



iimas



Universidad Nacional Autónoma de México

IIMAS

Series de Tiempo

Semestre 2025-2

Proyecto final :

**Series de Tiempo Discretas con modelos AR(1)
utilizando *Exponential Dispersion Models***

Allison Lara Nieva

Mayo 2025

Índice

1. Introducción	3
1.1. Objetivo General	4
1.2. Objetivos Específicos	4
2. Fundamento teórico	5
2.1. Series de tiempo discretas	5
2.2. Modelos AR(1)	5
2.3. Exponential Dispersion Models (EDM)	6
2.4. Thinning y operador de contracción	7
3. Modelo propuesto: AR(1) con márgenes EDM	8
3.1. Dinámica del modelo	8
3.2. Aplicación del operador de adelgazamiento	9
3.3. Desarrollo de ejemplos	10
4. Aplicación práctica	11
4.1. Descripción del dataset	11
4.2. Ajuste del modelo	12
4.3. Resultados	14
5. Discusión	16
5.1. Interpretación de los resultados	16
5.2. Ventajas y desventajas del modelo propuesto	17
6. Conclusiones	17
7. Bibliografía	18
8. Anexos	18

Series de Tiempo Discretas con modelos AR(1) utilizando Exponential Dispersion Models

Allison Lara Nieva

Mayo, 2025

1. Introducción

En el estudio de series de tiempo, los modelos autoregresivos clásicos han demostrado ser de gran utilidad para describir diversos fenómenos. Sin embargo, una limitación recurrente de los enfoques tradicionales, como el modelo AR(1), es su supuesta normalidad del error, asumiendo la mayor parte de las veces que sigue un comportamiento de ruido blanco. En la práctica, muchos fenómenos presentan características que violan este supuesto, como por ejemplo la discreción (que es el caso en particular que se analizará), asimetría y varianza no constante. Estos atributos dificultan un ajuste adecuado por parte de modelos que asumen errores gaussianos.

Para abordar esta limitación, este trabajo propone utilizar un modelo AR(1) estacionario con distribuciones marginales de Exponential Dispersion Models (EDM), enfocándose en distribuciones de tipo discretas como Poisson o Binomial Negativa. Este enfoque, basado en el trabajo de Mena, re-toma ideas centrales del marco desarrollado por Jørgensen y Song en 1998, quienes introdujeron el concepto de *thinning* a través del operador de contracción para generar dependencia temporal en procesos estacionarios con distribuciones marginales no normales.

Este modelo será aplicado a una serie de tiempo construida a partir de datos reales sobre siniestros en el sistema ferroviario mexicano, publicados por el gobierno de México a través de la plataforma datos.gob.mx. Estos datos permiten analizar el número de percances ocurridos desde 2016 hasta 2024 a lo largo de todo el territorio nacional, lo cual representa un caso concreto y relevante de serie temporal discreta.

1.1. Objetivo General

Desarrollar e implementar un modelo $AR(1)$ estacionario con distribuciones marginales en la familia de Exponential Dispersion Models (EDMs) para analizar series de tiempo discretas, mostrando su aplicabilidad práctica en un conjunto de datos real y comparando su desempeño frente a modelos tradicionales.

1.2. Objetivos Específicos

- Revisar la literatura teórica sobre series de tiempo discretas, modelos $AR(1)$ y la familia de Exponential Dispersion Models.
- Analizar el marco propuesto por Mena basado en el operador de contracción y el enfoque de Jørgensen y Song, para construir modelos $AR(1)$ con EDMs.
- Aplicar el modelo a un conjunto de datos discreto, en particular, a los siniestros del sistema ferroviario mexicano, identificando sus características, preparando la serie y estimando los parámetros necesarios.
- Comparar el ajuste del modelo $AR(1)$ -EDM con un modelo $AR(1)$ gaussiano tradicional, evaluando ventajas en términos de ajuste, interpretación y adecuación a la naturaleza discreta de los datos.

2. Fundamento teórico

2.1. Series de tiempo discretas

Las series de tiempo con valores discretos aparecen en numerosos contextos aplicados, particularmente cuando se observan conteos de eventos, objetos o individuos a lo largo del tiempo. Algunos ejemplos incluyen:

- El número de accidentes ferroviarios mensuales (caso aplicado que se estudiará).
- La cantidad de pacientes atendidos diariamente en un hospital.
- El conteo semanal de peces capturados en una región oceánica.
- La clasificación de días como secos o húmedos a partir de una serie continua, es decir, la discretización de una serie que por naturaleza tiene valores continuos.

Estas series se caracterizan por tener una naturaleza discreta, y frecuentemente muestran sobre-dispersión, es decir, su varianza excede a la media, lo cual viola los supuestos de muchos modelos clásicos como ARIMA. Para su modelado, se han desarrollado enfoques especializados que consideran esta estructura que no sigue una distribución marginal de tipo normal necesariamente.

2.2. Modelos AR(1)

Un modelo autoregresivo de orden 1, **AR(1)** de forma abreviada, se define mediante la siguiente expresión:

$$X_t = \phi X_{t-1} + \varepsilon_t$$

donde ε_t es un ruido blanco, usualmente asumido gaussiano. Este modelo presupone además:

- Linealidad en la dependencia temporal.
- Ruido gaussiano e independiente.

- Estacionariedad bajo $|\phi| < 1$.

En el contexto de datos discretos, estas suposiciones son inadecuadas. Específicamente, el uso de ruido continuo y normal puede resultar inapropiado para conteos discretos. Por ello, se han desarrollado modelos alternativos:

Modelos Discretos de Transición Markoviana

Introducidos por Raftery en 1985, modelan la probabilidad de transición entre estados como una mezcla:

$$f(x_t|x_{t-1}, \dots, x_{t-p}) = \sum_{j=1}^p w_j \cdot p(x_t|x_{t-j})$$

Donde w_j es un peso relacionado con la fuerza en que el j -ésimo rezago se incluye. Estos modelos permiten una representación flexible del pasado, pero su interpretación y ajuste pueden volverse complejos.

Modelos Discretos Basados en Adelgazamiento

Caso particular del modelo que se estudiará utilizando un adelgazamiento (*thinning*) binomial para reemplazar el término lineal de la ecuación original del modelo AR(1), que funciona principalmente con distribuciones marginales de tipo Poisson. Su representación es la siguiente:

$$X_t = \alpha * X_{t-1} + \varepsilon_t$$

con $\alpha * N = \sum_{i=1}^N B_i(\alpha)$, donde $\{B_i(\alpha)\}$ es una sucesión de variables aleatorias IID Bernoulli independientes de N , y para las cuales $\mathbb{P}(B_i(\alpha) = 1) = \alpha$. Esto garantiza que la variable siga siendo discreta, y permite construir procesos estrictamente estacionarios.

2.3. Exponential Dispersion Models (EDM)

Un **Exponential Dispersion Model (EDM)** es una familia de distribuciones que generaliza la familia exponencial clásica, incorporando un parámetro de dispersión λ que permite controlar de

forma separada la varianza respecto a la media. Formalmente, una variable aleatoria Z sigue un modelo EDM, denotado por $Z \sim ED^*(\theta, \lambda)$, si su densidad puede escribirse como:

$$f(x; \theta, \lambda) = c(x; \lambda) \exp\{\theta x - \lambda k(\theta)\}$$

donde:

- θ es el parámetro natural (relacionado con la media),
- λ es el parámetro de dispersión,
- $k(\theta)$ es la función cumulante, y
- $c(x; \lambda)$ es una función base que garantiza que la densidad esté bien definida.

Además, entre sus principales propiedades destacan las siguientes:

- **Cerradura bajo convolución:** si $Y_1 \sim ED^*(\theta, \tau_1)$ y $Y_2 \sim ED^*(\theta, \tau_2)$, entonces $Y_1 + Y_2 \sim ED^*(\theta, \tau_1 + \tau_2)$.
- **Infinitamente divisible:** para cualquier entero n , la variable $Z \sim ED^*(\theta, \lambda)$ puede representarse como la suma de variables independientes e idénticamente distribuidas con parámetros $(\theta, \lambda/n)$.
- **Momentos fáciles de calcular:** la media y varianza de $Z \sim ED^*(\theta, \lambda)$ están dadas por $\mathbb{E}(Z) = \lambda k'(\theta)$ y $Var(Z) = \lambda k''(\theta)$.

Ejemplos comunes de EDM son las distribuciones Normal, Poisson, Gamma, Binomial Negativa, Gaussiana Inversa, entre otras.

2.4. Thinning y operador de contracción

Operador de contracción

El concepto de **operador de contracción** fue introducido formalmente en el trabajo de Jørgensen.

Este operador surge al descomponer una variable aleatoria $X \sim ED^*(\theta, \tau_1 + \tau_2)$ como suma de dos variables independientes $Y \sim ED^*(\theta, \tau_1)$ y $X - Y \sim ED^*(\theta, \tau_2)$. La distribución condicional de Y dado $X = x$ se define como:

$$Y \mid X = x \sim G(\tau_1, \tau_2, x)$$

A esta distribución se le conoce como la **densidad de contracción**, y se caracteriza por ser independiente del parámetro θ . Esta operación permite dividir una observación x en dos componentes aleatorios, cada uno con su propia varianza (controlada por τ_1 y τ_2), pero compartiendo la misma forma de distribución. Intuitivamente, podemos pensar que la observación actual X tiene una componente heredado del pasado (Y) y otra generada de forma nueva ($X - Y$). Este mecanismo es fundamental para definir el operador de **thinning**, ya que nos da una forma estructurada y probabilísticamente coherente de separar una observación en una parte dependiente y otra independiente.

Adelgazamiento o *thinning*

El operador de **thinning** es una herramienta que permite generar dependencia entre observaciones preservando la forma de la distribución marginal. Supongamos que $X \sim ED^*(\theta, \lambda)$, entonces se define como:

$$A(X; \alpha) \mid X = x \sim G(\alpha\lambda, (1 - \alpha)\lambda, x)$$

donde $G(\cdot)$ es la **densidad de contracción**. Este operador permite separar a X en dos parte; una parte con distribución $ED^*(\theta, \alpha\lambda)$, que representa la dependencia heredada. Y otra parte con distribución $ED^*(\theta, (1 - \alpha)\lambda)$, que actúa como la innovación o ruido.

3. Modelo propuesto: AR(1) con márgenes EDM

3.1. Dinámica del modelo

La dinámica del modelo AR(1) con distribuciones marginales en la familia de *Exponential Dispersion Models* se basa en la siguiente ecuación presentada en el trabajo de Mena:

$$X_t = A_t(X_{t-1}, \rho) + \varepsilon_t,$$

donde:

- $A_t(X_{t-1}, \rho)$ es un operador aleatorio, condicionalmente distribuido como $G(\rho\lambda, (1 - \rho)\lambda, X_{t-1})$, es decir, una densidad de contracción aplicada sobre X_{t-1} .
- $\varepsilon_t \sim ED^*(\theta, (1 - \rho)\lambda)$ es una innovación independiente.
- $\rho \in (0, 1)$
- La distribución marginal del modelo **AR(1)** es $ED^*(\theta, \lambda)$.

La elección del operador A_t permite que $X_t \sim ED^*(\theta, \lambda)$ de forma marginal, manteniendo la estacionariedad de primer orden. Además, si la esperanza condicional sigue una estructura lineal en el pasado:

$$\mathbb{E}[X_t \mid X_{t-1} = x] = \rho x + (1 - \rho)\mu,$$

donde $\mu = \mathbb{E}[X]$ entonces el proceso tiene autocorrelación de orden h dada por:

$$\text{Corr}(X_{t+h}, X_t) = \rho^h.$$

Este resultado es análogo al caso gaussiano, pero también se prueba en el trabajo de Mena que es válido para cualquier distribución de la familia EDM tal que sea homogénea en el tiempo y que siga la estructura lineal presentada de la esperanza condicional.

3.2. Aplicación del operador de adelgazamiento

El componente principal que permite la definición del modelo es el uso del operador de contracción, denotado por $G(\tau_1, \tau_2, x)$, el cual corresponde a la distribución condicional de una variable $Y \sim ED^*(\theta, \tau_1)$ dado que $X = Y + Z \sim ED^*(\theta, \tau_1 + \tau_2)$.

El operador de adelgazamiento se define entonces por:

$$A_t(X_{t-1}, \rho) \mid X_{t-1} = x \sim G(\rho\lambda, (1 - \rho)\lambda, x).$$

Este operador permite que $A_t(X_{t-1}, \rho) \sim ED^*(\theta, \rho\lambda)$, y al sumarle una innovación $\varepsilon_t \sim ED^*(\theta, (1 - \rho)\lambda)$, se recupera $X_t \sim ED^*(\theta, \lambda)$ gracias a la propiedad de cerradura bajo convolución. De esta manera, la estructura marginal de la distribución se mantiene inalterada, y la dependencia temporal se introduce vía el operador de contracción.

3.3. Desarrollo de ejemplos

Distribución Poisson

Supongamos que $X_t \sim \text{Poisson}(\lambda)$. En este caso, el adelgazamiento corresponde a una binomial:

$$A_t(X_{t-1}, \rho) \mid X_{t-1} = x \sim \text{Binomial}(x, \rho),$$

y la innovación es:

$$\varepsilon_t \sim \text{Poisson}((1 - \rho)\lambda).$$

Esto debido a que la distribución Poisson puede interpretarse como una suma de conteos de eventos individuales que ocurren independientemente a lo largo del tiempo. Si se tiene una cantidad total de eventos $X_{t-1} = x$, y cada uno de estos eventos sobrevive al siguiente período con probabilidad ρ , entonces el número de eventos que permanecen se comporta como una variable aleatoria binomial con parámetros x y ρ . Esta construcción asegura que la parte heredada del pasado, $A_t(X_{t-1}, \rho)$, mantenga la naturaleza discreta del proceso y, además, que al sumarle una innovación Poisson independiente, el resultado total siga teniendo distribución marginal Poisson, gracias a la propiedad de cerradura bajo convolución de la familia EDM.

Por lo que el modelo $X_t = A_t(X_{t-1}, \rho) + \varepsilon_t$ queda con una distribución marginal $X_t \sim \text{Poisson}(\lambda)$ y el valor esperado condicional es $\mathbb{E}[X_t \mid X_{t-1}] = \rho X_{t-1} + (1 - \rho)\lambda$, cumpliendo la relación de autocorrelación ρ^h .

Distribución Binomial Negativa

Sea $X_t \sim \text{NB}(r, p)$, entonces el adelgazamiento queda dado por una **beta-binomial**:

$$A_t(X_{t-1}, \rho) \mid X_{t-1} = x \sim \text{Beta-Binomial}(x, \rho\lambda, (1 - \rho)\lambda)$$

Y la innovación es

$$\varepsilon_t \sim \text{NB}((1 - \rho)\lambda, p)$$

En este caso, la beta-binomial surge de forma natural debido a que la binomial negativa puede interpretarse como una Poisson con parámetro aleatorio. Al aplicar adelgazamiento sobre una Poisson con tasa aleatoria, la variabilidad extra hace que el número de eventos heredados no siga una binomial simple, sino una beta-binomial, que captura la incertidumbre adicional en la probabilidad de retención. Esta propiedad asegura que la estructura marginal se mantenga como una binomial negativa.

El modelo conserva la marginal $X_t \sim \text{NB}(r, p)$, y la esperanza condicional también cumple $\mathbb{E}[X_t | X_{t-1} = x] = \rho x + (1 - \rho)\mu$, donde $\mu = \mathbb{E}[X_t]$.

4. Aplicación práctica

4.1. Descripción del dataset

El conjunto de datos empleado en este trabajo proviene del portal de datos abiertos del Gobierno de México, específicamente del recurso titulado *Siniestros en el Sistema Ferroviario Mexicano* disponible en datos.gob.mx. Este dataset contiene un registro detallado de los accidentes ferroviarios ocurridos en el país, incluyendo información como la fecha del siniestro, ubicación, tipo de incidente y operador involucrado. Contiene 10,250 registros y se ve como sigue:

#	Fecha del siniestro	Entidad Federativa	Municipio	Grupo de siniestro	Tipo de siniestro
1	01-10-2016	Veracruz de Ignacio de la Llave	Medellín De Bravo	Grupo I. Cruces a Nivel	Arrollamiento de Vehículo
2	01-10-2016	Ciudad de México	Azcapotzalco	GRUPO II. Equipo ferroviario, infraestructura y operación	Rozamiento
3	01-10-2016	Michoacán de Ocampo	Morelia	GRUPO II. Equipo ferroviario, infraestructura y operación	Descarrilamiento
4	01-10-2016	Veracruz de Ignacio de la Llave	Xalapa	GRUPO II. Equipo ferroviario, infraestructura y operación	Inundación
5	02-10-2016	San Luis Potosí	Charcas	Grupo I. Cruces a Nivel	Arrollamiento de Vehículo
6	02-10-2016	Tamaulipas	Ciudad Madero	GRUPO II. Equipo ferroviario, infraestructura y operación	Descarrilamiento
7	02-10-2016	Michoacán de Ocampo	Morelia	GRUPO II. Equipo ferroviario, infraestructura y operación	Objetos Sobre la Vía
8	03-10-2016	Tabasco	Huimanguillo	GRUPO II. Equipo ferroviario, infraestructura y operación	Descarrilamiento
9	03-10-2016	Ciudad de México	Azcapotzalco	GRUPO II. Equipo ferroviario, infraestructura y operación	Descarrilamiento
10	03-10-2016	México	Tlalnepantla De Baz	GRUPO II. Equipo ferroviario, infraestructura y operación	Descarrilamiento
11	03-10-2016	Guanajuato	Irapuato	Grupo III. Muerte, lesión y otros	Persona Arrollada
12	03-10-2016	México	Tlalnepantla De Baz	GRUPO II. Equipo ferroviario, infraestructura y operación	Descarrilamiento
13	03-10-2016	Chiapas	Palenque	GRUPO II. Equipo ferroviario, infraestructura y operación	Descarrilamiento

Figura 1: Ejemplo de la base de datos

Para efectos de la serie, se consideraron los accidentes de todo el país, ya que si lo analizábamos por estado, había demasiados valores en cero a lo largo del tiempo. Construyendo así una serie de tiempo discreta con observaciones diarias desde octubre de 2016 hasta abril de 2024. La variable de interés, entonces, corresponde al número de siniestros diarios ocurridos en el país.

El uso de un modelo autoregresivo de primer orden **AR(1)** con distribuciones marginales EDM se justifica ya que la variable que estamos analizando es un conteo no negativo, que típicamente no sigue una distribución normal.

4.2. Ajuste del modelo

La preparación inicial de los datos consistió en la imputación de valores faltantes (en caso de días sin incidentes), y la división del conjunto en dos subconjuntos: **entrenamiento** (hasta diciembre de 2022) y **prueba** (a partir de enero de 2023). No se aplicaron transformaciones adicionales, dado que la serie ya se encuentra en escala natural de conteo.

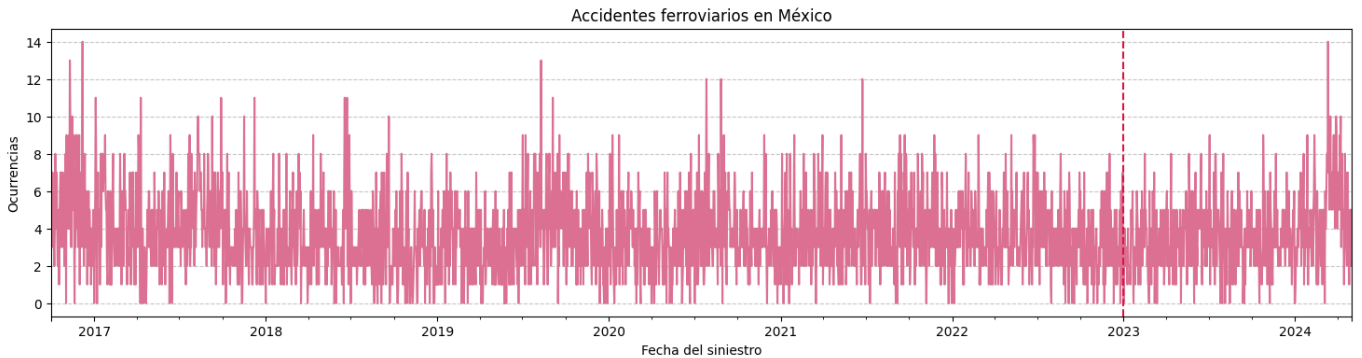


Figura 2: Serie de tiempo de los accidentes ferroviarios mexicanos, marcando el punto de división del conjunto de datos

Para estimar los parámetros del modelo **AR(1)** basados en los datos de entrenamiento, se recurrió al enfoque de **máxima verosimilitud**. En este contexto, la verosimilitud del modelo depende de la combinación del operador de adelgazamiento y de la innovación independiente proveniente de la

misma familia de la distribución marginal deseada.

Distribución Poisson

Dado un proceso AR(1) con distribución marginal $X_t \sim \text{Poisson}(\lambda)$, la verosimilitud condicional para cada observación se construye como:

$$X_t = A_t(X_{t-1}, \rho) + \varepsilon_t$$

donde $A_t(X_{t-1}, \rho) \sim \text{Binomial}(X_{t-1}, \rho)$, y $\varepsilon_t \sim \text{Poisson}((1 - \rho)\lambda)$.

La función de log-verosimilitud conjunta en este caso se define como:

$$\log \mathcal{L}(\lambda, \rho) = \sum_{t=1}^n \log \left(\sum_{k=0}^{\min(X_{t-1}, X_t)} \binom{X_{t-1}}{k} \rho^k (1 - \rho)^{X_{t-1}-k} \cdot \frac{e^{-(1-\rho)\lambda} ((X_t - k)^{(1-\rho)\lambda})}{(X_t - k)!} \right)$$

Distribución Binomial Negativa

Para el modelo con distribución marginal Binomial Negativa, la función de log-verosimilitud incorpora la densidad de la Beta-Binomial como operador de adelgazamiento, es decir $A_t(X_{t-1}, \rho) \mid X_{t-1} = x \sim \text{Beta-Binomial}(x, \rho\lambda, (1 - \rho)\lambda)$, y $\varepsilon_t \sim \text{NB}((1 - \rho)\lambda, p)$. Y la log-verosimilitud conjunta queda definida como:

$$\log \mathcal{L}(\lambda, p, \rho) = \sum_{t=1}^n \log \left(\sum_{k=0}^{\min(X_{t-1}, X_t)} f_{\text{BB}}(k \mid X_{t-1}, \rho\lambda, (1 - \rho)\lambda) \cdot f_{\text{NB}}(X_t - k \mid (1 - \rho)\lambda, p) \right)$$

donde f_{BB} y f_{NB} son las funciones de masa de la beta-binomial y binomial negativa, respectivamente.

Distribución Gaussiana

Dado un proceso AR(1) con distribución marginal gaussiana $X_t \sim \mathcal{N}(\mu, \sigma^2)$, la dinámica del modelo se especifica como:

$$X_t = \mu + \rho(X_{t-1} - \mu) + \varepsilon_t,$$

donde $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$, es decir, ruido blanco normal e independiente.

En este caso, la distribución condicional de X_t dado X_{t-1} es también normal, pues $X_t | X_{t-1} \sim \mathcal{N}(\mu + \rho(X_{t-1} - \mu), \sigma^2)$. La función de log-verosimilitud conjunta queda entonces definida por:

$$\log \mathcal{L}(\mu, \sigma, \rho) = \log f(X_0) + \sum_{t=1}^n \log f(X_t | X_{t-1}) = -\frac{n}{2} \log(2\pi\sigma^2) - \sum_{t=1}^n \frac{(X_t - \mu - \rho(X_{t-1} - \mu))^2}{2\sigma^2}$$

Estas expresiones fueron implementadas computacionalmente en Python, utilizando la librería de *scipy* para obtener los parámetros que maximizan la log-verosimilitud. La elección de este método permite incorporar restricciones sobre los parámetros (por ejemplo, $\rho \in [0, 1]$, $\lambda > 0$, $p \in (0, 1)$) de forma directa.

Para la evaluación del ajuste, se utilizaron dos métricas:

- El **criterio de información de Akaike (AIC)**:

$$\text{AIC} = 2k - 2 \log \mathcal{L}_{\text{máx}},$$

donde k es el número de parámetros estimados.

- El **Error Cuadrático Medio (MSE)** sobre el conjunto de prueba contra las predicciones hechas por los modelos.

Ambas métricas permiten comparar los modelos Poisson, Binomial Negativa y Gaussiano bajo un mismo criterio.

4.3. Resultados

Gráficas de predicción

A continuación se presentan las gráficas de predicción para cada uno de los modelos ajustados, comparando las predicciones generadas con los valores reales observados en el conjunto de prueba (2023–2024). Donde como línea base, se implementó también un modelo AR(1) clásico bajo el supuesto de ruido blanco gaussiano.

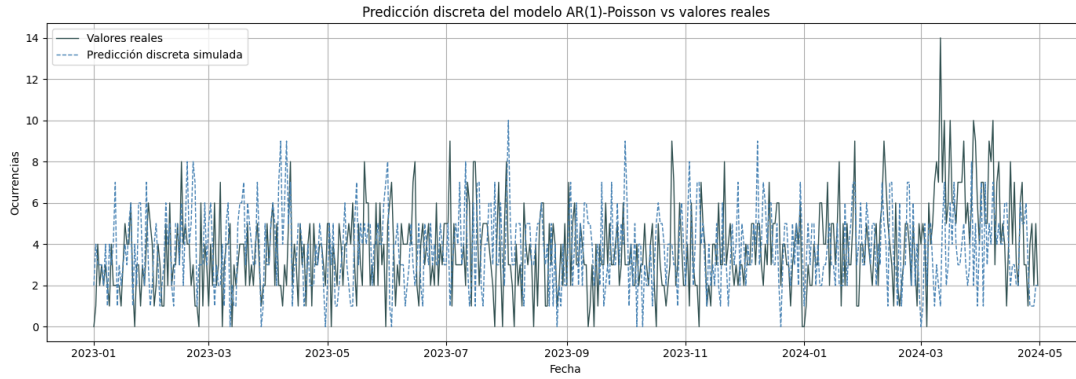


Figura 3: Predicción discreta del modelo AR(1)-Poisson vs valores reales.

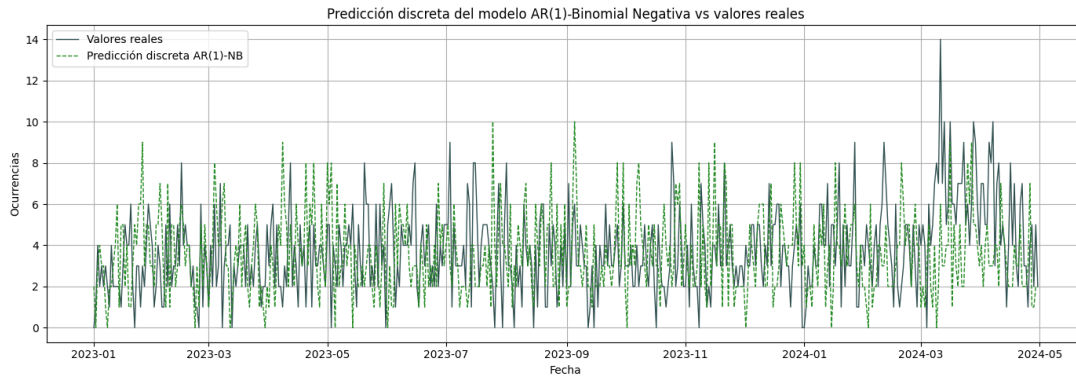


Figura 4: Predicción discreta del modelo AR(1)-Binomial Negativa vs valores reales.

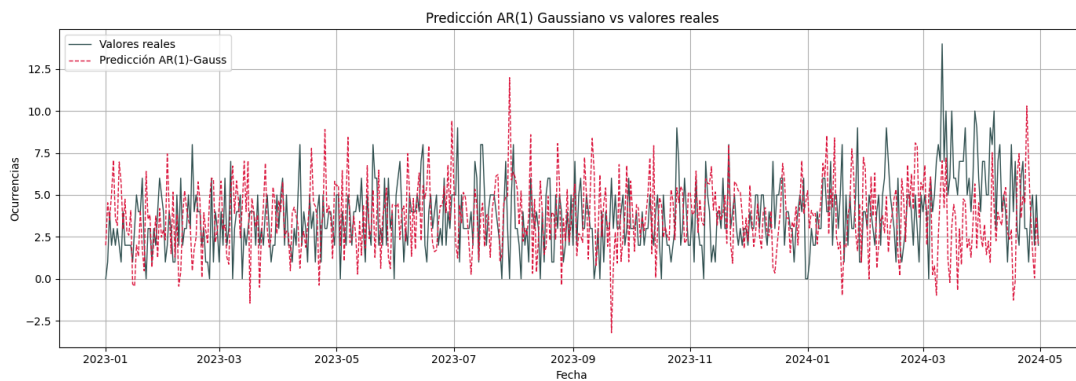


Figura 5: Predicción del modelo AR(1) Gaussiano vs valores reales.

Métricas de ajuste y validación

Como se mencionó en el apartado anterior, las métricas utilizadas para la comparación fueron el Error Cuadrático Medio (MSE) en el conjunto de prueba y el Criterio de Información de Akaike (AIC) calculado en el conjunto de entrenamiento. Los resultados se resumen en la siguiente tabla:

Modelo	MSE (test)	AIC (train)	Tiempo de ejecución
AR(1)-Poisson	7.728	9744.43	78.15 seg
AR(1)-Binomial Negativa	7.813	9695.78	271.12 seg
AR(1)-Gaussiano	8.975	9905.13	13.46 seg

Tabla 1: Comparación de métricas de desempeño entre los modelos AR(1) ajustados.

5. Discusión

5.1. Interpretación de los resultados

Por un lado, los resultados empíricos muestran que el modelo AR(1) con distribución marginal **Binomial Negativa** y adelgazamiento beta-binomial presenta el mejor desempeño global, al obtener el valor más bajo de AIC, lo que sugiere un mejor ajuste a los datos de entrenamiento. Aunque el modelo **Poisson** tiene un MSE ligeramente menor en el conjunto de prueba.

Por otro lado, el modelo **Gaussiano** mostró el peor desempeño tanto en MSE como en AIC. Además, sus predicciones continúan generando valores no enteros y hasta con valores negativos, lo que es inconsistente con la naturaleza discreta de los datos observados. Esto refuerza la idea de que asumir normalidad para series de conteo puede ser inapropiado y limitar la capacidad del modelo para capturar adecuadamente la dinámica del fenómeno.

5.2. Ventajas y desventajas del modelo propuesto

El uso de modelos $AR(1)$ con distribuciones marginales pertenecientes a la familia de Exponential Dispersion Models (EDM), junto con el uso de operadores de adelgazamiento (como la distribución binomial o la beta-binomial), presenta múltiples ventajas como la conservación de la estructura de los datos que por naturaleza son discretos. Además de que ya se observó que representan de mejor forma a este tipo de datos no sólo de forma cualitativa, sino también de forma cuantitativa.

Entre los principales retos se encuentran sobre todo el alto costo computacional que puede llegar a tener en series de mayor tamaño debido a que el cálculo de la verosimilitud conjunta para modelos no gaussianos requiere sumas internas. Además de que la estimación de parámetros mediante máxima verosimilitud no siempre es estable si se parte de malas condiciones iniciales.

6. Conclusiones

En este trabajo se exploraron modelos autorregresivos de primer orden $AR(1)$ adaptados a series de tiempo con datos discretos, utilizando distribuciones marginales dentro de la familia de los Exponential Dispersion Models (EDM). Se implementaron y compararon tres variantes principales: un modelo $AR(1)$ -Poisson con adelgazamiento binomial, un modelo $AR(1)$ -Binomial Negativa con adelgazamiento beta-binomial, y un modelo $AR(1)$ clásico con ruido gaussiano. Los resultados empíricos mostraron que el modelo con distribución marginal binomial negativa obtuvo el mejor desempeño global según el AIC, pero con un trade-off importante en el tiempo de ejecución. Mientras que el modelo gaussiano fue el menos adecuado tanto en términos de ajuste como de predicción.

La principal contribución de este proyecto es la implementación y comparación de modelos $AR(1)$ con las características ya mencionadas, aplicados a datos reales de siniestros ferroviarios en México. Donde el enfoque propuesto permite preservar la naturaleza discreta de los datos y mantener la distribución marginal deseada. Además, el marco de modelado desarrollado en este trabajo ofrece una alternativa robusta y de fácil entendimiento para trabajar con este tipo de datos.

7. Bibliografía

- Jørgensen, B. and Song, P. (1998). Stationary time series models with exponential dispersion model margins. *Journal of Applied Probability*, 35, 78-92.
- Vázquez Rivera, E. A. (2007). Series de tiempo con valores discretos [Tesis de licenciatura, Universidad Nacional Autónoma de México, Facultad de Ciencias]. Dirección General de Bibliotecas UNAM.
- Mena, R. H. (s.f.). AR(1) stationary models with marginal distributions in the $ED^*(\theta, \tau)$ family
- Mena, R. H. (s.f.). Clase 8: Modelos para Valores Discretos

8. Anexos

- Carpeta de Drive con la base de datos y análisis completo en Python: TS Proyecto Final

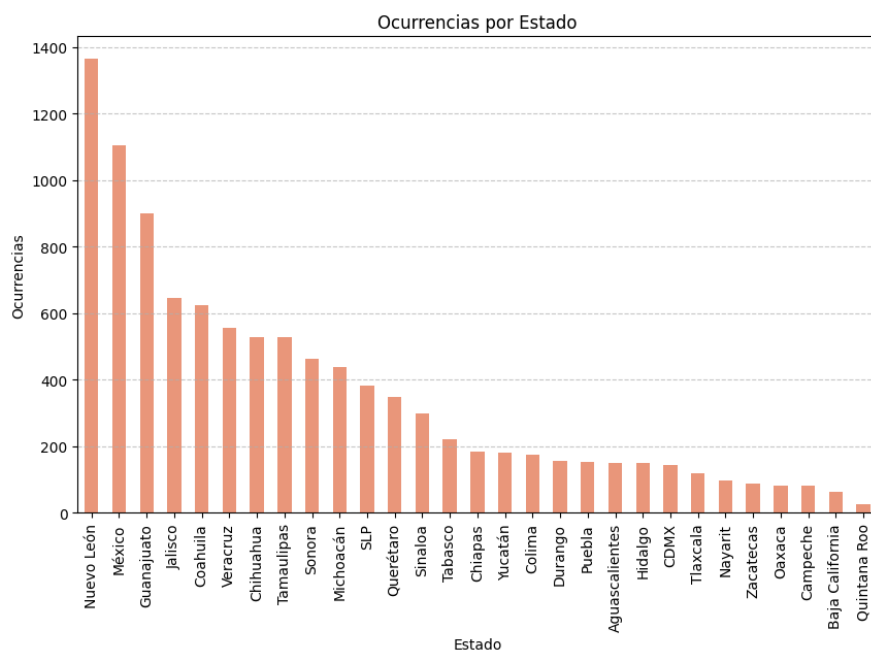


Figura 6: Cantidad de accidentes ferroviarios por estado

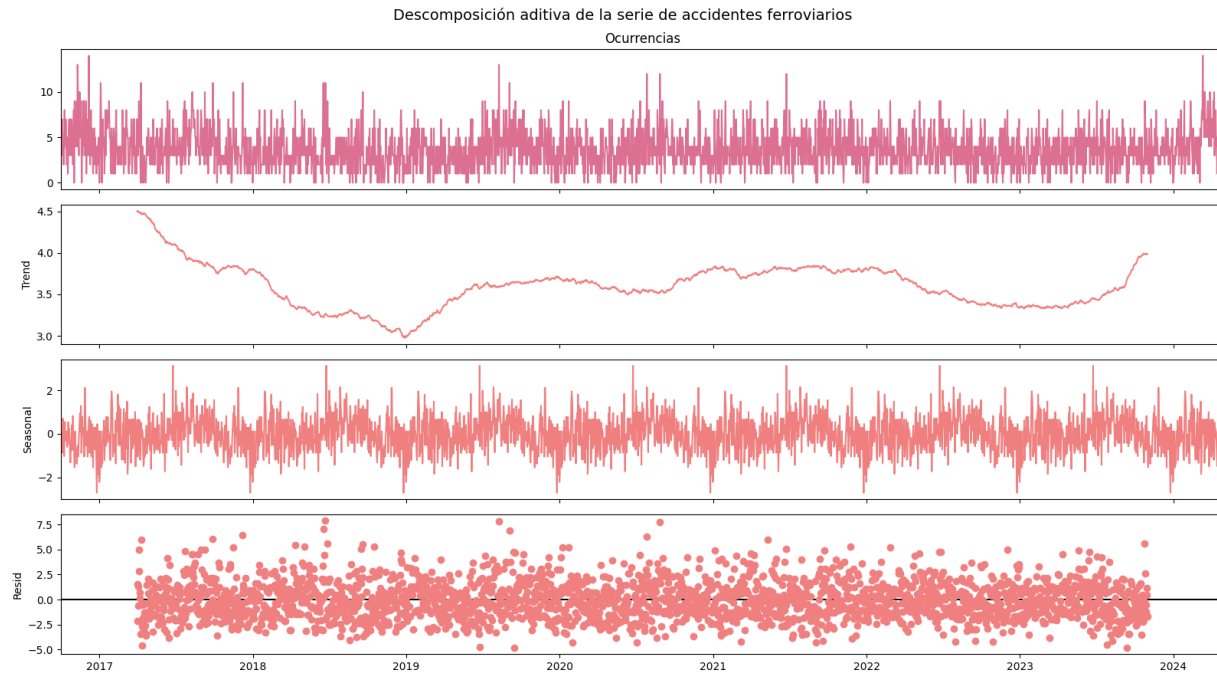


Figura 7: Descomposición aditiva de la serie de tiempo