



Pronóstico Probabilístico Basado en Predicción Conformal para Series Temporales: Comparación con Bootstrapping y DeepAR

Pedro José Leal Mesa

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Estadística
Bogotá D.C., Colombia
2026

Pronóstico Probabilístico Basado en Predicción Conformal para Series Temporales: Comparación con Bootstrapping y DeepAR

Pedro José Leal Mesa

Tesis presentada como requisito parcial para optar al título de:

Magíster en Ciencias - Estadística

Director:
Ph.D. Mario E. Arrieta-Prieto

Línea de Investigación:
Análisis de Series de Tiempo y Predicción Conformal

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Estadística
Bogotá D.C., Colombia
2026

Dedicatoria

A mi papá Antonio Leal, a mis compañeros y a mi director por su guía y apoyo constante en este proceso de formación.

“Essentially, all models are wrong, but some are useful.”

— George E. P. Box

Agradecimientos

Al Ph.D. Mario E. Arrieta-Prieto, director de esta tesis, por su invaluable orientación y paciencia. Quiero agradecerle especialmente por haberme incentivado a salir de mi zona de confort y presentar los resultados preliminares de este trabajo en el escenario internacional.

Un agradecimiento profundo y especial por la oportunidad de participar en el **45th International Symposium on Forecasting** en China. Esta experiencia marcó un hito en mi carrera profesional y fue posible gracias a un esfuerzo colectivo que jamás olvidaré: a la Universidad Nacional de Colombia por brindarme los fondos institucionales, y muy especialmente a mis compañeros, amigos, conocidos y familiares, quienes con un cariño inmenso organizaron y participaron en una rifa para completar el dinero necesario para este viaje. Su solidaridad fue el motor que me llevó al otro lado del mundo.

A mis compañeros de la Maestría en Estadística, con quienes compartí este camino académico, por las discusiones técnicas y el apoyo moral en los momentos de mayor reto.

A mi padre, Antonio Leal, por ser mi apoyo incondicional y por creer en este proyecto desde el primer día. Todo este esfuerzo es también suyo.

Con gratitud,
Pedro José Leal Mesa

Contents

Agradecimientos	vii
1 Introducción y Objetivos	1
1.1 Introducción	1
1.2 Planteamiento del problema	2
1.3 Justificación	3
1.4 Objetivos	4
1.4.1 Objetivo General	4
1.4.2 Objetivos Específicos	4
1.5 Estructura de la tesis	4
2 Marco Teórico	6
2.1 Pronóstico Probabilístico	6
2.1.1 Definición y Objetivos	6
2.1.2 Ventajas del Pronóstico Probabilístico	7
2.1.3 Representación de Distribuciones Predictivas	9
2.2 Métricas para Evaluación de Pronósticos Probabilísticos	9
2.2.1 Reglas de Puntuación Propias	9
2.2.2 Continuous Ranked Probability Score (CRPS)	11
2.2.3 Expected Continuous Ranked Probability Score (ECRPS)	12
2.3 Predicción Conformal por Intervalos: El Enfoque IIE	13
2.3.1 El Concepto de No-Conformidad	13
2.3.2 Protocolo de Construcción de Intervalos	14
2.4 Robustez ante la No-Intercambiabilidad: Aproximación de Barber	14
2.4.1 El Gap de Cobertura y Variación Total	15
2.4.2 Cuantiles Pesados y Decaimiento Temporal	15
2.5 Sistemas de Predicción Conformal (CPS): De Intervalos a Densidades	16
2.5.1 Formalización de la RPD y el Suavizado (τ)	16

2.5.2	La Máquina de Predicción de Mínimos Cuadrados (LSPM)	16
2.6	Sistemas de Predicción Conformal de Mondrian (MCPS)	17
2.6.1	Origen y Motivación: Validez Marginal vs. Condicional	17
2.6.2	La Taxonomía de Mondrian (κ)	18
2.6.3	Integración del Algoritmo MCPS	18
2.6.4	MCPS aplicado a la Máquina de Predicción de Mínimos Cuadrados (LSPM)	19
2.6.5	Garantías de Validez Condicional y Robustez	20
2.7	Sustento Teórico: Análisis de la Consistencia Universal de Vovk	20
2.7.1	Definición de Consistencia Universal	20
2.7.2	Mecanismo de la Demostración: El Enfoque de Histograma	21
2.7.3	La Distancia de Lévy y la Convergencia Débil	21
2.7.4	Implicaciones para el LSPM y la Eficiencia	22
2.8	Análisis Crítico sobre la Extensión a Series de Tiempo Ergódicas y Estacionarias	22
2.8.1	Hacia la superación del paradigma IID: El rol del Mixing	22
2.8.2	Mecanismos de Localización: Ventanas Móviles y Kernels	23
2.8.3	Estatus Teórico de la Extensión en la Presente Investigación	24
Bibliografía		25

List of Figures

2-1 Tipos de predicciones y su relación con la incertidumbre.	7
---	---

List of Tables

1 Introducción y Objetivos

En este primer capítulo se describe el contexto general de la investigación, enfocándose en la importancia de la cuantificación de la incertidumbre en el análisis de series temporales. Se presenta el planteamiento del problema, los retos actuales de la predicción probabilística y la justificación académica de este trabajo. Finalmente, se definen los objetivos general y específicos que guiarán el desarrollo de la tesis, se esbozan las hipótesis centrales y se presenta la estructura general del documento.

1.1 Introducción

La predicción es una tarea fundamental en diversas disciplinas, siendo de particular interés en el análisis de series de tiempo. Tradicionalmente, la literatura y la práctica se han centrado en estimaciones puntuales, como la media o el valor esperado de una variable futura. Sin embargo, este enfoque omite información crucial sobre la incertidumbre asociada a la predicción. Problemas clásicos, como la optimización de inventarios o la gestión de riesgos financieros, ilustran la necesidad de ir más allá de la media y considerar la distribución completa de los posibles resultados futuros para una toma de decisiones óptima.

La predicción probabilística aborda esta necesidad proporcionando no solo un valor central, sino una distribución de probabilidad o un conjunto de cuantiles sobre los valores futuros. En este contexto, surge la teoría de la *Predicción Conformal* (Conformal Prediction - CP) como un marco robusto para la construcción de intervalos y regiones de predicción con garantías de cobertura válidas bajo condiciones mínimas. A diferencia de otros métodos, la CP posee validez exacta y es aplicable a diversos modelos predictivos, lo que la convierte en una alternativa metodológica prometedora para el pronóstico probabilístico.

La relevancia comparativa de los métodos que se analizarán en este trabajo radica en sus características distintivas: mientras que la CP ofrece garantías teóricas de cobertura bajo

supuestos mínimos, los métodos de remuestreo como el *Bootstrapping* proporcionan flexibilidad no paramétrica para capturar la estructura de dependencia temporal, y los modelos de aprendizaje profundo como *DeepAR* (Salinas et al. 2020) destacan por su capacidad de modelar dependencias temporales complejas en múltiples series simultáneamente. Esta diversidad metodológica permite una evaluación integral de las fortalezas y limitaciones de cada enfoque.

No obstante, la aplicación de CP a series de tiempo presenta desafíos significativos debido a que los supuestos de intercambiabilidad de los datos suelen verse vulnerados por la dependencia temporal y la presencia de derivas distribucionales. Este trabajo busca explorar cómo adaptar estos métodos mediante mecanismos de ponderación temporal y contrastarlos frente a técnicas establecidas, contribuyendo así a cerrar una brecha importante en la literatura: la escasa aplicación y evaluación sistemática de CP con adaptaciones para no intercambiabilidad en escenarios de pronóstico de series temporales reales.

1.2 Planteamiento del problema

La predicción en series de tiempo enfrenta tres retos fundamentales que motivan esta investigación:

1. **Dependencia Temporal:** La naturaleza secuencial de los datos viola el supuesto de independencia fundamental en muchos métodos estadísticos estándar.
2. **No Estacionariedad:** Las características de los datos (tendencias, estacionalidad o cambios estructurales) cambian con el tiempo, lo que complica el modelado a largo plazo.
3. **Calibración de la Incertidumbre:** Generar pronósticos probabilísticos que sean simultáneamente precisos (calibrados) e informativos (eficientes) es una tarea compleja.

Los métodos actuales presentan limitaciones que este estudio busca contrastar. Por un lado, los modelos clásicos de tipo autorregresivo dependen fuertemente de supuestos sobre la estructura del proceso y la distribución del ruido (usualmente Gaussiano). Por otro lado, técnicas de remuestreo como el *Bootstrapping*, aunque flexibles, pueden tener dificultades cerca de límites de no estacionariedad. Finalmente, modelos modernos de aprendizaje profundo como *DeepAR* (Salinas et al. 2020) ofrecen un alto poder predictivo,

pero a menudo operan como “cajas negras” y carecen de las garantías teóricas formales de cobertura que ofrece el enfoque conformal.

El problema central de este trabajo se resume en la siguiente pregunta: **¿Cómo adaptar y evaluar un modelo de predicción probabilística basado en la teoría conformal para series de tiempo que considere la dependencia temporal, y cuál es su desempeño comparativo frente a métodos de referencia como Bootstrapping y DeepAR?**

1.3 Justificación

La cuantificación precisa de la incertidumbre es esencial para la toma de decisiones informada en áreas como la economía, meteorología y planificación de recursos. Adaptar la Predicción Conformal al dominio temporal representa un avance metodológico importante, ya que ofrece garantías de cobertura en muestra finita y una aplicabilidad general (libre de distribución).

A pesar de los desarrollos recientes en predicción conformal, existe un vacío notable en la literatura respecto a su aplicación sistemática en series temporales con características de no intercambiabilidad. La mayoría de los trabajos se han centrado en la construcción de intervalos de predicción, relegando el desarrollo de métodos para predicción distribucional completa. Además, la comparación rigurosa de CP adaptada mediante ponderación temporal frente a métodos establecidos (Bootstrapping) y estado del arte (DeepAR) utilizando métricas probabilísticas apropiadas como el *Continuous Ranked Probability Score* (CRPS) ha sido escasamente explorada.

Este trabajo contribuirá a cerrar estas brechas metodológicas al:

- Desarrollar una adaptación práctica de CP para series temporales que incorpore ponderación para manejar deriva distribucional.
- Proporcionar una evaluación comparativa sistemática en escenarios simulados y reales.
- Ofrecer evidencia empírica sobre las condiciones bajo las cuales CP puede ser competitiva o superior a métodos establecidos.

La comparación sistemática con métodos estándar (Bootstrapping) y estado del arte

(*DeepAR*) proporcionará una perspectiva clara sobre las fortalezas y debilidades del enfoque conformal en escenarios reales y simulados, facilitando la selección informada de métodos de pronóstico probabilístico en aplicaciones prácticas.

1.4 Objetivos

1.4.1 Objetivo General

Formular un modelo fundamentado en la teoría conformal para realizar predicciones probabilísticas en el contexto de series de tiempo, evaluando su desempeño mediante una comparación con modelos establecidos como *Bootstrapping* y *DeepAR*.

1.4.2 Objetivos Específicos

- Desarrollar un modelo de predicción probabilística que integre la teoría de predicción conformal con técnicas de modelado de series temporales, incorporando mecanismos de ponderación para abordar la no intercambiabilidad.
- Diseñar un entorno de simulación para evaluar el comportamiento del modelo propuesto bajo diversos escenarios de estructuras temporales y condiciones de ruido.
- Comparar el desempeño del modelo propuesto frente a modelos de referencia (*DeepAR* y *Bootstrapping*) utilizando métricas específicas de predicción probabilística como el *Continuous Ranked Probability Score* (CRPS) y análisis de calibración.
- Implementar la metodología en un caso de estudio con datos reales para cuantificar su relevancia práctica en aplicaciones de pronóstico.

1.5 Estructura de la tesis

El presente documento se organiza de la siguiente manera:

Capítulo 2: Marco Teórico. Se presenta una introducción al pronóstico probabilístico, se explican las principales métricas para evaluar el desempeño de los modelos (incluyendo

CRPS, calibración y nitidez), se desarrollan los fundamentos de los sistemas de predicción conformal, y se incluye una discusión sobre la extensión de la demostración de consistencia universal en escenarios estacionarios y ergódicos.

Capítulo 3: Metodología de Simulación. Se detalla el diseño del entorno de simulación, incluyendo la definición de los procesos generadores de datos, la formulación de los modelos predictivos (CP con ponderación, Bootstrapping y DeepAR), los cambios necesarios para su aplicación en este trabajo, y las extensiones de las simulaciones para diferentes escenarios.

Capítulo 4: Resultados de Simulación. Se presentan y analizan los resultados de las simulaciones diseñadas en el Capítulo 3, incluyendo comparaciones de ECRPS, análisis de cobertura, evaluación de calibración y pruebas de significancia estadística.

Capítulo 5: Aplicaciones a Datos Reales. Se implementan los métodos desarrollados en casos de estudio con datos reales, se presentan los resultados obtenidos y se discute su relevancia práctica.

Capítulo 6: Conclusiones. Se sintetizan los hallazgos principales del trabajo, se discuten las limitaciones del estudio, y se proponen direcciones para investigación futura.

2 Marco Teórico

En este capítulo se presentan los fundamentos teóricos que sustentan el desarrollo de esta investigación. Se inicia con una introducción al pronóstico probabilístico y su importancia en el análisis de series temporales, seguido de una discusión detallada sobre las métricas utilizadas para evaluar el desempeño predictivo. Posteriormente, se desarrollan los conceptos fundamentales de la predicción conformal y sus adaptaciones para series de tiempo.

2.1 Pronóstico Probabilístico

El pronóstico probabilístico representa un cambio de paradigma fundamental en la predicción estadística, pasando de estimaciones puntuales a distribuciones de probabilidad completas sobre cantidades futuras de interés. A diferencia de las predicciones puntuales tradicionales, que proporcionan únicamente un valor esperado o una estimación central, el pronóstico probabilístico cuantifica la incertidumbre asociada a la predicción mediante la especificación de una distribución predictiva completa.

2.1.1 Definición y Objetivos

Formalmente, sea Y_{t+h} una variable aleatoria que representa el valor de una serie temporal en el tiempo $t + h$, donde $h > 0$ denota el horizonte de predicción. Un pronóstico probabilístico es una distribución de probabilidad $F_{t+h|t}$ que caracteriza la incertidumbre sobre Y_{t+h} dado el conjunto de información disponible hasta el tiempo t , denotado por \mathcal{F}_t .

Gneiting y Raftery explican que el objetivo fundamental del pronóstico probabilístico es maximizar la nitidez de las distribuciones predictivas sujeto a calibración. Estos dos conceptos son fundamentales para entender la calidad de un pronóstico probabilístico:

- **Calibración:** Se refiere a la consistencia estadística entre las distribuciones pre-

dictivas y las observaciones. Una predicción está calibrada si las realizaciones son estadísticamente indistinguibles de muestras aleatorias de las distribuciones predictivas.

- **Nitidez:** Se refiere a la concentración de las distribuciones predictivas y es una propiedad exclusiva de los pronósticos. Cuanto más concentradas sean las distribuciones predictivas, mejor, siempre que se mantenga la calibración.

La Figura 2-1 ilustra la diferencia entre una predicción puntual, un intervalo de predicción y una distribución predictiva completa.

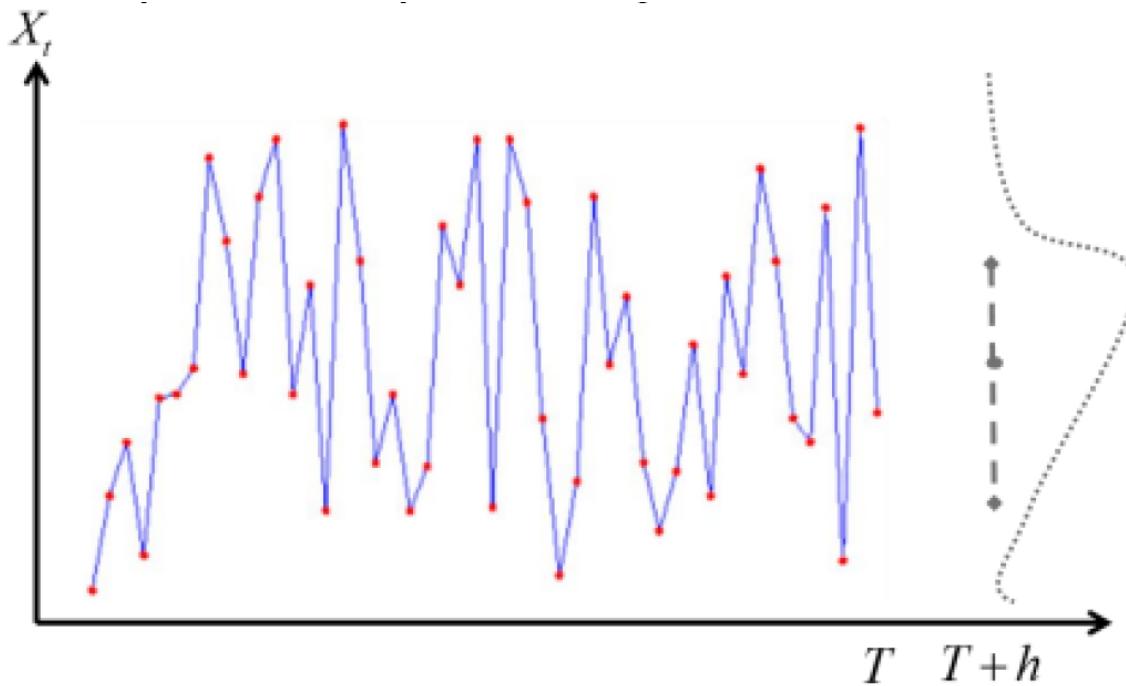


Figure 2-1: Tipos de predicciones y su relación con la incertidumbre.

2.1.2 Ventajas del Pronóstico Probabilístico

El pronóstico probabilístico ofrece múltiples ventajas sobre las predicciones puntuales tradicionales que justifican su adopción creciente en diversas aplicaciones:

1. **Cuantificación Completa de Incertidumbre:** A diferencia de las predicciones puntuales, que proporcionan únicamente un valor esperado, el pronóstico probabilístico caracteriza la incertidumbre de manera exhaustiva mediante distribuciones

de probabilidad. Esto permite a los tomadores de decisiones evaluar tanto la magnitud esperada de un evento como la variabilidad asociada, facilitando una comprensión más profunda de los riesgos y oportunidades.

2. **Soporte para Decisiones Óptimas:** En contextos donde las decisiones deben tomarse bajo incertidumbre, como la gestión de inventarios, la planificación de recursos energéticos, o la asignación de capital, las distribuciones predictivas completas son esenciales. Permiten la optimización de funciones de utilidad esperada y la implementación de estrategias que consideren explícitamente el trade-off entre riesgo y recompensa.
3. **Evaluación de Eventos Extremos:** Las predicciones puntuales son inherente-mente limitadas para caracterizar eventos raros o de cola. El pronóstico proba-bilístico, en cambio, permite estimar probabilidades de ocurrencia de eventos ex-tremos, información crucial para la gestión de riesgos financieros, la planificación de infraestructura ante eventos climáticos severos, y la evaluación de escenarios de baja probabilidad pero alto impacto.
4. **Agregación Temporal y Espacial Consistente:** Las distribuciones predictivas pueden agregarse de manera coherente a través de diferentes horizontes temporales o unidades espaciales. Por ejemplo, es posible derivar pronósticos mensuales a partir de pronósticos diarios manteniendo la consistencia probabilística, o agregar pronósticos locales en pronósticos regionales preservando las propiedades estadísticas apropiadas.
5. **Flexibilidad en la Comunicación de Incertidumbre:** Las distribuciones predic-tivas permiten múltiples formas de comunicación adaptadas a diferentes audiencias y necesidades: intervalos de predicción con diferentes niveles de confianza, probabili-dades de excedencia de umbrales críticos, percentiles específicos, o visualizaciones complejas de la distribución mediante gráficos de densidad o fan charts.
6. **Base para Post-procesamiento y Combinación:** Los pronósticos probabilísticos proporcionan la base natural para técnicas avanzadas de mejoramiento predictivo, como la combinación bayesiana de modelos, el post-procesamiento estadístico de ensembles numéricos, o la integración de múltiples fuentes de información mediante métodos de agregación de distribuciones.

Estas ventajas han motivado la transición hacia pronósticos probabilísticos en campos tan diversos como meteorología, finanzas, energía, epidemiología y gestión de cadenas de

suministro, donde la cuantificación rigurosa de la incertidumbre es fundamental para la toma de decisiones efectiva.

2.1.3 Representación de Distribuciones Predictivas

Las distribuciones predictivas pueden representarse de diversas formas, dependiendo del contexto y los requisitos de la aplicación:

- **Función de Distribución Acumulada (FDA):** $F_{t+h|t}(y) = P(Y_{t+h} \leq y | \mathcal{F}_t)$
- **Función de Densidad de Probabilidad:** $f_{t+h|t}(y)$ cuando existe
- **Cuantiles:** $q_\alpha = F_{t+h|t}^{-1}(\alpha)$ para $\alpha \in (0, 1)$
- **Intervalos de Predicción:** $[q_{\alpha/2}, q_{1-\alpha/2}]$ con cobertura nominal $(1 - \alpha) \times 100\%$
- **Conjuntos de Predicción:** Regiones de confianza más generales en espacios multidimensionales

En el contexto de series temporales, es particularmente importante considerar la dependencia temporal al construir distribuciones predictivas multivariadas para múltiples horizontes de predicción simultáneamente.

2.2 Métricas para Evaluación de Pronósticos Probabilísticos

La evaluación rigurosa del desempeño predictivo es fundamental para comparar metodologías de pronóstico y guiar mejoras en los modelos. En el contexto de pronósticos probabilísticos, las métricas de evaluación deben considerar tanto la calibración como la nitidez de las distribuciones predictivas.

2.2.1 Reglas de Puntuación Propias

Una *regla de puntuación* (scoring rule) es una función $S : \mathcal{F} \times \mathbb{R} \rightarrow \mathbb{R} \cup \{\infty\}$ que asigna una penalización numérica $S(F, y)$ a cada par formado por una distribución predictiva F

y una observación realizada y . En nuestra notación, valores más bajos de la puntuación indican mejor desempeño predictivo.

Propriety y Strict Propriety

La *propriety* es una característica fundamental que debe satisfacer toda métrica de evaluación de pronósticos probabilísticos para garantizar que incentive predicciones honestas y bien calibradas.

Definición (Regla de Puntuación Proper): Una regla de puntuación S es *proper* relativa a una clase \mathcal{F} de distribuciones de probabilidad si

$$\mathbb{E}_G[S(G, Y)] \leq \mathbb{E}_G[S(F, Y)] \quad (2-1)$$

para todas las distribuciones $F, G \in \mathcal{F}$, donde $Y \sim G$. Es decir, la puntuación esperada se minimiza cuando el pronóstico emitido es la verdadera distribución generadora de los datos.

Definición (Regla de Puntuación Strictly Proper): La regla de puntuación S es *strictly proper* si la desigualdad en (2-1) se cumple con igualdad únicamente cuando $F = G$.

La importancia de la *propriety* radica en que establece un principio de alineación de incentivos: si un pronosticador desea minimizar su puntuación esperada, su mejor estrategia es reportar sinceramente su verdadera distribución predictiva, sin distorsiones ni manipulaciones. Las reglas *improper* pueden incentivar al pronosticador a reportar una distribución diferente de su verdadera creencia, lo que conduce a evaluaciones sesgadas y conclusiones erróneas sobre el desempeño relativo de diferentes métodos.

La *strict propriety* es una condición más fuerte que garantiza que la verdadera distribución es el único minimizador de la puntuación esperada. Esta característica adicional es deseable porque proporciona un criterio único y bien definido para identificar la mejor predicción posible, facilita la identificación de diferencias entre métodos de pronóstico, y ofrece mayor poder discriminatorio en comparaciones empíricas.

2.2.2 Continuous Ranked Probability Score (CRPS)

El *Continuous Ranked Probability Score* (CRPS) es una de las reglas de puntuación estrictamente propias más utilizadas para evaluar pronósticos probabilísticos de variables continuas. Su popularidad se debe a su sólida fundamentación teórica, su interpretabilidad intuitiva, y su capacidad para evaluar simultáneamente calibración y nitidez.

Definiciones y Representaciones

El CRPS admite varias representaciones matemáticas equivalentes, cada una con sus propias ventajas conceptuales y computacionales.

Representación integral: La definición original del CRPS está dada por:

$$\text{CRPS}(F, y) = \int_{-\infty}^{\infty} (F(x) - \mathbb{1}\{y \leq x\})^2 dx \quad (2-2)$$

donde F es la función de distribución acumulada (FDA) de la distribución predictiva, y es la observación realizada, y $\mathbb{1}\{\cdot\}$ denota la función indicadora. Esta representación muestra que el CRPS mide el área entre la FDA predictiva y la FDA de la observación.

Representación basada en esperanzas: Una forma alternativa, más conveniente para cálculos, está dada por:

$$\text{CRPS}(F, y) = \mathbb{E}_F|X - y| - \frac{1}{2}\mathbb{E}_F|X - X'| \quad (2-3)$$

donde X y X' son variables aleatorias independientes con distribución F . Esta representación revela una interpretación intuitiva del CRPS: el primer término mide la distancia esperada entre la predicción y la observación, mientras que el segundo término penaliza la dispersión de la distribución predictiva.

Es importante notar que el CRPS es una métrica diseñada para comparar una distribución predictiva completa F contra un punto observado y . Esta es una distinción fundamental respecto al ECRPS, que se discutirá posteriormente.

Propiedades del CRPS

El CRPS posee varias propiedades deseables que explican su amplia adopción en la literatura de pronósticos:

1. **Strictly proper:** El CRPS es *strictly proper* relativo a la clase de todas las distribuciones de probabilidad en \mathbb{R} con primer momento finito. Esto garantiza que la puntuación esperada se minimiza únicamente cuando el pronóstico emitido coincide con la verdadera distribución generadora.
2. **Unidades consistentes:** El CRPS se expresa en las mismas unidades que la variable pronosticada, lo que facilita su interpretación práctica.
3. **Reducción al error absoluto:** Cuando F es una distribución degenerada (predicción puntual), el CRPS se reduce al error absoluto:

$$\text{CRPS}(\delta_x, y) = |x - y| \quad (2-4)$$

donde δ_x denota la medida de Dirac concentrada en x . Esta propiedad permite comparar directamente pronósticos puntuales y probabilísticos bajo un marco unificado.

4. **Sensibilidad dual:** El CRPS evalúa simultáneamente dos aspectos críticos del desempeño predictivo: calibración, mediante el término que penaliza desviaciones entre predicción y observación, y nitidez, mediante el término que premia distribuciones predictivas concentradas.

2.2.3 Expected Continuous Ranked Probability Score (ECRPS)

En aplicaciones prácticas de evaluación de pronósticos, típicamente disponemos de una secuencia de n pares pronóstico-observación: $(F_1, y_1), (F_2, y_2), \dots, (F_n, y_n)$, donde cada F_i representa una distribución predictiva y cada y_i es la observación correspondiente. El desempeño predictivo global se cuantifica mediante el *Expected Continuous Ranked Probability Score* (ECRPS), definido como la media aritmética de los CRPS individuales:

$$\text{ECRPS} = \frac{1}{n} \sum_{i=1}^n \text{CRPS}(F_i, y_i) \quad (2-5)$$

El ECRPS hereda todas las propiedades deseables del CRPS (*strict property*, unidades interpetables, sensibilidad dual a calibración y nitidez) y proporciona un resumen numérico único del desempeño predictivo sobre todo el conjunto de evaluación.

A diferencia del CRPS, que compara una distribución contra un punto observado, el ECRPS puede interpretarse como una métrica para comparar dos distribuciones cuando consideramos la distribución empírica de las observaciones. Si tenemos dos métodos de pronóstico que producen secuencias de distribuciones (F_1, \dots, F_n) y (G_1, \dots, G_n) respectivamente, sus ECRPS correspondientes nos permiten comparar el desempeño predictivo global de ambos métodos considerando toda la secuencia de pronósticos y observaciones.

2.3 Predicción Conformal por Intervalos: El Enfoque IIE

La predicción conformal clásica, introducida por Vovk et al. Vovk, Gammerman, and Shafer 2005, se fundamenta en la capacidad de generar conjuntos de predicción Γ^ϵ que garantizan una cobertura de confianza exacta para cualquier nivel de significancia $\epsilon \in (0, 1)$. A diferencia de los métodos estadísticos tradicionales que dependen de la asintótica (grandes muestras), la predicción conformal es válida para muestras finitas, siempre que se cumpla el supuesto de intercambiabilidad de los datos.

2.3.1 El Concepto de No-Conformidad

El núcleo de esta metodología es la *medida de no-conformidad* (NCM, por sus siglas en inglés). Una NCM es una función $A(B, z)$ que cuantifica el grado de "extrañeza" de un ejemplo z en relación con un multiconjunto (o *bag*) de ejemplos B . En el contexto de regresión, donde $z = (x, y)$, la medida de no-conformidad más común es el error absoluto de predicción, definido como:

$$\alpha_i = |y_i - \hat{y}_i| \quad (2-6)$$

donde \hat{y}_i es la estimación producida por un algoritmo de aprendizaje subyacente (denominado *underlying algorithm*). Es importante subrayar que la predicción conformal es agnóstica al modelo: puede envolver desde una regresión lineal simple hasta redes neuronales profundas, transformando sus predicciones puntuales en intervalos con validez estadística.

2.3.2 Protocolo de Construcción de Intervalos

Para construir un intervalo de predicción para un nuevo objeto x_n basado en un conjunto de entrenamiento z_1, \dots, z_{n-1} , el método IIE (Inducida por Errores) sigue un proceso de prueba de hipótesis inversa. Para cada valor potencial $y \in \mathbb{R}$:

1. **Aumentación del Conjunto:** Se asume hipotéticamente que la verdadera etiqueta de x_n es y , formando el conjunto aumentado $z_1, z_2, \dots, z_{n-1}, z_n$, donde $z_n = (x_n, y)$.
2. **Cálculo de Puntajes:** Se calculan los puntajes de no-conformidad $\alpha_1, \dots, \alpha_n$ para todos los elementos, incluyendo el ejemplo hipotético.
3. **Derivación del p-valor:** Se calcula la proporción de ejemplos que son "al menos tan extraños" como el nuevo ejemplo z_n :

$$p(y) = \frac{|\{i = 1, \dots, n : \alpha_i \geq \alpha_n\}|}{n} \quad (2-7)$$

4. **Inversión de la Región de Aceptación:** El intervalo de predicción $\Gamma^{1-\epsilon}$ se define como el conjunto de todos los valores y que no pueden ser rechazados al nivel de significancia ϵ :

$$\Gamma^\epsilon(x_1, y_1, \dots, x_{n-1}, y_{n-1}, x_n) = \{y \in \mathbb{R} : p(y) > \epsilon\} \quad (2-8)$$

Este procedimiento garantiza que $P(y_n \notin \Gamma^\epsilon) \leq \epsilon$. Si los puntajes α_i tienen una distribución continua (sin empates), la probabilidad de error es exactamente ϵ [Vovk, Gammerman, and Shafer 2005](#).

2.4 Robustez ante la No-Intercambiabilidad: Aproximación de Barber

Uno de los desafíos críticos en el análisis de series temporales es que el supuesto de intercambiabilidad rara vez se sostiene. Fenómenos como la autocorrelación, la heterocedasticidad y la deriva de parámetros (drift) invalidan la asunción de que el pasado y el futuro son estadísticamente idénticos. Barber et al. Barber et al. [2023](#) proponen una extensión fundamental para estos escenarios.

2.4.1 El Gap de Cobertura y Variación Total

Barber et al. formalizan la degradación de la validez conformal mediante el uso de la *Distancia de Variación Total* (d_{TV}). Si la distribución de los datos cambia en el tiempo, existe una brecha de cobertura (*coverage gap*). El teorema principal de Barber establece que la pérdida de cobertura está acotada por la suma de las distancias entre la distribución de los datos de entrenamiento y la distribución del dato de prueba:

$$\text{Error de Cobertura} \leq \epsilon + \sum_{i=1}^n w_i d_{TV}(Z_i, Z_{n+1}) \quad (2-9)$$

2.4.2 Cuantiles Pesados y Decaimiento Temporal

Para contrarrestar este efecto en series de tiempo, Barber et al. introducen los *Weighted Conformal Predictors*. En lugar de asignar un peso uniforme de $1/n$ a cada residuo histórico, se asignan pesos w_i que reflejan la relevancia del dato. En series no estacionarias, los datos más recientes son mejores predictores del futuro.

Se define comúnmente un decaimiento geométrico para los pesos:

$$w_i = \rho^{n-i}, \quad \rho \in (0, 1) \quad (2-10)$$

donde un ρ cercano a 1 asume una estabilidad lenta, mientras que un ρ menor reacciona rápidamente a cambios estructurales. El p-valor pesado se calcula como una suma ponderada de funciones indicadoras:

$$p^y = \frac{\sum_{i=1}^{n-1} w_i \mathbb{1}_{\alpha_i \geq \alpha_n} + w_n}{\sum_{j=1}^n w_j} \quad (2-11)$$

Este enfoque permite que la predicción conformal sea "adaptativa", manteniendo la cobertura cercana al nivel nominal incluso cuando la serie temporal experimenta cambios súbitos en su media o varianza Barber et al. 2023.

2.5 Sistemas de Predicción Conformal (CPS): De Intervalos a Densidades

El Capítulo 7 de la obra de Vovk Vovk, Gammerman, and Shafer 2005 marca la transición de la predicción de conjuntos a la predicción de distribuciones completas. Un *Sistema de Predicción Conformal* (CPS) no entrega un rango, sino una *Distribución Predictiva Aleatorizada* (RPD), denotada como $\Pi_n(y, \tau)$, que representa la probabilidad de que la verdadera etiqueta sea menor o igual a y .

2.5.1 Formalización de la RPD y el Suavizado (τ)

Para asegurar que la distribución resultante sea continua y cumpla con las propiedades de una FDA (Función de Distribución Acumulada), se introduce una variable de suavizado $\tau \sim U(0, 1)$. La función Π se define como:

$$\Pi_n(y, \tau) := \frac{|\{i : \alpha_i < \alpha_n^y\}| + \tau |\{i : \alpha_i = \alpha_n^y\}|}{n} \quad (2-12)$$

Es vital notar que aquí α_i son puntajes de *conformidad* (no de no-conformidad). Un ejemplo común en regresión es $\alpha_i = y_i - \hat{y}_i$. El uso de la variable τ garantiza la *calibración fuerte en probabilidad*: los valores de la RPD evaluados en la verdadera etiqueta son independientes y uniformes en $[0, 1]$, permitiendo una cuantificación exacta de la incertidumbre en cualquier punto de la distribución Vovk, Nouretdinov, et al. 2017.

2.5.2 La Máquina de Predicción de Mínimos Cuadrados (LSPM)

La *Least Squares Prediction Machine* (LSPM) es la aplicación primordial de los CPS al ámbito de la regresión. La LSPM utiliza la estructura de la regresión lineal para optimizar la eficiencia de la distribución predictiva.

Variantes de la LSPM

Vovk distingue tres formas de calcular los residuos dentro de una LSPM:

1. **LSPM Ordinaria:** Los puntajes son simplemente los residuos de entrenamiento.

Sin embargo, este enfoque tiende a ser demasiado optimista (sobreajuste), ya que el modelo ya ha ”visto” los datos de entrenamiento.

2. **LSPM Eliminada (Deleted):** Utiliza un esquema de validación cruzada interna (*leave-one-out*). Para cada dato i , se entrena un modelo omitiendo ese dato específico, asegurando que el residuo sea una medida honesta de la capacidad de generalización.
3. **LSPM Estudiantizada:** Es la variante más robusta y matemáticamente rigurosa. Ajusta cada residuo por su apalancamiento (*leverage*), h_i , proveniente de la diagonal de la matriz *hat*:

$$\alpha_i := \frac{y_i - \hat{y}_i}{\sigma\sqrt{1 - h_i}} \quad (2-13)$$

2.6 Sistemas de Predicción Conformal de Mondrian (MCPS)

A pesar de las sólidas garantías de validez marginal que ofrecen los Sistemas de Predicción Conformal (CPS) descritos en la sección 2.5, estos presentan una limitación teórica y práctica fundamental: la garantía de error es un promedio sobre todo el espacio de datos. Esto implica que el sistema puede ser extremadamente preciso en ciertas regiones del espacio de características y, simultáneamente, cometer errores sistemáticos en otras, siempre que el error global no supere el nivel ϵ . Los *Sistemas de Predicción Conformal de Mondrian* (MCPS, por sus siglas en inglés) introducen el concepto de *validez condicional por categorías*, permitiendo que la calibración se mantenga exacta dentro de subconjuntos específicos de los datos Vovk, Gammerman, and Shafer 2005.

2.6.1 Origen y Motivación: Validez Marginal vs. Condicional

El término ”Mondrian” hace referencia al pintor neerlandés Piet Mondrian, cuyas obras se caracterizan por particiones del lienzo en rectángulos de colores puros. De manera análoga, un sistema Mondriano partitiona el espacio de ejemplos Z en categorías mutuamente excluyentes.

La necesidad de este enfoque surge cuando existen grupos de datos con dificultades predictivas heterogéneas. Por ejemplo, en una serie temporal de demanda eléctrica, predecir

el consumo en un día festivo es intrínsecamente más difícil que en un día laboral. Un CPS global podría subestimar masivamente la incertidumbre en los días festivos, compensándola con una sobreestimación en los días laborales. El enfoque de Mondrian garantiza que la probabilidad de error sea exactamente ϵ tanto para los días laborales como para los festivos, de forma independiente Vovk, Nouretdinov, et al. 2017.

2.6.2 La Taxonomía de Mondrian (κ)

La base matemática de un MCPS es la *taxonomía*. Una taxonomía es una función medible $\kappa : \mathbb{N} \times (\mathbf{X} \times \mathbf{Y}) \rightarrow K$, donde K es un conjunto numerable de categorías. Para cada par de ejemplo (x_i, y_i) y su posición en la secuencia i , la taxonomía asigna una categoría κ_i .

Existen tres tipos principales de taxonomías aplicables a series temporales:

1. **Taxonomías de Objetos:** Dependen solo de las características x_i (ej. agrupar por niveles de volatilidad observada).
2. **Taxonomías de Etiquetas:** Dependen de la respuesta y_i . Esto da lugar a los *Label-Conditional Conformal Predictors*, vitales cuando el impacto de un error depende de la magnitud del valor (ej. errores en valores extremos son más costosos).
3. **Taxonomías Temporales:** Dependen del índice i . Este es el puente con el trabajo de Barber et al. Barber et al. 2023, donde la categoría de Mondrian puede ser una "ventana deslizante" de los datos más recientes para adaptarse a la no-intercambiabilidad.

2.6.3 Integración del Algoritmo MCPS

La integración de la lógica de Mondrian en un Sistema de Predicción Conformal se realiza modificando el cálculo del p-valor o de la RPD (Distribución Predictiva Aleatorizada). En lugar de comparar el puntaje del nuevo ejemplo α_n con todos los puntajes históricos, solo se compara con aquellos que pertenecen a su misma categoría.

Sea $\sigma = \{z_1, \dots, z_{n-1}\}$ el conjunto de entrenamiento y $z_n = (x_n, y)$ el ejemplo de prueba con etiqueta hipotética y . El proceso para generar la RPD de Mondrian Π_M es el siguiente:

1. Se identifica la categoría del nuevo ejemplo: $k = \kappa(n, (x_n, y))$.

2. Se filtran los índices de los ejemplos de entrenamiento que pertenecen a dicha categoría:

$$S_k = \{i \in \{1, \dots, n-1\} : \kappa(i, z_i) = k\} \quad (2-14)$$

3. Se calculan los puntajes de conformidad α_i solo para $i \in S_k \cup \{n\}$.

4. La RPD de Mondrian se define como:

$$\Pi_M(y, \tau) := \frac{|\{i \in S_k : \alpha_i < \alpha_n^y\}| + \tau |\{i \in S_k \cup \{n\} : \alpha_i = \alpha_n^y\}|}{|S_k| + 1} \quad (2-15)$$

El denominador $|S_k| + 1$ es clave: representa el tamaño de la "muestra local". Si una categoría tiene pocos ejemplos, la distribución predictiva será naturalmente más dispersa (reflejando mayor incertidumbre), mientras que categorías ricas en datos producirán densidades más nítidas Vovk, Gammerman, and Shafer 2005.

2.6.4 MCPS aplicado a la Máquina de Predicción de Mínimos Cuadrados (LSPM)

Cuando se integra Mondrian con la LSPM (sección 2.5), obtenemos un sistema capaz de generar densidades de probabilidad que se ajustan localmente. En este esquema, aunque el modelo de regresión puntual \hat{y} puede entrenarse con todos los datos para aprovechar la eficiencia global, el *ajuste de la escala de error* se realiza mediante Mondrian.

Si utilizamos residuos estudiantizados, la integración se define como:

$$\alpha_i = \frac{y_i - \hat{y}_i}{\hat{\sigma} \sqrt{1 - h_i}}, \quad \forall i \in S_{\kappa_n} \quad (2-16)$$

Esto permite que el sistema detecte que en ciertos regímenes (categorías), el modelo de mínimos cuadrados tiende a fallar más que en otros. El MCPS "estira" o "encoge" la distribución predictiva de la LSPM basándose únicamente en la evidencia histórica de esa categoría específica.

2.6.5 Garantías de Validez Condicional y Robustez

El resultado fundamental de los MCPS, demostrado por Vovk et al. Vovk, Gammerman, and Shafer 2005, es que para cada categoría $k \in K$:

$$P(y_n \notin \Gamma^\epsilon | \kappa_n = k) \leq \epsilon \quad (2-17)$$

Esta validez condicional es mucho más fuerte que la marginal. En el contexto de Barber et al. Barber et al. 2023, los MCPS ofrecen una defensa contra la no-intercambiabilidad si la taxonomía está bien diseñada para capturar la estructura de la deriva. Al segmentar los datos, el MCPS limita la propagación del "ruido" de ejemplos antiguos e irrelevantes hacia las predicciones actuales, asegurando que la densidad predictiva sea no solo estadísticamente válida, sino también contextualmente precisa para la toma de decisiones óptimas en cada estado del sistema.

2.7 Sustento Teórico: Análisis de la Consistencia Universal de Vovk

Para consolidar el marco teórico de esta investigación, es imperativo discutir el sustento matemático que garantiza que los Sistemas de Predicción Conformal (CPS) no solo son válidos en muestras finitas, sino también óptimos a medida que el volumen de datos aumenta. Este respaldo proviene de la demostración de la *consistencia universal* de Vovk Vovk 2019, formalizada en el Teorema 31 de su obra reciente.

2.7.1 Definición de Consistencia Universal

En el contexto de los CPS, la validez (propiedad R2) asegura que el sistema está calibrado independientemente de la distribución de los datos. Sin embargo, la validez por sí sola no garantiza que la distribución predictiva Π_n sea una buena aproximación a la verdadera distribución condicional de las etiquetas $P(y|x)$.

Vovk define un sistema predictivo como *universalmente consistente* si, para cualquier medida de probabilidad P (bajo el modelo IID) y para cualquier función continua acotada

f , se cumple que:

$$\int f d\Pi_n - \mathbb{E}_P(f|x_{n+1}) \rightarrow 0 \quad \text{en probabilidad cuando } n \rightarrow \infty \quad (2-18)$$

Esta propiedad implica que, asintóticamente, el CPS "encuentra" la verdadera distribución de probabilidad generadora de los datos, eliminando la incertidumbre epistémica conforme el tamaño de la muestra n tiende al infinito Vovk 2019.

2.7.2 Mecanismo de la Demostración: El Enfoque de Histograma

La prueba de Vovk sobre la existencia de un CPS universal se apoya en la construcción de un *Histogram Conformal Predictive System*. El argumento se divide en dos pilares fundamentales que vinculan la teoría de martingalas con la ley de los grandes números:

1. **Teorema de Convergencia de Martingalas de Lévy:** Vovk utiliza particiones anidadas del espacio de objetos X (celdas de histograma que se encogen conforme n crece). Según el teorema de Lévy, la esperanza condicional de la función sobre una celda que se reduce tiende al valor puntual de la esperanza condicional en el objeto de prueba x_{n+1} Vovk 2019.
2. **Ley de los Grandes Números (LGN):** Mientras las celdas se encogen para ganar resolución, el número de ejemplos dentro de cada celda debe tender a infinito ($nh_n \rightarrow \infty$, donde h_n es el ancho de la celda). Esto permite que la frecuencia empírica de las etiquetas dentro de la categoría de Mondrian converja a la esperanza real en esa región del espacio Vovk 2019.

2.7.3 La Distancia de Lévy y la Convergencia Débil

Un punto crítico de la demostración es el uso de la noción de Belyaev sobre secuencias de distribuciones que se aproximan débilmente. Vovk demuestra que bajo un CPS universal, la *Distancia de Lévy* entre la distribución predictiva conformal y la verdadera distribución condicional converge a cero en probabilidad Vovk 2019.

Este resultado es el que otorga rigor a la aplicación de CPS en problemas de alta criticidad, como el pronóstico de carga eléctrica o la gestión de riesgos financieros. Indica que el analista no tiene que elegir entre un modelo "seguro" (conformal) y un modelo "preciso"

(bayesiando/paramétrico); el CPS universal ofrece ambas ventajas simultáneamente:

- **A corto plazo:** Garantiza cobertura exacta mediante calibración fuerte.
- **A largo plazo:** Garantiza convergencia a la distribución real de los datos sin requerir asunciones paramétricas.

2.7.4 Implicaciones para el LSPM y la Eficiencia

Aunque el modelo de mínimos cuadrados (LSPM) estudiado en la sección 2.5 es eficiente bajo ruido gaussiano, Vovk advierte que no es universalmente consistente si la relación real entre X y Y no es lineal Vovk, Gammerman, and Shafer 2005. Por ello, el desarrollo de CPS basados en kernels o en métodos de vecinos cercanos (como se discute en el capítulo 4 de su obra) es lo que permite alcanzar la consistencia universal en espacios de características complejos. Esta conclusión justifica el uso de arquitecturas no lineales conformizadas en la presente tesis, ya que heredan la solidez de la prueba de consistencia de Vovk.

2.8 Análisis Crítico sobre la Extensión a Series de Tiempo Ergódicas y Estacionarias

Tras haber establecido la consistencia universal de los Sistemas de Predicción Conformal (CPS) en entornos de datos intercambiables, surge una interrogante fundamental para esta investigación: ¿es posible garantizar estas propiedades en el dominio de las series temporales? Dado que la dependencia temporal es la característica intrínseca de estos procesos, el supuesto de intercambiabilidad se debilita. En esta sección, se presenta una discusión teórica sobre la posibilidad de extender la consistencia universal hacia procesos estocásticos ergódicos, basándose en la relajación de los supuestos clásicos y la adopción de herramientas de estimación no paramétrica local.

2.8.1 Hacia la superación del paradigma IID: El rol del Mixing

La extensión de la teoría de Vovk a datos dependientes requiere sustituir la intercambiabilidad por condiciones de *mixing* (mezcla fuerte). La idea central analizada es que, si bien

las observaciones consecutivas están correlacionadas, esta dependencia debe desvanecerse a medida que la distancia temporal aumenta.

Se considera que un sistema predictivo puede alcanzar la consistencia universal en series de tiempo si el proceso cumple con los siguientes supuestos de regularidad:

- **Estacionariedad Estricta y Ergodicidad:** Se asume que las leyes probabilísticas que gobiernan la serie son constantes en el tiempo (estacionariedad) y que una sola realización de la serie es suficiente para inferir las propiedades estadísticas de todo el proceso (ergodicidad). Matemáticamente, esto permite que los promedios temporales converjan a los promedios del ensamble, un pilar para cualquier forma de aprendizaje asintótico.
- **Condición de α -mixing (Mezcla Fuerte):** Este supuesto es el más profundo de la discusión. Expresa cuantitativamente que la dependencia entre el pasado (Z_1, \dots, Z_t) y el futuro lejano (Z_{t+k}, \dots) converge a cero cuando $k \rightarrow \infty$. Si el decaimiento de esta correlación es lo suficientemente rápido (típicamente una tasa algebraica o exponencial), el sistema puede "olvidar" estados pasados irrelevantes y concentrarse en la dinámica actual.
- **Continuidad de Lipschitz en la Estructura de Residuos:** Se asume que la función de distribución acumulada de los residuos, $G(s|x)$, no presenta saltos abruptos ante pequeños cambios en las covariables, lo que permite que el suavizado por kernel sea efectivo.

2.8.2 Mecanismos de Localización: Ventanas Móviles y Kernels

La discusión se centra en un cambio de arquitectura del transductor conformal. En el caso IID, todos los residuos históricos tienen el mismo peso. En la extensión a series de tiempo, se propone un enfoque de *Conformalización Local*. Este mecanismo opera bajo dos dimensiones de adaptación:

1. **Adaptación Temporal mediante Ventana Creciente (L_n):** Se analiza el uso de una ventana de observación que selecciona solo los L_n ejemplos más recientes. Para asegurar la consistencia, el tamaño de esta ventana debe tender a infinito, pero a una tasa menor que el tamaño total de la muestra ($L_n = o(n)$), permitiendo que el sistema se adapte a posibles derivas lentas en la distribución.

2. **Adaptación Espacial mediante Kernels (h_n):** Dentro de la ventana temporal, los residuos $S_i = |Y_i - \mu(X_i)|$ se ponderan mediante una función kernel $K(\cdot)$ con un ancho de banda h_n . Esto garantiza que la distribución predictiva resultante sea "objeto-condicional", es decir, que refleje la incertidumbre específica del contexto actual de la serie (heterocedasticidad local).

2.8.3 Estatus Teórico de la Extensión en la Presente Investigación

Es imperativo aclarar que, si bien el análisis de sesgo y varianza para estimadores de kernel en procesos *mixing* sugiere que la distribución predictiva conformal converge débilmente en probabilidad a la verdadera distribución condicional, en el contexto de este trabajo, dicha conclusión se maneja como un **marco de discusión analítica**.

La complejidad de una demostración formal exhaustiva para el caso general de series de tiempo que requeriría el uso de teoremas límite centrales para campos aleatorios dependientes y un control riguroso de las colas de la distribución excede el alcance probatorio de este capítulo. No obstante, esta discusión proporciona el sustento lógico necesario para aplicar estos sistemas en los experimentos prácticos que siguen. Se establece así un puente teórico: la validez asintótica del sistema no se presenta como una certeza matemática absoluta en este documento, sino como una ruta altamente probable y teóricamente justificada bajo condiciones de estabilidad estocástica, lo que dota de rigor a la arquitectura de cuantificación de incertidumbre empleada en la tesis.

Bibliografía

- Alonso, A., D. Peña, and J. Romo (2002). “Forecasting Time Series With Sieve Bootstrap”. In: *Journal of Statistical Planning and Inference* 108.1-2, pp. 185–206. DOI: [10.1016/S0378-3758\(02\)00269-2](https://doi.org/10.1016/S0378-3758(02)00269-2).
- Arrieta Prieto, Mario Enrique (2017). “Evaluation of the Sieve Bootstrap’s performance in comparison with the classic approach for forecasting purposes in time series analysis”. In: *XXVII Simposio Internacional de Estadística / 5th International Workshop on Applied Statistics*. Poster Presentation. Medellín, Colombia.
- Barber, Rina Foygel et al. (2023). *Conformal Prediction Beyond Exchangeability*. arXiv preprint arXiv:2202.13415v5. Version 5. Accessed on January 3, 2026. arXiv: [2202.13415 \[stat.ME\]](https://arxiv.org/abs/2202.13415).
- Belyaev, Yuri and Sara Sjöstedt-de Luna (2000). “Weakly approaching sequences of random distributions”. In: *Journal of Applied Probability* 37.3, pp. 807–822.
- Bradley, Richard C. (2005). “Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions”. In: *Probability Surveys* 2, pp. 107–144. DOI: [10.1214/154957805100000104](https://doi.org/10.1214/154957805100000104).
- Bühlmann, Peter (1995). *Sieve Bootstrap for Time Series*. Technical report 431. University of California Berkeley.
- (1997). “Sieve Bootstrap for Time Series”. In: *Bernoulli* 3.2, pp. 123–148. DOI: [10.2307/3318610](https://doi.org/10.2307/3318610).
- Gneiting, Tilmann, Fadoua Balabdaoui, and Adrian E. Raftery (2007). “Probabilistic Forecasts, Calibration and Sharpness”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 69.2, pp. 243–268. DOI: [10.1111/j.1467-9868.2007.00587.x](https://doi.org/10.1111/j.1467-9868.2007.00587.x).
- Gneiting, Tilmann and Matthias Katzfuss (2014). “Probabilistic Forecasting”. In: *Annual Review of Statistics and Its Application* 1, pp. 125–151. DOI: [10.1146/annurev-statistics-062713-085831](https://doi.org/10.1146/annurev-statistics-062713-085831).

- Gneiting, Tilmann and Adrian E. Raftery (2007). “Strictly Proper Scoring Rules, Prediction, and Estimation”. In: *Journal of the American Statistical Association* 102.477, pp. 359–378. DOI: [10.1198/016214506000001437](https://doi.org/10.1198/016214506000001437).
- Härdle, Wolfgang K., Joel L. Horowitz, and Jens-Peter Kreiss (2003). “Bootstrap Methods for Time Series”. In: *International Statistical Review* 71.2, pp. 435–459. DOI: [10.1111/j.1751-5823.2003.tb00485.x](https://doi.org/10.1111/j.1751-5823.2003.tb00485.x).
- Hyndman, Rob J and George Athanasopoulos (2021). *Forecasting: Principles and Practice*. 3rd. Melbourne, Australia: OTexts. URL: <https://otexts.com/fpp3/>.
- Künsch, Hans R. (1989). “The Jackknife and the Bootstrap for General Stationary Observations”. In: *The Annals of Statistics* 17.3, pp. 1217–1241. DOI: [10.1214/aos/1176347265](https://doi.org/10.1214/aos/1176347265).
- Lévy, Paul (1937). *Théorie de l'addition des variables aléatoires*. Paris: Gauthier-Villars.
- Politis, Dimitris N. and Joseph P. Romano (1994). “The Stationary Bootstrap”. In: *Journal of the American Statistical Association* 89.428, pp. 1303–1313. DOI: [10.1080/01621459.1994.10476870](https://doi.org/10.1080/01621459.1994.10476870).
- Salinas, David et al. (2020). “DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks”. In: *International Journal of Forecasting* 36.3, pp. 1181–1191. DOI: [10.1016/j.ijforecast.2019.07.001](https://doi.org/10.1016/j.ijforecast.2019.07.001). arXiv: [1704.04110](https://arxiv.org/abs/1704.04110).
- Thorarinsdottir, Thordis L. and Nina Schuhen (Oct. 2017). “Verification: Assessment of Calibration and Accuracy”. In: *Norwegian Computing Center SAMBA/17/17*.
- Vovk, Vladimir (2019). “Universally consistent conformal predictive distributions”. In: *Proceedings of the Eighth Workshop on Conformal and Probabilistic Prediction and Applications (COPA 2019)*. Vol. 105. Proceedings of Machine Learning Research. PMLR, pp. 105–122. URL: <http://proceedings.mlr.press/v105/vovk19a.html>.
- Vovk, Vladimir, Alexander Gammerman, and Glenn Shafer (2005). *Algorithmic Learning in a Random World*. New York: Springer. DOI: [10.1007/b138548](https://doi.org/10.1007/b138548).
- Vovk, Vladimir, Ilia Nouretdinov, et al. (2017). *Conformal predictive distributions with kernels*. Working paper 20. Working Paper 20. Accessed on January 3, 2026. On-line compression modelling project (new series). URL: <http://alrw.net/articles/CPDK.pdf>.