que

$$\hat{\sigma}^2 := \frac{\mathbf{Y}' \mathbf{R}' (\mathbf{R} \mathbf{C} \mathbf{R}')^{-1} \mathbf{R} \mathbf{Y}}{n-2}$$

es un estimador insesgado para σ^2 .

Demostración.

Haciendo uso del teorema de formas cuadráticas, se tiene:

$$\mathbf{Z}' \mathbf{Z} = \frac{\mathbf{Y}' \mathbf{R}' (\mathbf{R} \mathbf{C} \mathbf{R}')^{-1/2}}{\sigma} \frac{(\mathbf{R} \mathbf{C} \mathbf{R}')^{-1/2} \mathbf{R} \mathbf{Y}}{\sigma} \sim \chi_{n-2}^2$$

$$\mathbf{Z}'\mathbf{Z} = \frac{\mathbf{Y}'\mathbf{R}'(\mathbf{RCR}')^{-1}\mathbf{RY}}{\sigma^2} \sim \chi_{n-2}^2$$

Entonces $E(\mathbf{Z}'\mathbf{Z}) = n - 2$, y por linealidad de la esperanza se cumple

$$E\left(\frac{\mathbf{Y}'\mathbf{R}'(\mathbf{RCR}')^{-1}\mathbf{RY}}{n-2}\right) = \sigma^2 \quad \square$$

Nota. Teniendo la estimación de la varianza, se puede evaluar la verosimilitud y tener criterios de evaluación mediante las medidas de información tales como:

$$AIC := -2\log\left(N_{n-2}(\mathbf{0}, \mathbb{I})\left[\frac{(\mathbf{RCR}')^{-1/2}}{\sigma}\mathbf{RY}\right]\right) + 2 * k$$

$$BIC := -2\log\left(N_{n-2}(\mathbf{0}, \mathbb{I})\left[\frac{(\mathbf{RCR}')^{-1/2}}{\sigma}\mathbf{RY}\right]\right) + k\log(n)$$

Con $N_p(\mathbf{0}, \mathbb{I})[\cdot]$ la distribución normal multivariada, con vector de medias cero y matriz de varianzas y covarianzas la identidad, evaluada en un vector dado.

Un problema al que podemos enfrentarnos es someter la hipótesis de igualdad de volatilidad en dos puentes brownianos fraccionarios independientes, para esto se puede considerar el siguiente test, este test se construye basados en el test clásico de la distribución F , también se cumple que el test de igualdad de varianzas bajo la distribución F es el test uniformemente más potente.

Sea $y_0^{(1)} = 0, Y_{t_1^{(1)}}, Y_{t_2^{(1)}}, \dots, Y_{t_{n-1}^{(1)}}, y_{t_n^{(1)}} = 0$ una muestra aleatoria parcial de un puente browniano fraccionario con H conocido entre $[0, 1]$ de parámetro σ_1 y sea $y_0^{(2)} = 0, Y_{t_1^{(2)}}, Y_{t_2^{(2)}}, \dots, Y_{t_{m-1}^{(2)}}, y_{t_m^{(2)}} = 0$ una muestra aleatoria parcial de un puente browniano fraccionario con el mismo H del primer puente y de parámetro σ_2 , se desea someter el siguiente contraste de hipótesis:

$$\begin{cases} H_0 : \sigma_1 = \sigma_2 \\ H_1 : \sigma_1 \neq \sigma_2 \end{cases}$$

Como

$$\frac{(n-2)\hat{\sigma}_1^2}{\sigma_1^2} \sim \chi_{n-2}^2 \quad y \quad \frac{(m-2)\hat{\sigma}_2^2}{\sigma_2^2} \sim \chi_{m-2}^2$$

y dichos puentes son independientes, entonces, el estadístico para someter el contraste es:

$$F \stackrel{H_0}{=} \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \sim F_{m-2}^{n-2}$$

Con esta estadística a dos colas se puede someter el contraste de interés.

3.6 Estimación del parámetro H

Para estimar H , existen varios enfoques. En primer lugar, se puede abordar una estimación analítica según se describe en [40]. Sin embargo, esta estimación implica la consideración de la integral estocástica, lo cual presenta un desafío, ya que en la vida real solo se dispone de una realización de dicha integral. Otra alternativa es emplear el método de máxima verosimilitud. En este caso, se puede realizar una partición del intervalo $[0, 1]$ y determinar el valor H^* que maximiza la verosimilitud. Este enfoque ofrece una aproximación práctica y eficiente para estimar el parámetro fractal H a partir de los datos disponibles.

Para ilustrar este método, se considera $H = 0.8$. Se genera una grilla de 20 valores y se identifica el valor de H que maximiza la verosimilitud. La Figura 4-6 muestra los resultados obtenidos.

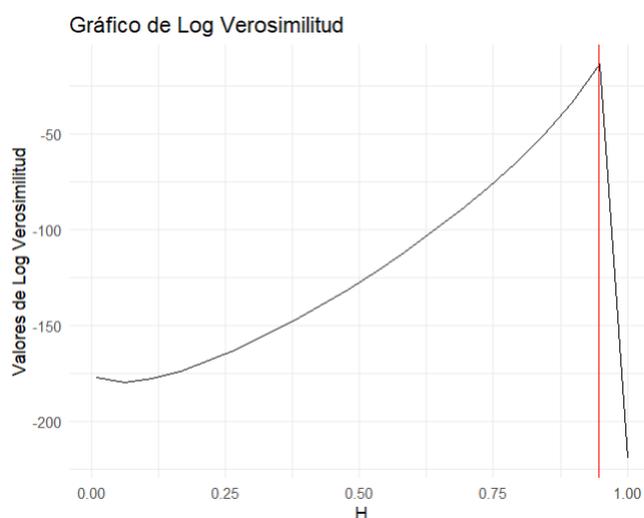


Figura 3-1: Resultados de la maximización de la verosimilitud.

Al obtener un valor estimado de $\hat{H} = 0.94$, se propone para futuras investigaciones evaluar la calidad de la convergencia y las propiedades de esta estimación.

Aunque no será considerado en el contexto de este trabajo, se propone como una dirección para investigaciones futuras el uso de la estimación bayesiana para H . Esta aproximación implica tomar una distribución previa triangular $\Delta(0, 1)$ o $beta(\alpha, \beta)$, ya que estas distribuciones tienen un soporte en el intervalo $(0,1)$. Explorar la estimación bayesiana podría brindar una perspectiva adicional y potencialmente mejorar la precisión de las estimaciones de H en entornos estocásticos.

Existe una amplia variedad de distribuciones cuyo soporte se encuentra en el intervalo $(0, 1)$; sin embargo, no siempre las distribuciones posteriores del parámetro serán conocidas por su núcleo, por lo que se sugiere para futuras investigaciones el empleo de métodos MCMC como Metropolis-Hastings, que no requiere de marginalización conjugada.

4 Simulación y aplicaciones

4.1 Datos funcionales con ruidos basados en puentes

Uno de los desafíos significativos en el análisis de datos funcionales reside en la generación de curvas de confianza. Varios autores han abordado este tema; por ejemplo, Goldsmith [12] detalla la construcción de regiones funcionales de confianza utilizando el Análisis en Componentes Principales Funcional (ACPF). Sin embargo, es importante destacar que estas curvas de confianza no tienen en cuenta la restricción del punto inicial y final fijo.

En el contexto de sensores utilizados para capturar trayectorias espinales, tales como la flexión lateral derecha, la flexión lateral izquierda, la flexión frontal, entre otras. Para profundizar en las metodologías ópticas de estos movimientos puede ver [42] para un análisis más profundo sobre los movimientos espinales. En estos movimientos se conserva la fijación del punto inicial y final. Esto implica que la variabilidad aumenta considerablemente lejos de los puntos inicial y final. En este trabajo final se presentará tanto una metodología teórica como una metodología empírica para estimar bandas de confianza de estos datos funcionales. Dichas metodologías abordarán la consideración del punto inicial y final fijo, un aspecto crucial en el análisis de este tipo de datos.

Proposición. 11 Sea $\chi_1(t), \chi_2(t) \dots, \chi_n(t)$ una muestra aleatoria funcional de $\chi(t) := \mu(t) + B(t)$ con $\{B(t)\}_{t \geq 0}$ un puente browniano clásico con parámetro de volatilidad desconocido, se cumple que

$$\hat{\mu}(t) := \frac{1}{n} \sum_{i=1}^n \chi_i(t)$$

es un estimador funcional insesgado para $\mu(t)$.

Demostración.

Considere la esperanza del estimador funcional

$$E(\hat{\mu}(t)) = E\left(\frac{1}{n} \sum_{i=1}^n \chi_i(t)\right) = \frac{1}{n} \sum_{i=1}^n E(\chi_i(t)) = \frac{1}{n} \sum_{i=1}^n \mu(t) = \mu(t)$$

□

Para garantizar la validez estadística de la inferencia sobre la función media, se propone la siguiente metodología:

- i. Se realiza una verificación del ajuste del modelo sobre cada una de las realizaciones parciales de los puentes, asegurando que la rotación de datos, de acuerdo con la proposición 10, siga el supuesto de normalidad. Luego, se determina la proporción de puentes que cumplen con este supuesto, lo cual actúa como un indicador de la bondad de ajuste del modelo.
- ii. Otro supuesto relevante del modelo es que las realizaciones del puente presentan un parámetro de volatilidad constante. Con este propósito, se sugiere la realización de pruebas estadísticas como el test de Bartlett, la razón de verosimilitud, Hartley, Levene, entre otros, para evaluar la homogeneidad en la variabilidad de las diferentes rotaciones de los puentes.
- iii. Para determinar si un modelo fraccionario se ajusta mejor, se puede emplear criterios de información específicos presentados en la sección anterior.

Al considerar una muestra parcial de estos procesos estocásticos con un soporte funcional finito, es posible verificar si se satisfacen las condiciones del teorema central del límite de Lindeberg-Feller. Para una mayor profundización sobre estos resultados, se puede consultar el libro de Kokoszka [16], entonces una región para estimar la función media puede ser definida como:

$$RC_{1-\alpha}(\mu(t)) \cong \bar{\chi}(t) \pm z_{\alpha/2} \hat{V}ar(\bar{\chi}(t))$$

Otra alternativa para generar curvas de confianza sobre la función media es estimar el parámetro y realizar simulaciones utilizando el método de Monte Carlo.

Supongamos que se busca poner a prueba la simetría en los movimientos espinales, es decir, la igualdad de las funciones medias de dos movimientos. En términos estadísticos, se plantea lo siguiente:

Sea $\chi_1(t), \chi_2(t), \dots, \chi_n(t)$ muestra funcional de $\chi(t) := \mu(t) + B_1(t)$ y sea $\Psi_1(t), \Psi_2(t), \dots, \Psi_m(t)$ muestra funcional de $\Psi(t) := \eta(t) + B_2(t)$ con $\{B_1(t)\}$ y $\{B_2(t)\}$ dos puentes brownianos clásicos independientes. El contraste de hipótesis de interés es:

$$\begin{cases} H_0 : \mu(t) = \eta(t) \\ H_1 : \mu(t) \neq \eta(t) \end{cases}$$

La simetría de los movimientos espinales hacia la derecha y la izquierda es de gran interés. En este sentido, surge la siguiente interrogante: ¿Son iguales los parámetros de volatilidad? Para abordar esta cuestión, se sugiere el siguiente procedimiento.

En primer lugar, se estima $\mu(t)$ y $\eta(t)$ a partir de los datos. Posteriormente, se procede a centrar los datos y se somete a prueba la siguiente hipótesis: $H_0 : \sigma_1 = \sigma_2$ vs $H_1 : \sigma_1 \neq \sigma_2$. En caso de que esta hipótesis no sea rechazada, se estima el parámetro de volatilidad de ambos puentes utilizando la varianza combinada entre los dos estimadores, según el siguiente estimador insesgado:

$$\hat{\sigma}^2 = \frac{(n-2)\hat{\sigma}_1^2 + (m-2)\hat{\sigma}_2^2}{n+m-4}$$

Utilizando este estimador, se pueden generar simulaciones para construir bandas de confianza a través del método de Monte Carlo. Si los parámetros de volatilidad son significativamente diferentes, entonces se realiza la estimación de las varianzas de manera independiente.

Tomando como función media la función seno definida en el intervalo $[0, \pi]$, y considerando como ruidos muestras simuladas de puentes brownianos clásicos y puentes brownianos fraccionarios, es importante notar que los movimientos espinales pueden presentar incrementos dependientes. Por esta razón, se recomienda comparar la bondad de ajuste entre el puente browniano clásico y el puente browniano fraccionario.

El contexto de los movimientos espinales se puede realizar este contraste para verificar si se cumple que los movimientos a derecha y a izquierda son simétricos. Por seguridad de los datos en este trabajo final la metodología se ilustrará con datos sintéticos definidos de la siguiente manera:

$$\chi_i(t) = \text{sen}(t) + B_i(t), i = 1, \dots, n$$

Con $B_1(t), \dots, B_n(t)$ puentes brownianos fraccionarios con H fijo, independientes con parámetro de volatilidad igual σ . Por otro lado sean

$$\Psi_i(t) = \text{sen}(t) + W_i(t), i = 1, \dots, m$$

Con $W_1(t), \dots, W_m(t)$ puentes brownianos fraccionarios con H fijo, independientes con parámetro de volatilidad igual σ . Note que son dos muestras funcionales con las mismas

características, si se requieren bandas de confianza de dichos procesos, realizando estimación de σ , H y $\mu(t)$, por el método de Monte Carlo algunas realizaciones de dicho proceso lucen de la siguiente manera:

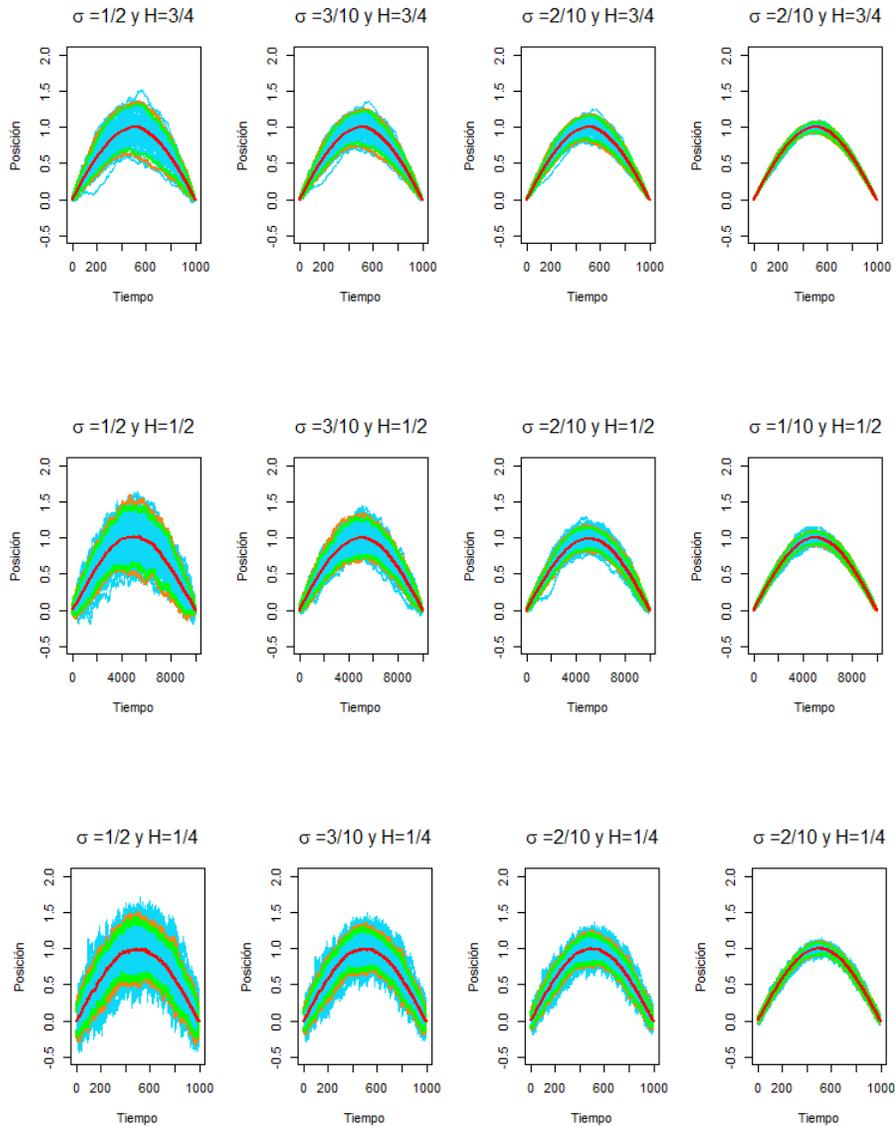


Figura 4-1: Realizaciones del modelo funcional con ruidos fraccionarios, considerando diferentes parámetros de Hurst y de volatilidad, junto con la función media y bandas de confianza al 95% y 99%. https://github.com/jrangelg/Trabajo_final_maestr-a/tree/main

El modelado de la estructura de covarianza utilizando la base cuadrada de Fourier implica la expansión de una función en términos de una base ortogonal de funciones cuadráticas. Esta base, denotada como F , contiene funciones que son cuadrados de las funciones clásicas de la base del espacio funcional considerado. Sea n la dimensión de esta base, se puede observar que la cantidad total de funciones en F es $\dim(F) = n^2$. Cada par de funciones en F es ortogonal y, por lo tanto, permite representar las relaciones de covarianza entre las distintas componentes de la función bajo estudio.

La utilización de la base cuadrada de Fourier para modelar la estructura de covarianza implica el empleo de n^2 parámetros para capturar la variabilidad y dependencia de los datos funcionales. Esta cantidad exponencial de parámetros está relacionada con la dimensionalidad cuadrática de la base, lo que puede conducir a problemas de sobreajuste y una representación excesiva de la información contenida en los datos.

En contraste, el puente browniano fraccional, caracterizado por su parámetro de Hurst H , representa una alternativa más parsimoniosa en el modelado de datos funcionales. Este enfoque utiliza un parámetro continuo H para modelar la estructura de dependencia y variabilidad. A través de la manipulación de H , el puente browniano fraccional puede adaptarse a diferentes niveles de suavidad y memoria a largo plazo en los datos funcionales, lo que permite una representación más eficiente y concisa de la estructura de covarianza.

La capacidad del puente browniano fraccional para capturar la estructura de dependencia con un número reducido de parámetros, en contraposición a la alta dimensionalidad presente en la base cuadrada de Fourier, destaca su utilidad en el análisis de datos funcionales. Además, su sensibilidad al parámetro de Hurst proporciona flexibilidad para modelar una amplia gama de comportamientos, lo que lo convierte en una herramienta valiosa en aplicaciones estadísticas que implican el análisis de datos funcionales con complejidades variables, mientras que la base cuadrada de Fourier representa una aproximación detallada pero altamente dimensional, el puente browniano fraccional ofrece una alternativa más eficiente y flexible para el modelado de estructuras de covarianza en datos funcionales, permitiendo un análisis más conciso y adaptativo en diversas aplicaciones estadísticas.

Con el propósito de establecer intervalos de confianza para el parámetro funcional $\mu(t) - \eta(t)$, se propone una metodología específica. Para ello, se consideran realizaciones parciales de $\chi_1(t), \chi_2(t), \dots, \chi_n(t)$ como muestras funcionales de $\chi(t) := \mu(t) + B_1(t)$, y $\Psi_1(t), \Psi_2(t), \dots, \Psi_m(t)$ como muestras funcionales de $\Psi(t) := \eta(t) + B_2(t)$, donde $B_1(t)$ y $B_2(t)$ representan dos puentes brownianos clásicos independientes con el mismo parámetro H .

Estas instancias de la función pueden ser organizadas sistemáticamente en una tabla con el propósito de realizar inferencia estadística sobre las funciones medias de manera adecuada

Tabla 4-1: Organización de las realizaciones parciales de datos funcionales

i	t_1	t_2	\cdots	t_k
1	x_{1,t_1}	x_{1,t_2}	\cdots	x_{1,t_k}
2	x_{2,t_1}	x_{2,t_2}	\cdots	x_{2,t_k}
\vdots	\vdots	\vdots	\ddots	\vdots
n	x_{n,t_1}	x_{n,t_2}	\cdots	x_{n,t_k}

i	t_1	t_2	\cdots	t_k
1	y_{1,t_1}	y_{1,t_2}	\cdots	y_{1,t_k}
2	y_{2,t_1}	y_{2,t_2}	\cdots	y_{2,t_k}
\vdots	\vdots	\vdots	\ddots	\vdots
m	y_{m,t_1}	y_{m,t_2}	\cdots	y_{m,t_k}

Es importante notar que, en esta instancia, se analizan realizaciones de datos funcionales de igual tamaño. Sin embargo, como una perspectiva para trabajos futuros, se plantea la exploración de métodos para tratar con realizaciones de datos funcionales de tamaño irregular en distintos individuos. En el contexto actual, se considera una serie de datos funcional regular.

Es relevante destacar que $\chi(t) - \Psi(t)_t$ bajo la hipótesis nula asume la forma de un puente browniano con parámetros H y σ desconocidos. Para ajustar el coeficiente fraccionario, se evalúa la adecuación del modelo mediante el intervalo de cuantiles del valor p, definido como el intervalo de tamaño $1 - \alpha$ de los n valores p de cada evaluación de normalidad. Este proceso se lleva a cabo en una grilla que abarca el intervalo (0,1). En caso de un completo desconocimiento sobre la dependencia del puente, un valor menor a 1/2 sugiere una similitud con el ruido propio de un proceso de ruido blanco, lo cual indicaría una posible adecuación a un modelo de regresión clásica. Por otro lado, un valor cercano a 1/2 implicaría incrementos independientes, sugiriendo la presencia de una memoria prolongada en el puente.

Siguiendo la línea de razonamiento planteada en el test de permutaciones [15], se puede considerar el siguiente esquema para considerar las diferencias entre los datos funcionales. Es de relevancia notar que esta organización tiene como supuesto que los datos parciales se tomaron de manera regular

Tabla 4-2: Organización de datos combinando las realizaciones parciales de las dos muestras funcionales

i	t_1	t_2	\cdots	t_k
1,1	$x_{1,t_1} - y_{1,t_1}$	$x_{1,t_2} - y_{1,t_2}$	\cdots	$x_{1,t_k} - y_{1,t_k}$
1,2	$x_{1,t_1} - y_{2,t_1}$	$x_{1,t_2} - y_{2,t_2}$	\cdots	$x_{1,t_k} - y_{2,t_k}$
\vdots	\vdots	\vdots	\ddots	\vdots
n, m	$x_{n,t_1} - y_{m,t_1}$	$x_{n,t_2} - y_{m,t_2}$	\cdots	$x_{n,t_k} - y_{m,t_k}$

Sobre la presente realización del puente, se procede a la construcción de bandas de confianza mediante técnicas de remuestreo, considerando diversos valores de H . Posteriormente, se evalúa, en esos diferentes modelos, el cuantil de los valores p con el fin de determinar cuál modelo se ajusta mejor a los datos. Además, se verifica la validez de las bandas de confianza en el contexto de cada modelo seleccionado.

Para evaluar la idoneidad de los intervalos de los valores p en un contexto donde las observaciones se presumen originadas a partir de un proceso de puente browniano ajustado a un valor fraccionario específico, es esencial considerar la estimación del parámetro de volatilidad y examinar la bondad de ajuste del modelo. Una de las alternativas propuestas es llevar a cabo una prueba de normalidad multivariada. Sin embargo, Srivastava [31] señala que los tests de normalidad son susceptibles a fallos en situaciones de alta dimensionalidad.

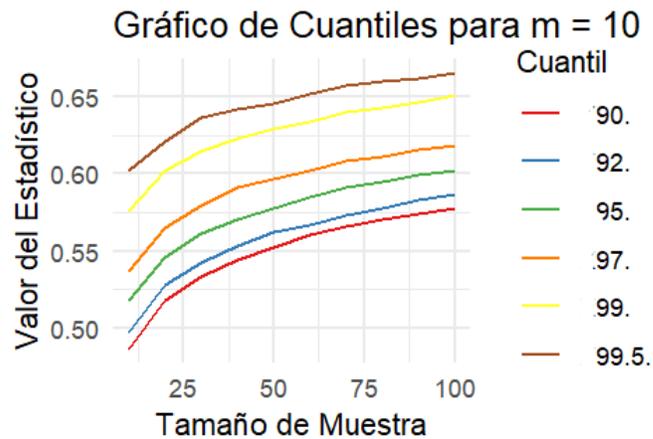
Otra alternativa planteada sería utilizar la propiedad de Bonferroni [23]. No obstante, si el tamaño de la muestra es reducido, como por ejemplo 20, mantener una significancia individual de 0.01 implicaría una significancia grupal de 0.2, un nivel que resulta elevado. Una estrategia diferente es emplear la técnica estadística de envelope o envolvimiento, la cual consiste en tomar una estadística para cada individuo, observar su comportamiento bajo la hipótesis nula y comparar estos estadísticos resultantes en nuestra muestra con los simulados bajo la hipótesis nula. Para un análisis más detallado sobre las técnicas de envolvimiento, se puede consultar el artículo de Dennis Cook [7], donde se expone teóricamente esta metodología.

La mayoría de las pruebas de normalidad consideran una distancia entre la distribución empírica y la teórica. Esto justifica el uso de la distancia de Kolmogorov-Smirnov como medida de error para validar la bondad de ajuste. Esta motivación inspira la realización del proceso de embedding sobre dicha estadística para evaluar su comportamiento. Para validar esto, se llevaron a cabo 10,000 simulaciones de realizaciones parciales de puentes brownianos, incrementando el tamaño en unidades de 10 hasta llegar a 100 en cada simulación. Durante este proceso, se evaluó un soporte regular de tamaño 10. Es importante destacar que el enfoque para muestras con $m > 10$ sigue el mismo principio. Se extrajeron realizaciones parciales de puentes brownianos, manteniendo un coeficiente fraccionario constante. Posteriormente, se procedió a realizar la rotación, siguiendo el procedimiento descrito en la Proposición 13. Finalmente, se llevó a cabo una prueba de normalidad con media cero y se calcularon los cuantiles.

Tabla 4-3: Cuantiles del test Kolmogorov Smirnov para $m = 10$

n	Cuantiles					
	90%	92%	95%	97%	99%	99.5%
10	0.486	0.497	0.518	0.537	0.576	0.602
20	0.518	0.528	0.546	0.565	0.602	0.621
30	0.533	0.542	0.561	0.579	0.615	0.636
40	0.544	0.553	0.570	0.591	0.623	0.642
50	0.552	0.562	0.578	0.597	0.629	0.645
60	0.560	0.567	0.585	0.602	0.634	0.652
70	0.566	0.573	0.591	0.608	0.640	0.657
80	0.570	0.578	0.595	0.611	0.643	0.660
90	0.574	0.583	0.599	0.616	0.646	0.662
100	0.578	0.587	0.602	0.618	0.651	0.665

Para evaluar si bajo un soporte de $m = 10$, la simulación del estadístico crítico de envolvimiento depende del tamaño de muestra se pueden graficar los valores del estadístico para los diferentes niveles de confianza y los diferentes tamaños de muestra

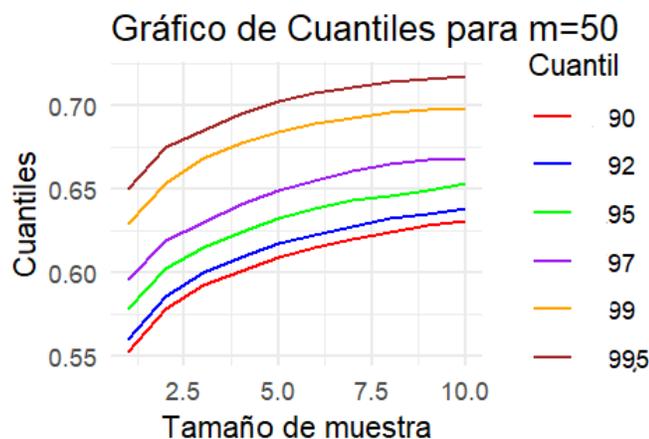
**Figura 4-2:** Cuantiles del test Kolmogorov Smirnov para $m = 10$

https://github.com/jrangelg/Trabajo_final_maestr-a/tree/main

Se observa que a medida que aumenta el tamaño de muestra, el estadístico de envolvimiento también aumenta, lo cual es intuitivo. Se puede notar que el estadístico tiende a ser más grande al aumentar el nivel de confianza deseado. Asimismo, se busca determinar estos límites de envolvimiento para diferentes tamaños de muestra (m). En este trabajo, se presentarán resultados para tamaños parciales del soporte iguales a $m = 50$. No obstante, el procedimiento es el mismo para otros tamaños de soporte.

Tabla 4-4: Cuantiles del test Kolmogorov Smirnov para $m = 50$

n	Cuantiles					
	90%	92%	95%	97%	99%	99.5%
10	0.552	0.560	0.578	0.596	0.629	0.650
20	0.578	0.586	0.602	0.619	0.653	0.675
30	0.592	0.600	0.615	0.630	0.668	0.685
40	0.601	0.609	0.624	0.641	0.677	0.695
50	0.609	0.617	0.632	0.649	0.684	0.702
60	0.615	0.622	0.638	0.655	0.689	0.707
70	0.620	0.627	0.643	0.661	0.692	0.711
80	0.624	0.632	0.646	0.665	0.696	0.714
90	0.628	0.635	0.649	0.667	0.697	0.716
100	0.631	0.638	0.653	0.668	0.698	0.717

**Figura 4-3:** Cuantiles del test Kolmogorov Smirnov para $m = 50$

https://github.com/jrangelg/Trabajo_final_maestr-a/tree/main

Se evidencia un comportamiento similar al que se veía en $m = 10$; sin embargo, la escala aumenta, indicándonos que si se aumenta el soporte sobre los datos funcionales, el comportamiento de la estadística también aumenta. Para otro tamaño diferente del soporte, la técnica de envolvimiento se realiza de la misma manera. Aunque la técnica de envolvimiento, en comparación con la cota de Bonferroni, tiene la flexibilidad de utilizarse con tamaños de muestra grandes, tiene el defecto de requerir una cantidad elevada de cómputo.

4.2 Puente browniano fraccionario para modelar trayectorias de migración

Otra aplicación de los puentes estocásticos se encuentra en el modelado de las trayectorias seguidas por una población de una determinada especie biológica. Sin embargo, en la literatura, se suele emplear comúnmente el puente browniano clásico. En este estudio, se propone la utilización del puente browniano fraccional con $H > 1/2$, ya que las trayectorias generadas por este son menos erráticas.

Para aplicar el puente browniano fraccional, se desarrolla teóricamente un homeomorfismo en el intervalo $[0, 1]$, conectando este intervalo con un segmento de recta que parte desde el punto inicial. Supongamos una situación en la cual una especie se desplaza desde el punto $(1, 1)$ al punto $(5, 8)$, donde la especie descansa. Posteriormente, parte desde este punto hacia $(8, 6)$.

Se hace uso del teorema que establece que todo segmento de recta es homeomorfo al intervalo $(0, 1)$; mediante simulaciones del puente clásico, se lleva a cabo la transformación al espacio deseado.

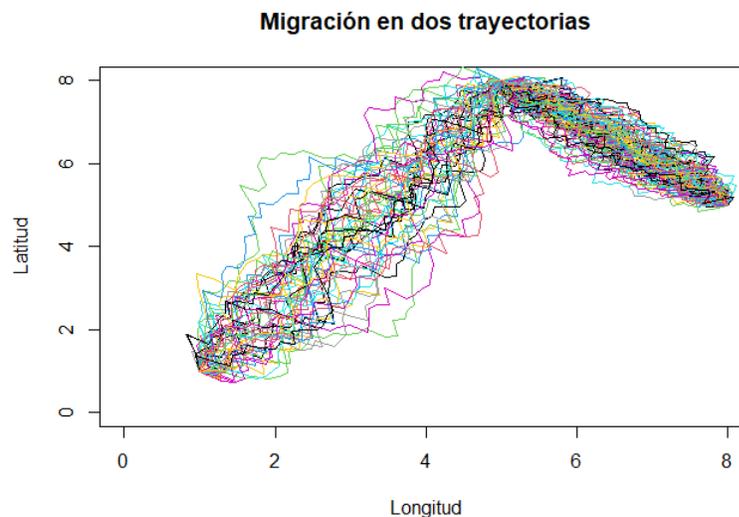


Figura 4-4: Modelo de migración con puente browniano clásico

https://github.com/jrangelg/Trabajo_final_maestr-a/tree/main

En el trabajo de referencia [19], se aborda la exploración del uso del puente clásico en el análisis de trayectorias migratorias de diversas especies. No obstante, se ha evidenciado que el movimiento de algunas especies no es completamente impredecible, como se modela con el puente browniano clásico. Por esta razón, este estudio propone la adopción del puente browniano fraccional, ya que el parámetro H regula la suavidad de la trayectoria.

Es esencial destacar que si $H > 1/2$, la trayectoria tiende a ser más suave, mientras que si $H < 1/2$, la trayectoria tiende a ser más errática.

La metodología presentada en el trabajo de Kranstauber [19] se describe de la siguiente manera:

- i. Seguir a n individuos en un soporte parcial del espacio.
- ii. Realizar la rotación de los datos para obtener una muestra aleatoria sin correlación entre los incrementos del movimiento.
- iii. Estimar el parámetro de volatilidad.
- iv. Simular la región de confianza del movimiento de las aves.

En el ejercicio de simulación mencionado anteriormente, se muestran las regiones de confianza correspondientes, como se ilustra en la Figura

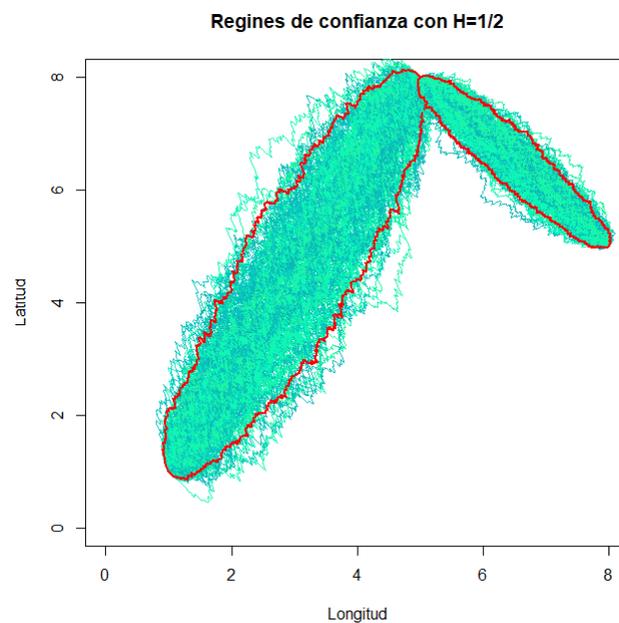


Figura 4-5: Región de confianza del 95% bajo un $H = 1/2$

https://github.com/jrangelg/Trabajo_final_maestr-a/tree/main

Sin embargo, el puente browniano clásico solo considera este comportamiento de volatilidad. En el contexto de esta investigación, ampliaremos nuestra perspectiva al explorar diversos parámetros fraccionarios y variaciones en los parámetros de volatilidad. Este enfoque nos permitirá realizar un análisis comparativo exhaustivo de los comportamientos resultantes.

Al introducir parámetros fraccionarios, estamos extendiendo la modelización clásica para capturar fenómenos más complejos y realistas. La inclusión de diferentes valores de fracción de Hurst (H) nos permitirá explorar cómo la rugosidad de las trayectorias del

punte browniano se ve afectada por estas variaciones. Asimismo, al ajustar los parámetros de volatilidad, estamos considerando la influencia directa en la magnitud de las fluctuaciones en el proceso estocástico. Al analizar diferentes niveles de volatilidad, podremos identificar patrones y tendencias emergentes, así como evaluar la sensibilidad del modelo a estos cambios.

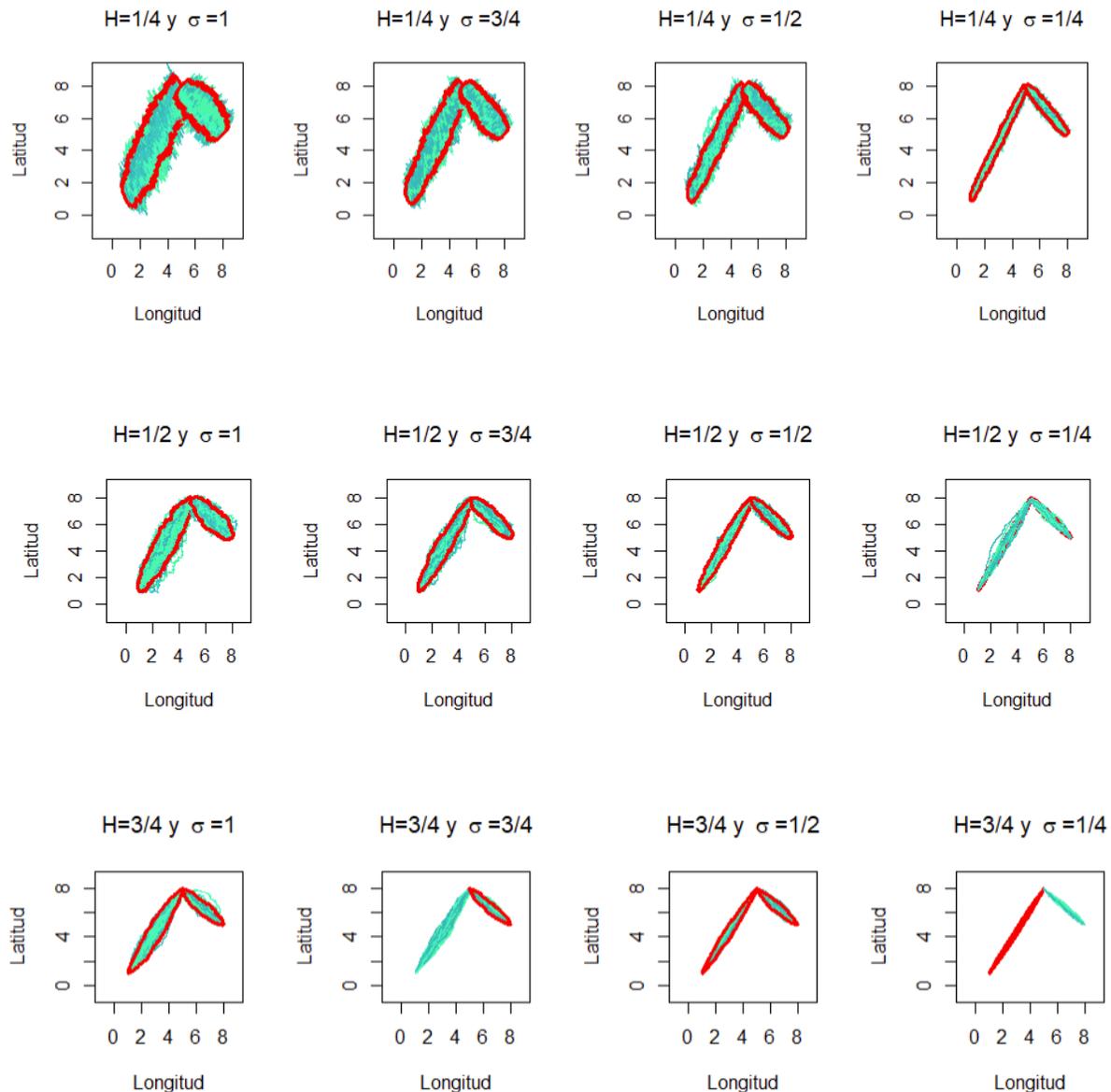


Figura 4-6: Realizaciones del puente browniano fraccionario para modelar migraciones, considerando diferentes parámetros de Hurst y de volatilidad, junto con las bandas de confianza al 95%.

https://github.com/jrangelg/Trabajo_final_maestr-a/tree/main

La observación de que, a diferencia del modelado con puente browniano clásico, al utilizar un parámetro de Hurst (H) cercano a cero el ruido tiende a ser similar a un ruido blanco, mientras que si es cercano a uno la dependencia y la volatilidad no aumentan tan rápido, es muy interesante y revela características fundamentales del puente browniano fraccionario.

Cuando H se acerca a cero, el proceso se asemeja al comportamiento de un ruido blanco. En este caso, las fluctuaciones aleatorias en el sistema son independientes entre sí y no presentan una memoria a largo plazo. En otras palabras, cada incremento o decremento en el proceso no está fuertemente correlacionado con los valores anteriores, lo que se traduce en un comportamiento más aleatorio y menos predecible. Por otro lado, cuando H se acerca a uno, el proceso exhibe una mayor persistencia y memoria a largo plazo. Las fluctuaciones anteriores tienen una influencia más pronunciada en las futuras, lo que resulta en una trayectoria más suave y menos propensa a cambios abruptos.

Este comportamiento refleja una mayor dependencia en el pasado, y la volatilidad no aumenta tan rápidamente como en el caso de H cercano a cero. Estas observaciones son coherentes con la interpretación del parámetro de Hurst en el contexto del puente browniano fraccionario. Un valor bajo de H indica una menor autocorrelación y una mayor semejanza con un ruido blanco, mientras que un valor alto de H implica una mayor persistencia y dependencia a largo plazo en el proceso.

Hacer uso del puente browniano fraccionario en la modelación migratoria de aves ofrece una valiosa expansión en la variedad de puentes disponibles para este propósito. En comparación con el puente browniano clásico, la inclusión de parámetros fraccionarios introduce una mayor flexibilidad en la representación de la volatilidad, permitiendo capturar de manera más precisa la complejidad de los patrones migratorios de las aves.

Este enfoque innovador no solo enriquece la gama de herramientas disponibles, sino que también proporciona una visión más completa y detallada de las características del puente browniano en el contexto de la migración de aves. Al explorar diferentes conjuntos de parámetros fraccionarios, podemos analizar cómo variaciones en la rugosidad de las trayectorias impactan en la predicción de movimientos migratorios. Aunque esta metodología ha sido abordada por Kranstauber para resolver problemas asociados con la anisotropía y la inclusión de covariables, aún no se ha incorporado el coeficiente fraccionario, ni las técnicas inferenciales, ni las pruebas de bondad de ajuste del modelo. Estos elementos podrían representar áreas potenciales de mejora. Para contrastar estos comportamientos, se recurre al método de Cholesky. Este método se utiliza para visualizar y analizar el comportamiento de los movimientos brownianos fraccionarios, proporcionando una perspectiva más detallada y precisa sobre la naturaleza de las trayectorias observadas.

El método de Cholesky [9] es valioso en este contexto ya que permite descomponer una matriz de covarianza en el producto de una matriz triangular inferior y su transpuesta. Al aplicar este método a los datos de movimientos brownianos fraccionarios, se puede obtener una representación más clara de las relaciones y dependencias entre las diferentes etapas del movimiento. Este enfoque contribuye a una comprensión más profunda de las estructuras subyacentes en las trayectorias, permitiendo identificar patrones, dependencias y cambios en la volatilidad a lo largo del tiempo. En consecuencia, el método de Cholesky se convierte en una herramienta valiosa para el análisis detallado de los movimientos brownianos fraccionarios y proporciona una visión enriquecida que puede complementar y mejorar la interpretación de los resultados obtenidos mediante otros métodos.

Es relevante destacar que hasta el momento, la aplicación de métodos de puente browniano fraccionario en la modelación migratoria de aves no ha sido sometida a un análisis riguroso en cuanto a la elección de tipos de coordenadas. Diferentes sistemas de coordenadas, como esféricas, planas o de acuerdo con geoide, pueden tener un impacto significativo en la representación y predicción de los movimientos migratorios. En la teoría de estadística espacial, se ha avanzado en la incorporación de diversos tipos de coordenadas para adaptarse a la complejidad de los datos geoespaciales. Sin embargo, este enfoque aún no se ha trasladado completamente a la modelación migratoria utilizando puente browniano fraccionario.

Un área de investigación futura y fascinante podría derivarse de la exploración de cómo la elección del sistema de coordenadas afecta la validez y eficacia de la metodología del puente browniano fraccionario en la migración de aves. Esto podría incluir el análisis de coordenadas esféricas para tener en cuenta la curvatura de la Tierra, la utilización de coordenadas planas para áreas geográficas más limitadas y la consideración de sistemas de coordenadas basados en el geoide para una representación más precisa de la topografía terrestre.

Este trabajo futuro podría no solo mejorar la aplicabilidad de la metodología en diferentes contextos geográficos, sino también proporcionar una comprensión más profunda de cómo las elecciones de coordenadas afectan la interpretación de los patrones migratorios. En última instancia, esta investigación adicional contribuirá al avance y refinamiento de los modelos de puente browniano fraccionario en el campo de la ecología y la migración de aves.

4.3 Aplicación en lenguaje natural

Una aplicación innovadora de los puentes estocásticos es su implementación para introducir ruido en redes neuronales con diversas arquitecturas. En particular, se destaca el artículo publicado en 2023 [34], que presenta el método COLOR (Coherent dialOgue pLanning approach via BrOwnian bRidge) para aplicar ruido modelado por un puente estocástico clásico a un chatbot. La contribución adicional a este modelo implica la inclusión del puente browniano fraccionario como componente de ruido, denominando a esta versión mejorada como COLFOR (Coherent dialOgue pLanning approach via Fractional BrOwnian bRidge). Este enfoque busca enriquecer la capacidad del chatbot para manejar la variabilidad en las interacciones, permitiendo una adaptación más efectiva a la diversidad de situaciones en las conversaciones. Al incorporar el puente browniano fraccionario, COLFOR busca capturar de manera más precisa la complejidad temporal de los datos y las dinámicas no lineales presentes en las interacciones del chatbot.

La introducción de este componente adicional busca mejorar la robustez y la capacidad de generalización del modelo, proporcionándole una mayor flexibilidad para comprender y responder de manera más eficiente a las variaciones en el lenguaje y en el contexto de las conversaciones. Este enfoque representa un paso hacia adelante en la aplicación de métodos estocásticos avanzados en el desarrollo de modelos de procesamiento del lenguaje natural, ofreciendo potencialmente mejoras significativas en la calidad y adaptabilidad de los chatbots.

La mayoría de modelos de lenguaje natural usan el embedding el cual es un mecanismo que consiste en representar de manera semántica y compacta objetos, como palabras o frases, en un espacio vectorial. En este espacio, la proximidad entre vectores refleja la similitud semántica entre los objetos que representan. Estos vectores, también conocidos como embeddings, capturan relaciones y significados contextuales, facilitando el análisis y procesamiento de información compleja. Ahora bien, un embedding con ruido se refiere a la incorporación de componentes estocásticos o aleatorios en el proceso de generación de los vectores embeddings. En términos estadísticos, esto implica introducir variabilidad controlada en la representación vectorial de palabras o frases. Un enfoque específico es el embedding con ruido profundo, donde se incorpora ruido a lo largo de múltiples capas o niveles del modelo, permitiendo una modelación más sofisticada de la incertidumbre inherente en los datos.

Desde la perspectiva de la tesis, la implementación de embeddings con ruido profundo busca mejorar la capacidad del modelo para manejar la variabilidad y la incertidumbre en los datos lingüísticos. Este enfoque puede contribuir a la robustez del modelo frente a la variabilidad del lenguaje natural y aportar a la capacidad de generalización del modelo,

permitiéndole adaptarse de manera más efectiva a diferentes contextos y matices semánticos presentes en el corpus de datos. La inclusión de componentes estocásticos también puede ayudar a capturar la complejidad estadística subyacente en las relaciones semánticas, mejorando así la calidad y la capacidad predictiva del modelo de embedding.

Un modelo dirigido por objetivos es una categoría de modelos de generación de diálogos que se concentra en la consecución de metas específicas durante una conversación. A diferencia de los modelos de generación de diálogos generales, que pueden ser más abiertos y no poseer un objetivo específico, los modelos dirigidos por objetivos están diseñados para seguir una ruta planificada que conduzca a la realización de un objetivo predefinido.

Estos modelos encuentran aplicación en diversas áreas, como sistemas de recomendación, asistentes virtuales y chatbots, donde es crucial dirigir la conversación hacia una meta específica, como proporcionar información, efectuar recomendaciones o llevar a cabo una tarea determinada. El proceso de generación de diálogos dirigidos por objetivos implica la planificación de una ruta lógica que conecte el contexto actual con el objetivo deseado en cada turno de la conversación. Esto conlleva tomar decisiones sobre los temas a abordar, las acciones a llevar a cabo y el orden en que se deben tratar. El propósito es lograr una transición fluida y coherente hacia el objetivo final.

En [38] se explica que estos modelos suelen hacer uso de técnicas de planificación y aprendizaje automático para determinar la mejor ruta de diálogo y generar respuestas pertinentes que se alineen con el objetivo establecido. También pueden integrar conocimientos específicos relacionados con el objetivo, como bases de datos o información contextual, para mejorar la calidad y relevancia de las respuestas generadas.

En este estudio, se plantea la utilización de un puente browniano fraccionario debido a la presencia de una dependencia secuencial inherente en los modelos de procesamiento del lenguaje natural. A diferencia del puente browniano clásico, que no toma en cuenta esta dependencia secuencial, la elección del puente browniano fraccionario se fundamenta en la necesidad de capturar de manera más precisa la estructura temporal compleja presente en los datos lingüísticos. El puente browniano fraccionario es especialmente adecuado para abordar la dependencia secuencial en modelos de lenguaje natural, ya que permite modelar correlaciones a largo plazo y capturar patrones temporales no lineales. Esto es crucial en contextos donde la secuencia de palabras o tokens en una oración o texto juega un papel esencial para comprender el significado y la coherencia semántica.

La introducción del puente browniano fraccionario en el modelo contribuye a una representación más realista de la dinámica temporal subyacente en el lenguaje natural. Este enfoque

es particularmente valioso en aplicaciones donde la captura precisa de la dependencia secuencial es esencial, como en la generación de texto coherente, la traducción automática o cualquier tarea que involucre la comprensión y producción de secuencias lingüísticas. La representación de un embedding básico se puede expresar mediante un vector matemático.

Basados en el artículo de Xu (2021) [39] se generaliza de la siguiente manera el embedding con ruidos. Sea \mathbf{v} la representación vectorial de una palabra o frase, el embedding básico se define como:

$$\mathbf{v} = [v_1, v_2, \dots, v_n]$$

donde v_i es la i -ésima dimensión del vector.

Para incorporar ruido estocástico en el embedding, introducimos un término adicional. Sea \mathbf{v}_r el vector de embedding con ruido, y \mathbf{v} el embedding básico, entonces:

$$\mathbf{v}_r = \mathbf{v} + \boldsymbol{\epsilon}$$

donde $\boldsymbol{\epsilon}$ representa el vector de ruido estocástico.

Para modelar la dependencia secuencial, consideramos un puente browniano fraccionario. La adición de este componente se expresa como:

$$\boldsymbol{\epsilon} = \int_0^t B_H(t-s) dW_s$$

donde $B_H(t-s)$ es el kernel del puente browniano fraccionario y dW_s es la diferencial de un proceso de Wiener.

La integración de ruido, particularmente mediante el uso de un puente browniano fraccionario, en los embeddings ha demostrado ser una estrategia efectiva para enriquecer la representación de la dependencia secuencial en modelos de procesamiento del lenguaje natural. Esta técnica permite capturar de manera más precisa las complejidades temporales inherentes en el lenguaje, mejorando así la capacidad del modelo para comprender y generar texto de manera más coherente y contextual.

A pesar de los avances logrados con la incorporación de ruido en los embeddings, se enfrenta a un desafío crítico y de alto interés en la investigación: la evaluación de la bondad de ajuste en estos modelos. La presencia de redes neuronales que componen los diversos componentes de ruido complica la tarea de evaluar la calidad del ajuste del modelo. La naturaleza intrincada de estas arquitecturas hace que sea difícil desentrañar la contribución individual de cada elemento de ruido, dificultando así la validación precisa

del modelo.

Este problema abierto representa un área crucial para el avance de la investigación en embeddings con ruido. La comunidad científica se encuentra desafiada a desarrollar métodos innovadores y robustos para evaluar la bondad de ajuste en modelos de lenguaje natural mejorados con ruido estocástico. La resolución de este desafío no solo contribuirá al perfeccionamiento de las técnicas existentes, sino que también abrirá nuevas vías para el desarrollo de modelos más confiables y generalizables en el ámbito del procesamiento del lenguaje natural. La superación de este obstáculo promete impulsar significativamente la eficacia y aplicabilidad de los embeddings con ruido en diversas aplicaciones prácticas.

Tabla 4-5: Simulaciones de embeddings de la frase La universidad nacional es una de las mejores universidades en Colombia con movimientos brownianos con $\sigma = 0.1$

No.	Frase Reconstruida
1	La universidad nacional es una de las mejores universidades en Colombia
2	Nacional una universidades es mejores Colombia las en de universidad
3	En una universidad la mejores Colombia universidades es nacional de
4	Las una Colombia universidad en nacional mejores universidades es de
5	Una universidades en mejores nacional Colombia de la es universidad
6	Mejores nacional de es universidades las en la una Colombia
7	Nacional mejores las de universidades una universidad la Colombia es
8	Las de mejores universidad una nacional en universidades Colombia es
9	Colombia nacional universidad de en es la las mejores una
10	Mejores Colombia la de nacional en es universidad una las

Una pregunta que surge es si hay un cambio significativo al modificar el parámetro a 0.05. Veamos las simulaciones para explorar cómo esta variación afecta los resultados. Al ajustar este parámetro, es posible que se experimenten diferencias notables en la dinámica del sistema o en los resultados obtenidos. Esto podría tener implicaciones importantes, ya que los parámetros a menudo influyen en el comportamiento y la estabilidad de los modelos o simulaciones. Por lo tanto, es crucial examinar de cerca cómo la variación en este parámetro específico impacta en las condiciones simuladas y si estas variaciones tienen alguna relevancia práctica o teórica. Este análisis más detallado nos permitirá comprender mejor la sensibilidad del sistema a diferentes configuraciones de parámetros y, en última instancia, mejorar nuestra comprensión global del fenómeno o modelo que estamos estudiando.

Tabla 4-6: Simulaciones de embeddings de la frase La universidad nacional es una de las mejores universidades en Colombia con movimientos brownianos con $\sigma = 0.05$

No.	Frase Original
1	La universidad nacional es una de las mejores universidades en Colombia
2	Nacional una universidades es mejores Colombia las en de universidad
3	En una universidad la mejores Colombia universidades es nacional de
4	Las una Colombia universidad en nacional mejores universidades es de
5	Una universidades en mejores nacional Colombia de la es universidad
6	Mejores nacional de es universidades las en la una Colombia
7	Nacional mejores las de universidades una universidad la Colombia es
8	Las de mejores universidad una nacional en universidades Colombia es
9	Colombia nacional universidad de en es la las mejores una
10	Mejores Colombia la de nacional en es universidad una las

La apreciación sobre la estacionariedad de los incrementos en el movimiento browniano y la reducida volatilidad en las fronteras del puente browniano refleja una comprensión profunda de estas dinámicas estocásticas. Mientras el movimiento browniano exhibe incrementos estacionarios a lo largo del tiempo, el puente browniano, al introducir anclajes en sus extremos, logra mitigar la variabilidad extrema en esas áreas. Esta característica hace que el puente browniano sea una herramienta valiosa para modelar procesos que requieren una transición suave entre estados o momentos clave, proporcionando una representación más realista de las fluctuaciones en ciertos contextos.

La investigación acerca de si el puente browniano mejora la cohesión en los embeddings revela una perspectiva esencial para comprender cómo las modificaciones en la generación de ruido afectan la calidad de la representación semántica. Al incorporar ruido mediante el puente browniano, se plantea la posibilidad de suavizar las transiciones entre palabras o conceptos cercanos en el espacio semántico. Este modelo fue propuesto por Jian Wang en 2023; ahora, examinemos algunas simulaciones bajo esta propuesta.

Tabla 4-7: Simulaciones de embeddings de la frase La universidad nacional es una de las mejores universidades en Colombia con puentes brownianos con $\sigma = 0.1$

No.	Frase Original
1	La universidad nacional es una de las mejores universidades en Colombia
2	Nacional una universidades es mejores Colombia las en de universidad
3	En una universidad la mejores Colombia universidades es nacional de
4	Las una Colombia universidad en nacional mejores universidades es de
5	Una universidades en mejores nacional Colombia de la es universidad
6	Mejores nacional de es universidades las en la una Colombia
7	Nacional mejores las de universidades una universidad la Colombia es
8	Las de mejores universidad una nacional en universidades Colombia es
9	Colombia nacional universidad de en es la las mejores una
10	Mejores Colombia la de nacional en es universidad una las

En esta sección, es importante aclarar que no se tiene la intención de calibrar el modelo COLFOR. La calibración de dicho modelo requeriría una base de datos considerablemente extensa de conversaciones dirigidas que permitan entrenar el modelo de manera efectiva. Sin embargo, se propone y explica la implementación de un embedding utilizando un ejemplo de juguete con ruido browniano fraccionario. Esta aproximación amplía las posibilidades, ya que en [34] solo se considera el puente browniano clásico. Es relevante destacar que los modelos de lenguaje natural, dada su naturaleza de dependencia al generar texto, suelen basarse en redes neuronales recurrentes. Por esta razón, resulta más adecuado incorporar un ruido con memoria a largo plazo, como lo proporciona el puente fraccionario. No obstante, se plantea como un problema de investigación interesante la implementación y evaluación de este enfoque sobre un conjunto significativo de entrenamiento.

Adicionalmente, es fundamental comprender que la propuesta aquí presentada no busca resolver la calibración completa del modelo COLFOR, sino más bien introduce una técnica novedosa que expande las posibilidades y considera aspectos más complejos del proceso de generación de texto. La utilización de ruido browniano fraccionario ofrece una alternativa atractiva y se espera que futuras investigaciones exploren su viabilidad y rendimiento en conjuntos de datos más extensos. Además, se destaca la necesidad de considerar la memoria a largo plazo en modelos de lenguaje natural, lo cual puede ser crucial para capturar adecuadamente las complejidades del lenguaje y mejorar la calidad de las generaciones textuales.

Finalmente, queda planteada como una tarea futura de investigación la evaluación del comportamiento del embedding propuesto sobre un conjunto considerable de entrenamiento. Esta evaluación permitirá comprender mejor las ventajas y limitaciones del enfoque, así

como su aplicabilidad en escenarios más realistas. Se anima a la exploración continua de nuevas estrategias y técnicas en el campo de modelos de lenguaje natural, con la esperanza de mejorar la capacidad predictiva y generativa de estos modelos en diversos contextos y aplicaciones.

5 Conclusiones

Este trabajo ofrece un análisis comparativo de los puentes estocásticos, explorando diversas metodologías inferenciales basadas en la estadística clásica para la estimación del parámetro de volatilidad asociado a dichos puentes. Se aborda la simulación de datos funcionales con ruido modelado a partir de puentes, y se introducen regiones de confianza. Además, se presenta un método basado en técnicas de involucramiento para validar la bondad de ajuste del modelo propuesto. La investigación también incluye la simulación de trayectorias de animales mediante el uso de puentes brownianos fraccionarios, incorporando la velocidad con la que se dispersan o concentran. Como último aporte, se propone la aplicación del puente browniano fraccionario para mejorar la calibración del modelo COLOR. Estas diversas contribuciones resaltan la versatilidad y aplicabilidad de los puentes estocásticos en distintos contextos y problemáticas, ofreciendo herramientas valiosas para la modelación y análisis de procesos estocásticos complejos.

En el tercer capítulo, se detalló la metodología para llevar a cabo inferencia clásica tanto sobre el puente browniano clásico como sobre el puente browniano fraccionario. Se desarrollaron estimadores insesgados, se propuso un método para evaluar la bondad de ajuste del modelo y se delineó la forma de realizar contrastes de hipótesis. Sin embargo, como trabajo futuro, se plantea la exploración de la estadística bayesiana aplicada a los parámetros de las ecuaciones diferenciales, tanto en el contexto del puente browniano como en otras ecuaciones, teniendo en cuenta la limitación de contar con solo una realización parcial del proceso.

Estas estimaciones pretenden incorporar tanto la perspectiva de expertos como datos históricos, permitiendo alimentar el modelo de manera secuencial. Este enfoque también abre la puerta a la inclusión de ecuaciones diferenciales estocásticas en áreas más amplias de la estadística, promoviendo un análisis más completo y flexible de diversos fenómenos estocásticos. Un problema de gran interés radica en la búsqueda de un estimador analítico para el coeficiente fraccionario. Esta búsqueda no solo tiene el potencial de reducir significativamente los recursos computacionales requeridos, sino que también proporcionaría una precisión esencial en la selección del parámetro H . Encontrar una solución analítica para este estimador abriría la puerta a una metodología más eficiente y precisa en la estimación de este componente clave de los puentes estocásticos.

En la sección 4.1 se presenta una propuesta para la aplicación de puentes estocásticos como ruidos en datos funcionales. Los datos funcionales suelen residir en espacios de Hilbert, y el operador de covarianza que caracteriza la autocorrelación de los datos opera en el producto tensorial cuadrático del espacio. Esto conlleva la necesidad de estimar un gran número de parámetros. Una ventaja del uso de puentes estocásticos como ruido radica en la reducción de la cantidad de parámetros a estimar, limitándose a dos: el coeficiente de volatilidad y el coeficiente fraccionario.

Es importante destacar que la viabilidad de utilizar puentes estocásticos como ruido está condicionada a la variabilidad de la función media, que debe ser pequeña en los extremos del soporte y aumentar a medida que se aleja de la frontera. Se subraya que, aunque esta propuesta simplifica la estimación de parámetros, el proceso clásico con bases funcionales describe de manera más precisa el ruido, especialmente para las bandas de confianza. Adicionalmente, se explora el uso de la técnica de embedding para evaluar la bondad de ajuste. En contraste con Bonferroni, esta técnica muestra una mayor potencia sin perderla fácilmente. Además, demuestra ser más robusta que los tests de normalidad multivariada, conocidos por perder robustez en entornos de alta dimensionalidad.

En la sección 4.2, se realiza una comparación del uso de diversos puentes estocásticos en trayectorias que conectan dos puntos. A diferencia de los enfoques tradicionales que consideran exclusivamente el puente browniano clásico en estudios espaciales, se destaca la ausencia de procedimientos de bondad de ajuste en estos modelos. Esta carencia puede limitar las posibilidades de capturar aumentos flexibles en la volatilidad. Para abordar esta limitación, se introduce el puente browniano fraccionario, que permite una mayor flexibilidad en los aumentos en la volatilidad. Su inclusión se revela como un componente crucial para enriquecer la capacidad explicativa del modelo y capturar de manera más precisa la complejidad de las trayectorias entre dos puntos específicos. Como trabajo futuro, queda evaluar el puente browniano para trayectorias anisotrópicas y puentes fraccionarios con covariables. Recientemente, Krastaumber ha avanzado en la investigación para el puente clásico; sin embargo, queda como un problema interesante generalizar estos trabajos al puente browniano fraccionario.

En la sección 4.3, se proporciona una aproximación innovadora al uso de puentes estocásticos, como ruido en la modelización de datos funcionales. Aunque la propuesta no busca abordar la calibración completa del modelo COLFOR, plantea una técnica prometedora que amplía las posibilidades en el campo de modelos de lenguaje natural. Como trabajo futuro, se propone evaluar la aplicación del puente browniano fraccionario en escenarios más realistas y con conjuntos de datos extensos. Esta evaluación proporcionará una comprensión más profunda de las ventajas y limitaciones del enfoque, permitiendo así su refinamiento

y optimización para mejorar su rendimiento en situaciones prácticas. La investigación continua en este ámbito promete contribuir al avance de las estrategias utilizadas en modelos de lenguaje natural, explorando vías para una generación de texto más precisa y adaptable a diversas aplicaciones.

Un desafío futuro de interés radica en desarrollar una metodología inferencial para abordar ecuaciones diferenciales estocásticas multivariadas asociadas al puente browniano multivariado. Este trabajo constituye una extensión analógica de la hoja browniana, y su importancia radica en la capacidad para manejar ecuaciones diferenciales estocásticas que involucran múltiples variables. La complejidad inherente a este contexto requiere enfoques específicos que permitan obtener estimaciones robustas y precisas de los parámetros asociados al puente browniano multivariado. Investigaciones en este sentido no solo contribuirían al desarrollo teórico, sino que también abrirían nuevas posibilidades en la modelización de sistemas complejos y la comprensión de la dinámica estocástica en múltiples dimensiones.

En conclusión, los puentes estocásticos son una familia de procesos que modelan ruido cuando la volatilidad es pequeña en las fronteras y lejana de ellas, permitiendo una descripción más precisa de la variabilidad en diferentes segmentos del soporte. La comparación entre el puente browniano clásico y el puente browniano fraccionario destaca la flexibilidad que este último aporta, especialmente al considerar aumentos flexibles en la volatilidad. Esta característica es fundamental al modelar trayectorias entre dos puntos específicos, ya que proporciona una representación más rica y realista de la variación en el proceso. Además, la propuesta de utilizar el puente browniano fraccionario como ruido en datos funcionales abre nuevas perspectivas en la modelización de sistemas dinámicos, especialmente cuando la memoria a largo plazo es crucial. En resumen, la elección adecuada del puente estocástico según las características del fenómeno en estudio puede enriquecer significativamente la capacidad predictiva, descriptiva e inferencial del modelo.

6 Referencias

- [1] ARAGÓN URREGO, Daniel: Valoración de opciones americanas por el método de malla estocástica bajo movimiento Browniano fraccional del activo subyacente (American Option Pricing by the Stochastic Mesh Method Under Fractional Brownian Movement of the Underlying Asset). (2018)
- [2] BRAS, Pierre ; KOHATSU-HIGA, Arturo: Simulation of reflected Brownian motion on two dimensional wedges. In: *Stochastic Processes and their Applications* 156 (2023), S. 349-378
- [3] CARLSON, Max ; KIRBY, Robert M. ; SUNDAR, Hari: A scalable framework for solving fractional diffusion equations. In: *Proceedings of the 34th ACM International Conference on Supercomputing*, 2020, S. 1-11
- [4] CASELLA, George ; FERRÁNDIZ, Juan ; PEÑA, Daniel ; INSUA, David R. ; BERNARDO, José M ; GARCÍA-LÓPEZ, PA ; GONZÁLEZ, A ; BERGER, J ; DAWID, AP ; DICICCIO, Thomas J. u. a.: Statistical inference and Monte Carlo algorithms. In: *Test* 5 (1996), S. 249-344
- [5] CASTAÑEDA, Liliana B. ; ARUNACHALAM, Viswanathan ; DHARMARAJA, Selvamuthu: *Introduction to probability and stochastic processes with applications*. John Wiley & Sons, 2012
- [6] CHOW, Winston C.: Brownian bridge. In: *Wiley interdisciplinary reviews: computational statistics* 1 (2009), Nr. 3, S. 325-332
- [7] COOK, R D.: Envelope methods. In: *Wiley Interdisciplinary Reviews: Computational Statistics* 12 (2020), Nr. 2, S. e1484
- [8] DASGUPTA, Anirban: *Asymptotic theory of statistics and probability*. Bd. 180. Springer, 2008
- [9] DIEKER, Ton: *Simulation of fractional Brownian motion*, Masters Thesis, Department of Mathematical Sciences, University of Twente ..., Diss., 2004
- [10] EMBRECHTS, Paul ; MAEJIMA, Makoto: An introduction to the theory of self-similar stochastic processes. In: *International journal of modern physics B* 14 (2000), Nr. 12n13, S. 1399-1420
- [11] FRIEDRICH, Jan ; GALLON, Sebastian ; PUMIR, Alain ; GRAUER, Rainer: Stochastic interpolation of sparsely sampled time series via multipoint fractional Brownian bridges. In: *Physical Review Letters* 125 (2020), Nr. 17, S. 170602

-
- [12] GOLDSMITH, Jeff ; GREVEN, Sonja ; CRAINICEANU, CIPRIAN: Corrected confidence bands for functional data using principal components. In: *Biometrics* 69 (2013), Nr. 1, S. 41-51
- [13] GORGENS, Maik: Conditioning of Gaussian processes and a zero area Brownian bridge. In: *arXiv preprint arXiv:1302.4186* (2013)
- [14] GOSSET, William S.: William Sealy Gosset. In: *Biographical Encyclopedia of Mathematicians* 1 (1908), S. 239
- [15] HOLLANDER, Myles ; WOLFE, Douglas A. ; CHICKEN, Eric: *Nonparametric statistical methods*. John Wiley & Sons, 2013
- [16] KOKOSZKA, Piotr ; REIMHERR, Matthew: *Introduction to functional data analysis*. CRC press, 2017
- [17] KOLMOGOROV, Andrei N.: Wiener'sche spiralen und einige andere interessante kurven in hilbertscen raum, cr (doklady). In: *Acad. Sci. URSS (NS)* 26 (1940), S. 115-118
- [18] KRANSTAUBER, Bart ; KAYS, Roland ; LAPOINT, Scott D. ; WIKELSKI, Martin ; SAFI, Kamran: A dynamic Brownian bridge movement model to estimate utilization distributions for heterogeneous animal movement. In: *Journal of Animal Ecology* 81 (2012), Nr. 4, S. 738-746
- [19] KRANSTAUBER, Bart ; KAYS, Roland ; LAPOINT, Scott D. ; WIKELSKI, Martin ; SAFI, Kamran: A dynamic Brownian bridge movement model to estimate utilization distributions for heterogeneous animal movement. In: *Journal of Animal Ecology* 81 (2012), Nr. 4, S. 738-746
- [20] LAMPERTI, John: Semi-stable stochastic processes. In: *Transactions of the American mathematical Society* 104 (1962), Nr. 1, S. 62-78
- [21] MALLIAVIN, Paul: *Stochastic analysis*. Bd. 313. Springer, 2015
- [22] MANSUY, Roger ; YOR, Marc: *Aspects of Brownian motion*. Springer Science & Business Media, 2008
- [23] NAPIERALA, Matthew A.: What is the Bonferroni correction? In: *Aaos Now* (2012), S. 40-41
- [24] NUALART, David: *The Malliavin calculus and related topics*. Bd. 1995. Springer, 2006
- [25] ØKSENDAL, Bernt ; ØKSENDAL, Bernt: *Stochastic differential equations*. Springer, 2003
- [26] ÖZAK, Myriam M. n. ; CASTAÑEDA, Liliana B.: *Introducción a la teoría avanzada de la probabilidad*. Bd. 2. Univ. Nacional de Colombia, 2002
- [27] RAJU, Tonse N.: William Sealy Gosset and William A. Silverman: two "students" of science. In: *Pediatrics* 116 (2005), Nr. 3, S. 732-735
- [28] RINCÓN, Luis: *Introducción a los procesos estocásticos*. UNAM, Facultad de Ciencias, 2012

-
- [29] RINCÓN, Luis: *Introducción a la probabilidad*. (2014)
- [30] ROSTEK, S ; SCHÖBEL, R: A note on the use of fractional Brownian motion for financial modeling. In: *Economic Modelling* 30 (2013), S. 30–35
- [31] SRIVASTAVA, Muni S.: Multivariate theory for analyzing high dimensional data. In: *Journal of the Japan Statistical Society* 37 (2007), Nr. 1, S. 53–86
- [32] SURYAWAN, Herry P. ; GUNARSO, Bobby: Self-intersection local times of generalized mixed fractional Brownian motion as white noise distributions. In: *Journal of Physics: Conference Series* Bd. 855 IOP Publishing, 2017, S. 012050
- [33] TAQQU, Murad: Weak convergence to fractional Brownian motion and to the Rosenblatt process. In: *Advances in Applied Probability* 7 (1975), Nr. 2, S. 249–249
- [34] WANG, Jian ; LIN, Dongding ; LI, Wenjie: Dialogue Planning via Brownian Bridge Stochastic Process for Goal-directed Proactive Dialogue. In: *arXiv preprint arXiv:2305.05290* (2023)
- [35] WANG, Shiyan ; RAMKRISHNA, Doraiswami ; NARSIMHAN, Vivek: Exact sampling of polymer conformations using Brownian bridges. In: *The Journal of Chemical Physics* 153 (2020), Nr. 3, S. 034901
- [36] WASSERMAN, Larry: *All of nonparametric statistics*. Springer Science & Business Media, 2006
- [37] WEI, Bo-Cheng: *Exponential family nonlinear models*. Bd. 1. Springer, 1998
- [38] XU, Mengjia: Understanding graph embedding methods and their applications. In: *SIAM Review* 63 (2021), Nr. 4, S. 825–853
- [39] XU, Mengjia: Understanding graph embedding methods and their applications. In: *SIAM Review* 63 (2021), Nr. 4, S. 825–853
- [40] YERLIKAYA-ÖZKURT, Fatma ; VARDAR-ACAR, Ceren ; YOLCU-OKUR, Yeliz ; WEBER, G-W: Estimation of the Hurst parameter for fractional Brownian motion using the CMARS method. In: *Journal of Computational and Applied Mathematics* 259 (2014), S. 843–850
- [41] YUAN, Chenggui ; MAO, Xuerong: Convergence of the Euler–Maruyama method for stochastic differential equations with Markovian switching. In: *Mathematics and Computers in Simulation* 64 (2004), Nr. 2, S. 223–235
- [42] ZHU, Chenyao ; LUO, Lan ; LI, Rui ; GUO, Junhui ; WANG, Qining: Wearable Motion Analysis System for Thoracic Spine Mobility with Inertial Sensors. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* (2024)