

Modelado de KRIs utilizando Difusiones de Jacobi

Olivares Cruz Yael Fabián

3 de diciembre de 2025

Índice

1	Introducción	2
2	Objetivo	3
3	Marco teórico	4
4	Metodología	8
5	Análisis de resultados	11
6	Conclusiones	17
7	Bibliografía	18
8	Apéndice. Repositorio de código	19

1. Introducción

En la gestión de riesgos, la precisión no es sólo un requisito regulatorio, sino una necesidad estratégica. Las instituciones dependen de Indicadores Clave de Riesgo (KRIs), como índices de capacidad o de obsolescencia, para tomar decisiones críticas. Sin embargo, existe un desafío técnico al modelar muchos de estos indicadores, su naturaleza acotada.

Los KRIs son métricas diseñadas para monitorear y anticipar tendencias adversas en los procesos, la tecnología o la infraestructura. Su función es alertar sobre deterioros antes de que se materialicen pérdidas. En este trabajo se analizará el KRI ACM0016.MX. Que cuantifica el nivel de ocupación del almacenamiento de backups en la plataforma corporativa para respaldos (Datadomain) de producción ubicada en México. Este indicador se alerta en amarillo cuando alcanza un valor mayor o igual a 0.8 y en rojo cuando es mayor o igual a 0.9.

A diferencia de los precios de activos o tipos de cambio, que teóricamente pueden extenderse indefinidamente, la mayoría de los KRIs son porcentajes o ratios estrictamente confinados al intervalo $[0, 1]$. Utilizar modelos tradicionales, como el movimiento browniano geométrico puede resultar inadecuado, ya que no garantiza que las proyecciones se mantengan dentro de límites realistas, lo que puede llevar a subestimar la volatilidad o generar escenarios imposibles.

Este reporte aborda dicha problemática mediante la implementación del proceso de difusión de Jacobi. Este modelo estocástico es ideal para variables acotadas, ya que asegura matemáticamente que la evolución de riesgo permanezca siempre dentro del rango permitido. Más allá de la elección del modelo, este trabajo enfrenta un segundo problema en la industria, la calidad de la información. Dado que los datos de riesgos operativos suelen presentar huecos debido a que se reportan mensual o trimestralmente, se ha desarrollado una metodología que combina puentes de difusión para la reconstrucción de datos faltantes y el algoritmo Gibbs Sample para la calibración de parámetros.

El objetivo final de este análisis es demostrar cómo estas técnicas bayesianas y estocásticas permiten no solo simular trayectorias de riesgo realistas, sino también inferir los verdaderos parámetros de tendencia y volatilidad con mayor robustez que los métodos de estimación lineal tradicionales.

2. Objetivo

El objetivo de este proyecto es desarrollar e implementar un marco metodológico robusto basado en el proceso de difusión de Jacobi para el modelado, simulación y calibración de KRIs acotados en el intervalo $[0, 1]$, garantizando la consistencia matemática de las proyecciones y mejorando la precisión en la medición del riesgo operacional.

Se establecen la siguientes metas técnicas:

1. **Simulación de procesos acotados.** Implementar el esquema de discretización de Milstein para generar trayectorias estocásticas del proceso de Jacobi, asegurando que el indicador de riesgo no salga del intervalo $[0, 1]$.
2. **Reconstrucción de información incompleta.** Aplicar técnicas de puentes de difusión para completar datos faltantes en series de tiempo irregulares, permitiendo un análisis de riesgo con mayor información.
3. **Calibración Bayesiana de parámetros.** Estimar los parámetros de riesgo, tales como velocidad de reversión a la media, tendencia a largo plazo y volatilidad. Utilizando algoritmos de muestreo Gibbs.
4. **Validación y comparación.** Evaluar el desempeño del modelo propuesto comparando los parámetros estimados sobre la serie completa, la serie reducida y la reconstruida, comparando la precisión de los estimadores.
5. **Aplicación a un KRI de capacidad de red local.** El modelo se utilizará para diagnosticar el estado de la red respondiendo a cuatro interrogantes mediante el análisis de parámetros estadísticos:
 - a) **Resiliencia Operativa.** ¿Qué tan rápido se recupera la red tras un pico de demanda?
 - b) **Evaluación de Capacidad.** ¿La infraestructura actual está sobredimensionada o subdimensionada?
 - c) **Naturaleza del tráfico.** ¿El comportamiento de la red es estable y predecible o volátil?
 - d) **Proyección de Riesgo.** ¿Cuál es la probabilidad de alcanzar el umbral de alerta en los próximos 30 días?

3. Marco teórico

En esta sección se desarrollan las bases matemáticas y computacionales necesarias para la simulación del proceso de Jacobi. Comenzaremos definiendo lo que es un proceso de difusión.

Definición 3.1 (Procesos de difusión)

Un proceso de difusión es un proceso estocástico continuo en el tiempo, definido como la solución de una ecuación diferencial estocástica (EDE) de la forma:

$$dX_t = b_\alpha(X_t) dt + \sigma_\beta(X_t) dW_t, \quad X_0 = x_0,$$

donde:

- $b_\alpha(X_t)$ es el **coeficiente de deriva**, que describe la tendencia determinista del proceso y depende de un vector de parámetros $\alpha = (\alpha_1, \dots, \alpha_p)$;
- $\sigma_\beta(X_t)$ es el **coeficiente de difusión**, que controla la magnitud de las fluctuaciones aleatorias y depende de un vector de parámetros $\beta = (\beta_1, \dots, \beta_q)$;
- W_t es un movimiento browniano estándar unidimensional definido en un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbf{P})$.

La dinámica del proceso está determinada por la interacción entre estos dos componentes:

$$b_\alpha(x) = \text{drift (componente determinista)}, \sigma_\beta(x) = \text{diffusion (componente estocástico)}.$$

El primer término, $b_\alpha(x) dt$, representa el desplazamiento promedio del proceso en un intervalo infinitesimal de tiempo dt , mientras que el segundo término, $\sigma_\beta(x) dW_t$, introduce la variabilidad aleatoria, proporcional a la intensidad de la difusión $\sigma_\beta(x)$.

El objetivo estadístico consiste en estimar los parámetros desconocidos α y β a partir de observaciones del proceso X_t . En particular, cuando las observaciones son discretas y el modelo pertenece a la familia de procesos de Itô, la inferencia puede realizarse a partir de las propiedades de la función de deriva y del coeficiente de difusión, utilizando métodos de máxima verosimilitud, de momentos o de inferencia bayesiana.

A continuación presentaremos el modelo de Jacobi como proceso de difusión.

Definición 3.2 (Modelo de Jacobi)

El modelo de Jacobi es un proceso de difusión que tiene la siguiente expresión.

$$dX_t = \theta(\mu - X_t)dt + \sigma\sqrt{X_t(1 - X_t)}dW_t \quad X_t, \mu \in (0, 1), \quad \theta, \sigma > 0$$

Consideremos el modelo de Jacobi bajo la forma reparametrizada considerando

$$\alpha_1 = \theta\mu, \quad \alpha_2 = \theta.$$

entonces el modelo queda expresado de la siguiente forma

$$dX_t = (\alpha_1 - \alpha_2 X_t) dt + \sigma\sqrt{X_t(1 - X_t)} dW_t, \quad X_t \in (0, 1), \quad \alpha_1, \alpha_2, \sigma > 0.$$

El parámetro α_1 regula el nivel de atracción hacia los valores altos del proceso, mientras que α_2 controla la velocidad de reversión media. El término de difusión $\sigma\sqrt{X_t(1 - X_t)}$ garantiza que las trayectorias permanezcan dentro del intervalo $(0, 1)$, ya que la volatilidad se anula en los extremos $X_t = 0$ y $X_t = 1$.

Un aspecto fundamental en el análisis de datos estocásticos es la estimación de los parámetros que rigen la dinámica del fenómeno observado, permitiendo así ajustar el modelo teórico con la información real. A continuación, se expondrán los fundamentos teóricos para la estimación de parámetros mediante el método de máxima verosimilitud y variación cuadrática.

Definición 3.3 (Estimador máximo verosimil)

El estimador de máxima verosimilitud (EMV) $\hat{\alpha}_{MV}$ se define como la solución de

$$L_T(\hat{\alpha}_{MV}) = \sup_{\alpha \in \mathcal{A}} L_T(\alpha),$$

donde \mathcal{A} es el espacio paramétrico de (α_1, α_2) .

Cuando no es posible encontrar los estimadores directamente mediante esta vía, los parámetros asociados al coeficiente de difusión pueden estimarse a partir de la variación cuadrática del proceso:

$$\langle X, X \rangle_T = \lim_{n \rightarrow \infty} \sum_{i=1}^n (X_{t_i} - X_{t_{i-1}})^2 = \int_0^T \hat{\sigma}^2(X_s) ds.$$

Dado que el estimador del coeficiente de difusión es asintóticamente normal, podemos concluir que es insesgado.

El resto de los parámetros, contenidos en el coeficiente de deriva, pueden ser calculados mediante máxima verosimilitud, cuya función de verosimilitud está dada por:

$$L_T(\alpha) = \exp \left(\int_0^T \frac{b_\alpha(X_t)}{\hat{\sigma}^2(X_t)} dX_t - \frac{1}{2} \int_0^T \left[\frac{b_\alpha(X_t)}{\hat{\sigma}(X_t)} \right]^2 dt \right),$$

donde $b_\alpha(X_t) = \alpha_1 - \alpha_2 X_t$.

Otro enfoque para la estimación de parámetros es el método de momentos, para ello utilizamos la esperanza condicional del modelo de Jacobi, discretizado a paso Δt , está dada por

$$\mathbf{E}[X_{t+\Delta t} | X_t] = X_t e^{-\alpha_2 \Delta t} + \frac{\alpha_1}{\alpha_2} (1 - e^{-\alpha_2 \Delta t}).$$

Sean

$$a := e^{-\alpha_2 \Delta t}, \quad b := \frac{\alpha_1}{\alpha_2} (1 - e^{-\alpha_2 \Delta t}).$$

Esto sugiere estimar a y b mediante mínimos cuadrados ordinarios sobre el modelo lineal

$$X_i = aX_{i-1} + b + \varepsilon_i.$$

Los estimadores para α_1 y α_2 se obtienen entonces como:

$$\hat{\alpha}_2 = -\frac{\ln(\hat{a})}{\Delta t}, \quad \hat{\alpha}_1 = \frac{\hat{b} \hat{\alpha}_2}{1 - \hat{a}}.$$

Por último, para estimar el parámetro σ utilizamos estimación por variación cuadrática, que está dada de la siguiente forma.

Definición 3.4 (Variación cuadrática)

Dado un proceso de difusión de Itô

$$dX_t = \mu(X_t, t) dt + \sigma(X_t, t) dW_t,$$

con W_t un movimiento browniano estándar, la variación cuadrática en el intervalo $[0, T]$ se define como

$$[X]_T = \lim_{|\pi| \rightarrow 0} \sum_{i=0}^{n-1} (X_{t_{i+1}} - X_{t_i})^2,$$

donde $\pi = \{t_0 < t_1 < \dots < t_n\}$ es una partición del intervalo.

Para un proceso de difusión de Itô, se cumple que

$$[X]_T = \int_0^T \sigma^2(X_s, s) ds.$$

En el caso particular del modelo de Jacobi reparametrizado:

$$dX_t = (\alpha_1 - \alpha_2 X_t) dt + \sigma \sqrt{X_t(1 - X_t)} dW_t,$$

entonces

$$[X]_T = \int_0^T \sigma^2 X_s(1 - X_s) ds.$$

De la definición discreta de la variación cuadrática,

$$\sum_{i=1}^n (X_{t_i} - X_{t_{i-1}})^2 \longrightarrow \int_0^T \sigma^2 X_s(1 - X_s) ds, \quad \max \Delta_i \rightarrow 0, \quad \Delta_i = t_i - t_{i-1}.$$

Por tanto,

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_{t_i} - X_{t_{i-1}})^2}{\sum_{i=1}^n X_{t_{i-1}}(1 - X_{t_{i-1}}) \Delta_i}, \quad \hat{\sigma} = \sqrt{\hat{\sigma}^2}.$$

Para asegurar que el proceso de Jacobi permanezca estrictamente dentro del intervalo $(0, 1)$ es necesario que el coeficiente de difusión se anule en los extremos y que el drift sea suficientemente fuerte para evitar la absorción en las fronteras

Definición 3.5 (Condiciones tipo Feller)

El modelo

$$dX_t = \theta(\mu - X_t) dt + \sigma \sqrt{X_t(1 - X_t)} dW_t,$$

cumple las condiciones tipo Feller siempre que

$$2\theta\mu \geq \sigma^2 \quad y \quad 2\theta(1 - \mu) \geq \sigma^2.$$

Estas desigualdades garantizan que las fronteras 0 y 1 sean inalcanzables en tiempo finito, haciendo que la dinámica del proceso esté confinada de manera natural. En el contexto de KRIs, esto implica que el indicador no alcanzará valores inválidos o físicamente imposibles fuera del rango permitido.

Una propiedad fundamental de los procesos con reversión a la media es el tiempo característico que tarda la serie en volver a su nivel de equilibrio. Consideremos la parte determinista del proceso de Jacobi,

$$dX_t = \theta(\mu - X_t) dt,$$

cuyo comportamiento describe un decaimiento exponencial hacia la media μ . La solución de esta ecuación diferencial ordinaria es

$$\mathbf{E}[X_t] - \mu = (X_0 - \mu) e^{-\theta t}.$$

Esto implica que la distancia al nivel de equilibrio se reduce exponencialmente. La vida media, se define como el tiempo $t_{1/2}$ en el que dicha distancia se reduce exactamente a la mitad:

$$e^{-\theta t_{1/2}} = \frac{1}{2}.$$

Resolviendo para $t_{1/2}$ obtenemos

$$t_{1/2} = \frac{\ln(2)}{\theta}.$$

Esta cantidad proporciona una medida directa de la rapidez con la que el indicador regresa a su comportamiento típico después de un evento. Valores grandes de θ implican una rápida reversión, mientras que valores pequeños corresponden a un proceso más persistente.

Es importante señalar que esta interpretación se obtiene considerando únicamente la dinámica determinista del modelo; es decir, corresponde al comportamiento del drift en ausencia del término estocástico. Por lo tanto la vida media debe entenderse como una aproximación útil para interpretar el parámetro θ , pero no como una descripción completa de la dinámica total del proceso, la cual está influida también por la difusión.

Una vez obtenida la trayectoria reconstruida del modelo, es importante validar que la aplicación del proceso de difusión sea consistente con los datos observados, para esto consideramos lo siguiente. Un proceso de difusión es de la forma

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t$$

Mediante la transformación de Lamperti, es posible reescribir el proceso en términos de una variable transformada con difusión uno de la forma:

$$dY_t = \alpha(Y_t)dt + dW_t$$

En versión discreta se ve de la forma

$$Y_{t_i} - Y_{t_{i-1}} - \alpha(Y_{t_i}) = W_{t_i} - W_{t_{i-1}}, \quad W_{t_k} \sim N(0, \Delta t)$$

Entonces, si definimos

$$Z_i = Y_{t_i} - Y_{t_{i-1}} - \alpha(Y_{t_i})$$

Entonces los residuos transformados

$$Z_1, \dots, Z_n \sim N(0, \Delta t)$$

Deberían comportarse como variables independientes con distribución normal, lo que permite evaluar la calidad del modelo mediante pruebas de normalidad y análisis de resultados.

4. Metodología

En esta sección se detallan los pasos matemáticos necesarios para aplicar el algoritmo Gibbs Sample al modelo de Jacobi. Comenzamos con la transformación de Lamperti para simplificar la difusión.

La transformación de Lamperti está dada por

$$\begin{aligned} h(x) &= \int_{x^*}^x \frac{1}{\sigma_\beta(y)} dy; \quad \sigma_\beta(y) = \sigma \sqrt{y(1-y)} \\ &= \int_{x^*}^x \frac{1}{\sigma \sqrt{y(1-y)}} dy \\ &= \frac{1}{\sigma} \int_{x^*}^x \frac{1}{\sqrt{y(1-y)}} dy \\ &= \frac{1}{\sigma} \arcsen(2x - 1) \end{aligned}$$

De modo que

$$h(x) = \frac{1}{\sigma} \arcsen(2x - 1) \text{ y } h^{-1}(y) = \frac{1}{2}(\sen(\sigma y) + 1)$$

Sea $\eta_y := h^{-1}(y)$

Por fórmula de Íto sabemos que

$$dY_t = M_{\alpha,\beta}(Y_t)dt + dW_t$$

$$\text{donde } M_{\alpha,\beta}(Y_t) = \frac{b_\alpha(h^{-1}(Y_t))}{\sigma_\beta(h_\beta^{-1}(Y_t))} - \frac{1}{2}\sigma'(h_\beta^{-1}(Y_t))$$

$$\text{Como } b_\alpha(x) = \alpha_1 - \alpha_2 x \text{ y } \sigma_\beta(y) = \sigma \sqrt{y(1-y)}, \text{ entonces } \sigma'_\beta(y) = \frac{1-2x}{2\sqrt{x(1-x)}}\sigma$$

La función de verosimilitud para los parámetros $\alpha = (\alpha_1, \alpha_2)$, dados los datos observados Y_{obs} y los datos faltantes Y_{mis} , pueden escribirse en términos de la familia exponencial como:

$$\mathcal{L}(\theta; Y_{obs}, Y_{mis}) = \exp\left\{\alpha^t \nu_\beta - \frac{(\alpha^t)^2 \Lambda_\beta}{2}\right\}$$

Para expresar la deriva en la forma canónica de la familia exponencial, consideramos

$$b_\alpha(x) = \alpha_1 - \alpha_2 x$$

Definimos las funciones base como $w_1(x) = 1, w_2(x) = -x$. De esta forma, la deriva se reescribe como

$$b_\alpha(x) = \alpha_1 w_1(x) + \alpha_2 w_2(x)$$

Además las derivadas de estas funciones base son $w'_1(x) = 0, w'_2(x) = -1$

Las funciones κ son necesarias para calcular las estadísticas ν_β y Λ_β , se definen de la siguiente forma:

$$\kappa_k^1 = \frac{1}{2}[w'_k(h_\beta^{-1}(y)) - 2w_k(h_\beta^{-1}(y))[ln(\sigma_\beta(h_\beta^{-1}(y)))]']$$

$$\kappa_{i,j}^2 = \frac{w_i(h_\beta^{-1}(y))w_j(h_\beta^{-1}(y))}{\sigma_\beta^2(h_\beta^{-1}(y))}$$

Sustituyendo las expresiones específicas para el modelo de Jacobi, obtenemos:

$$\kappa_1^1(y) = -\frac{1}{2} \frac{\eta'_y(1-2\eta_y)}{\eta_y(1-\eta_y)}$$

$$\kappa_2^1(y) = -\frac{1}{2} + \frac{\eta'_y(1-2\eta_y)}{2(1-\eta_y)}$$

$$\kappa_{1,1}^2(y) = \frac{1}{\sigma^2 \eta_y(1-\eta_y)}$$

$$\kappa_{2,2}^2(y) = \frac{\eta_y}{\sigma^2(1-\eta_y)}$$

$$\kappa_{1,2}^2(y) = \kappa_{2,1}^2(y) = -\frac{1}{\sigma^2(1-\eta_y)}$$

Donde $\eta_y^i(s) = \frac{1}{2}(\sen(\sigma y) + 1)$ y su derivada es $\frac{d}{ds}\eta_y^i(s) = \frac{\sigma}{2}\cos(\sigma y)$

El vector ν_β tiene dos componentes ν_{β^1} y ν_{β^2} , definidos por

$$\nu_\beta^j = \int_{x_1}^{x_n} \frac{w_j(y)}{\sigma_\beta^2(y)} dy - \sum_{i=1}^n \int_{t_{i-1}}^{t_i} k_j^1(z_s + l_\beta^i(s)) ds, \quad j = 1, 2$$

Para la primera componente

$$\begin{aligned} \nu_\beta^1 &= \int_{x_1}^{x_n} \frac{1}{\sigma^2 y(1-y)} dy - \sum_{i=1}^n \int_{t_{i-1}}^{t_i} -\frac{1}{2} \frac{\frac{d}{ds} \eta_y^i(s)(1-2\eta_y^i(s))}{\eta_y^i(s)(1-\eta_y^i(s))} ds \\ &= \frac{1}{\sigma^2} (\ln|y| - \ln|y-1|) \Big|_{x_1}^{x_n} + \frac{1}{2} \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \frac{\frac{d}{ds} \eta_y^i(s)(1-2\eta_y^i(s))}{\eta_y^i(s)(1-\eta_y^i(s))} ds \\ &= \frac{1}{\sigma^2} D_1 + \frac{1}{2} \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \frac{\frac{d}{ds} \eta_y^i(s)(1-2\eta_y^i(s))}{\eta_y^i(s)(1-\eta_y^i(s))} ds \end{aligned}$$

Donde $D_1 = \ln \left| \frac{x_n(x_1-1)}{x_1(x_n-1)} \right|$

Para la segunda componente

$$\begin{aligned} \nu_\beta^2 &= \int_{x_1}^{x_n} \frac{-y}{\sigma^2 y(1-y)} dy - \sum_{i=1}^n \int_{t_{i-1}}^{t_i} -\frac{1}{2} + \frac{\frac{d}{ds} \eta_y^i(s)(1-2\eta_y^i(s))}{2(1-\eta_y^i(s))} ds \\ &= -\frac{1}{\sigma^2} (-\ln|1-y|) \Big|_{x_1}^{x_n} + \frac{1}{2} \sum_{i=1}^n s \Big|_{t_{i-1}}^{t_i} - \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \frac{\frac{d}{ds} \eta_y^i(s)(1-2\eta_y^i(s))}{2(1-\eta_y^i(s))} ds \\ &= \frac{1}{\sigma^2} D_2 + \frac{1}{2} (t_n - t_0) - \frac{1}{2} \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \frac{\frac{d}{ds} \eta_y^i(s)(1-2\eta_y^i(s))}{1-\eta_y^i(s)} ds \end{aligned}$$

Donde $D_2 = \ln \left| \frac{1-x_n}{1-x_1} \right|$,

$$\eta_y^i(s) = \frac{1}{2} (\sin(\sigma(z_s^i + l_\beta^i(s))) + 1),$$

$$\frac{d}{ds} \eta_y^i(s) = \frac{\sigma}{2} \cos(\sigma(z_s^i + l_\beta^i(s))).$$

La matriz Λ_β se calcula integrando las funciones $\kappa_{i,j}^2$ a lo largo del camino propuesto.

$$\Lambda_\beta^{i,j} = \sum_{k=1}^n \int_{t_{k-1}}^{t_k} k_{i,j}^2(z_s^i + l_\beta^i(s)) ds$$

Calculamos sus componentes.

$$\begin{aligned} \Lambda_\beta^{1,1} &= \sum_{k=1}^n \int_{t_{k-1}}^{t_k} k_{1,1}^2(z_s^i + l_\beta^i(s)) ds \\ &= \sum_{k=1}^n \int_{t_{k-1}}^{t_k} \frac{1}{\sigma^2 \eta_y^i(s)(1-\eta_y^i(s))} ds \\ &= \frac{1}{\sigma^2} \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \frac{1}{\eta_y^i(s)(1-\eta_y^i(s))} ds \end{aligned}$$

$$\begin{aligned} \Lambda_\beta^{2,2} &= \sum_{k=1}^n \int_{t_{k-1}}^{t_k} k_{2,2}^2(z_s^i + l_\beta^i(s)) ds \\ &= \frac{1}{\sigma^2} \sum_{k=1}^n \int_{t_{k-1}}^{t_k} \frac{\eta_y^i(s)}{1-\eta_y^i(s)} ds \end{aligned}$$

$$\begin{aligned} \Lambda_\beta^{1,2} &= \Lambda_\beta^{2,1} = \sum_{k=1}^n \int_{t_{k-1}}^{t_k} k_{1,2}^2(z_s^i + l_\beta^i(s)) ds \\ &= -\frac{1}{\sigma^2} \sum_{k=1}^n \int_{t_{k-1}}^{t_k} \frac{1}{1-\eta_y^i(s)} ds \end{aligned}$$

Donde $\eta_y^i(s) = \frac{1}{2}(\text{sen}(\sigma(z_s^i + l_\beta^i(s))) + 1)$

El término l_β^i representa la corrección lineal del puente browniano y se calcula como:

$$\begin{aligned} l_\beta^i(\beta, \beta_0) &= \frac{1}{\Delta} \left[(t_i - t) \left(\frac{x_{i-1}}{\beta} - \frac{x_{i-1}}{\beta_0} \right) + (t - t_{i-1}) \left(\frac{x_i}{\beta} - \frac{x_i}{\beta_0} \right) \right]; t_i - t_{i-1} = \Delta, t \in (t_{i-1}, t_i) \\ &= -\frac{(t_i - t)x_{i-1}}{\beta_0 \Delta} - \frac{t_i - t_{i-1}}{\beta_0 \Delta} x_i + \frac{1}{\beta} \left(\frac{(t_i - t)x_{i-1} + (t - t_{i-1})x_i}{\Delta} \right) \\ &= C_1^i + \frac{1}{\beta} C_2^i \end{aligned}$$

Donde C_1^i y C_2^i son constantes que dependen de los datos observados y del parámetro inicial β_0 .

Las expresiones finales para ν_β y Λ_β son: $D_1 = \ln \left| \frac{x_n(x_1-1)}{x_1(x_n-1)} \right|$
 $D_2 = \ln \left| \frac{1-x_n}{1-x_1} \right|$
 $\eta_y^i(s) = \frac{1}{2}(\text{sen}(\sigma(z_s^i + l_\beta^i(s))) + 1)$
 $\frac{d}{ds} \eta_y^i(s) = \frac{\sigma}{2} \cos(\sigma(z_s^i + l_\beta^i(s)))$

$$\begin{aligned} \nu_\beta &= \left(\frac{1}{\sigma^2} D_1 + \frac{1}{2} \sum_{i=1}^n \int_{t_{k-1}}^{t_k} \frac{\frac{d}{ds} \eta_y^i(s)(1 - 2\eta_y^i(s))}{\eta_y^i(s)(1 - \eta_y^i(s))} ds, \frac{1}{\sigma^2} D_2 + \frac{1}{2}(t_n - t_0) - \frac{1}{2} \sum_{i=1}^n \int_{t_{k-1}}^{t_k} \frac{\frac{d}{ds} \eta_y^i(s)(1 - 2\eta_y^i(s))}{1 - \eta_y^i(s)} ds \right) \\ \Lambda_\beta &= \begin{pmatrix} \frac{1}{\sigma^2} \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \frac{1}{\eta_y^i(s)(1 - \eta_y^i(s))} ds & -\frac{1}{\sigma^2} \sum_{k=1}^n \int_{t_{k-1}}^{t_k} \frac{1}{1 - \eta_y^i(s)} ds \\ -\frac{1}{\sigma^2} \sum_{k=1}^n \int_{t_{k-1}}^{t_k} \frac{1}{1 - \eta_y^i(s)} ds & \frac{1}{\sigma^2} \sum_{k=1}^n \int_{t_{k-1}}^{t_k} \frac{\eta_y^i(s)}{1 - \eta_y^i(s)} ds \end{pmatrix} \end{aligned}$$

El Algoritmo Gibbs Sample para estimar los parámetros $\alpha = (\alpha_1, \alpha_2)$ y reconstruir los caminos procede de la siguiente manera:

Definición 4.1 (Algoritmo Gibbs Sample)

1. Generar $\alpha \sim N_{T(0)}(\alpha_0, \Sigma_0)$ (Normal truncada en cero) y $\beta \sim \exp\{\beta_0\}$
Hacer $k = 1$ y $\theta^k = (\alpha^k, \beta^k)$, con $\alpha^k = \alpha$ y $\beta^k = \beta$
2. Generar z_t^i con $i = 1, \dots, n$. Dados X^d (Observaciones) y θ^k
3. $\alpha^{k+1} \sim N_{T(0)}((\nu_\beta + \alpha_0 \Sigma_0^{-1})(\Lambda_\beta + \Sigma_0^{-1})^{-1}, (\Lambda_\beta + \Sigma_0^{-1})^{-1})$
 $\beta^{k+1} \sim \pi_0(\theta_\beta) \mathcal{L}(\theta^k; Y_{obs, Y_{mis}})$
4. $k = k + 1$ y regresamos al paso 2

Se simuló una trayectoria estocástica del proceso de Jacobi con parámetros fijos conocidos ($\alpha_1 = 4,5, \alpha_2 = 15, \sigma = 0,1$) utilizando el esquema de Milstein, representando el comportamiento ideal de un KRI durante un periodo de observación. Posteriormente se eliminó el 90% de los datos de la trayectoria generada, simulando un escenario severo de pérdida de información.

El algoritmo de Gibbs Sample se ejecutó durante 1,000 iteraciones para asegurarla estabilidad de la estimación, se aplicó un análisis visual de las cadenas para determinar el punto de convergencia y descartar el periodo de calentamiento. La eficacia del modelo se evaluará contrastando la capacidad de reconstrucción de la trayectoria y la precisión de los estimadores frente a los parámetros reales y la estimación clásica.

Finalmente se emplearon datos sintéticos que replican el comportamiento de la red local de una organización, con el propósito de caracterizar el perfil de riesgo asociado a este indicador. El análisis de los hallazgos y su interpretación operativa se presentan en la siguiente sección.

5. Análisis de resultados

A continuación se presentan los resultados obtenidos al aplicar el modelo de Jacobi a los datos sintéticos y los relacionados a la capacidad de almacenamiento de backups. El análisis abarca la validación de la generación de trayectorias, la precisión en la reconstrucción de las mismas y la estimación de parámetros, concluyendo con la interpretación de las implicaciones operativas para la organización.

Para evaluar el comportamiento del modelo, se generaron trayectorias sintéticas utilizando el método de Milstein con los parámetros $\alpha_1 = 4.5$, $\alpha_2 = 15$ y $\sigma = 0.2$ empleando una discretización de 1,000 puntos equidistantes.

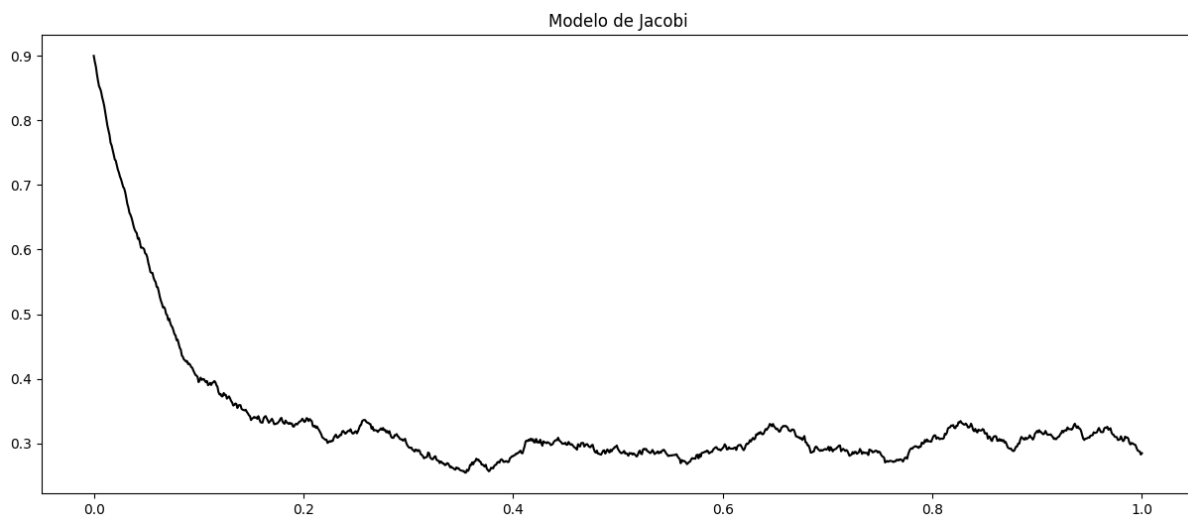


Figura 1: Grafico de trayectoria simulada con Milstein

La trayectoria simulada refleja adecuadamente la dinámica esperada del proceso de Jacobi, manteniéndose dentro del intervalo $(0,1)$ y mostrando reversión a la media consistente con los parámetros especificados

Posteriormente, la trayectoria completa se redujo al 10 % de sus puntos, a partir de esta versión incompleta, se reconstruyó la trayectoria completa utilizando puentes estocásticos con los parámetros verdaderos del modelo.

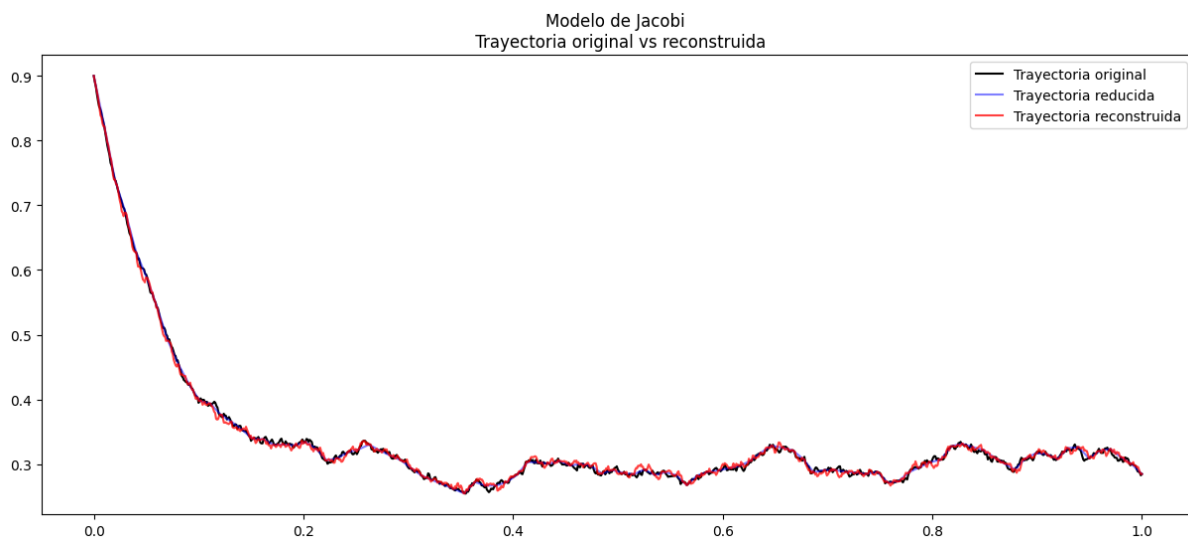


Figura 2: Validación de reconstrucción de trayectorias mediante puentes estocásticos

La trayectoria reconstruida reproduce de manera eficiente la tendencia general y las fluctuaciones principales de la trayectoria original, mostrando la efectividad de los puentes estocásticos cuando tenemos información incompleta.

En cada trayectoria se estimaron los parámetros utilizando regresión lineal sobre la esperanza condicional para los parámetros α_1 y α_2 , mientras que σ se estimó mediante variación cuadrática. A continuación se muestran los resultados:

Parámetros	Reales	Completa	Reducida	Reconstruida
α_1	4.5	4.479960	4.165158	4.587631
α_2	15.0	15.221981	14.138781	15.539451
σ	0.2	0.225064	0.415448	0.235087

Cuadro 1: Comparación de parámetros por trayectoria.

Para obtener intervalos de los parámetros y una estimación más robusta en presencia de información limitada, se aplicó muestreo de Gibbs sobre la trayectoria reducida. Se estimaron α_1 y α_2 y se construyeron intervalos que contienen el 90 % de los puntos de las simulaciones generadas.

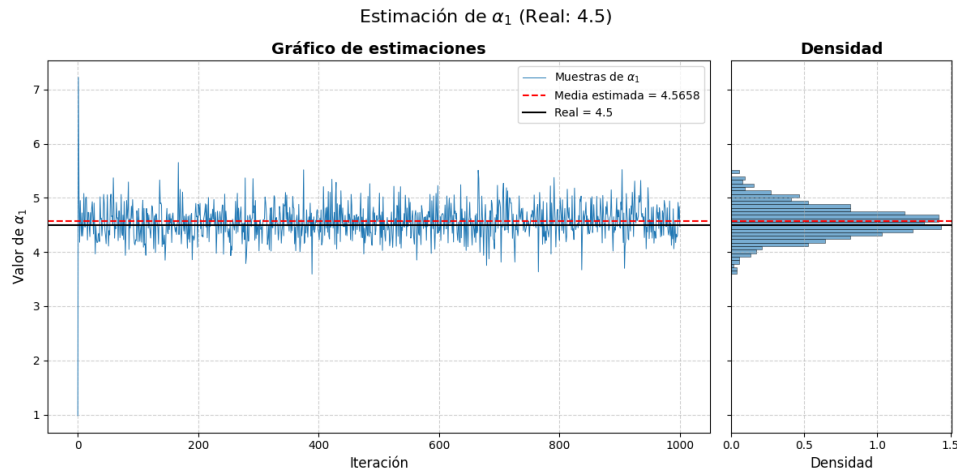


Figura 3: Estimación α_1

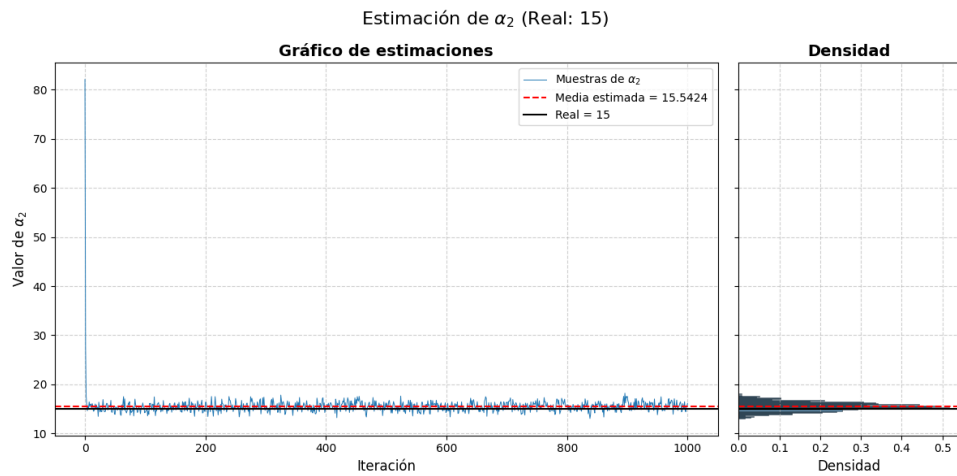


Figura 4: Estimación α_2

Posteriormente se aplicó el mismo esquema bayesiano al indicador correspondiente a la capacidad de

Parámetros	Reales	Gibbs Sample	IC Inferior	IC Superior
a_1	4.5	4.565757	4.111996	5.0645
a_2	15.0	15.542387	14.09607	17.056751

Cuadro 2: Estimación de α_1 y α_2

almacenamiento del sistema de backups, el cual tiene la siguiente tendencia obtenida con datos capturados mensualmente.

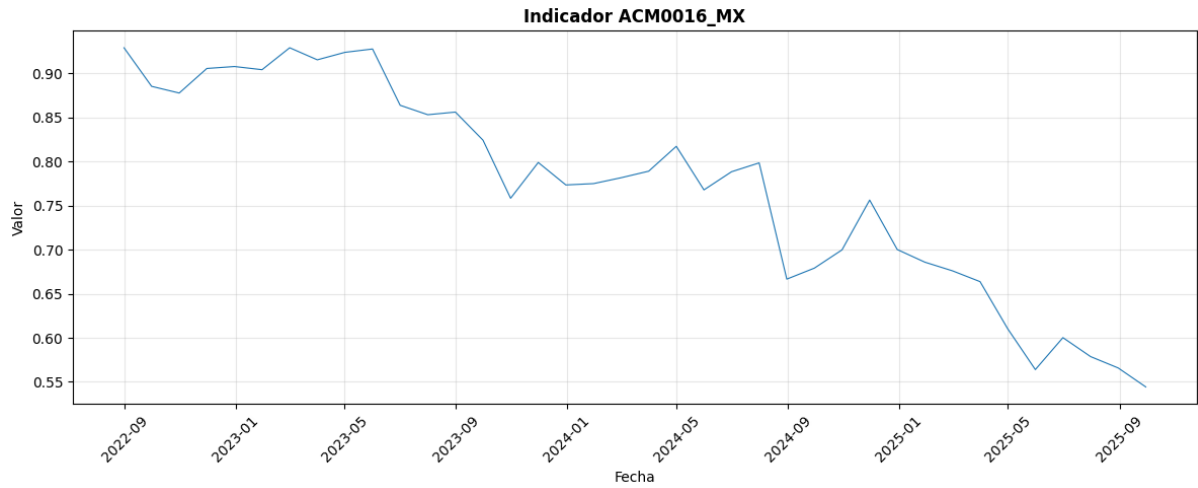


Figura 5: Tendencia del indicador

A partir de los datos históricos se estimaron los parámetros α_1 y α_2 , mediante muestreo de Gibbs. Para σ se utilizó el estimador por variación cuadrática, obteniendo los siguientes resultados.

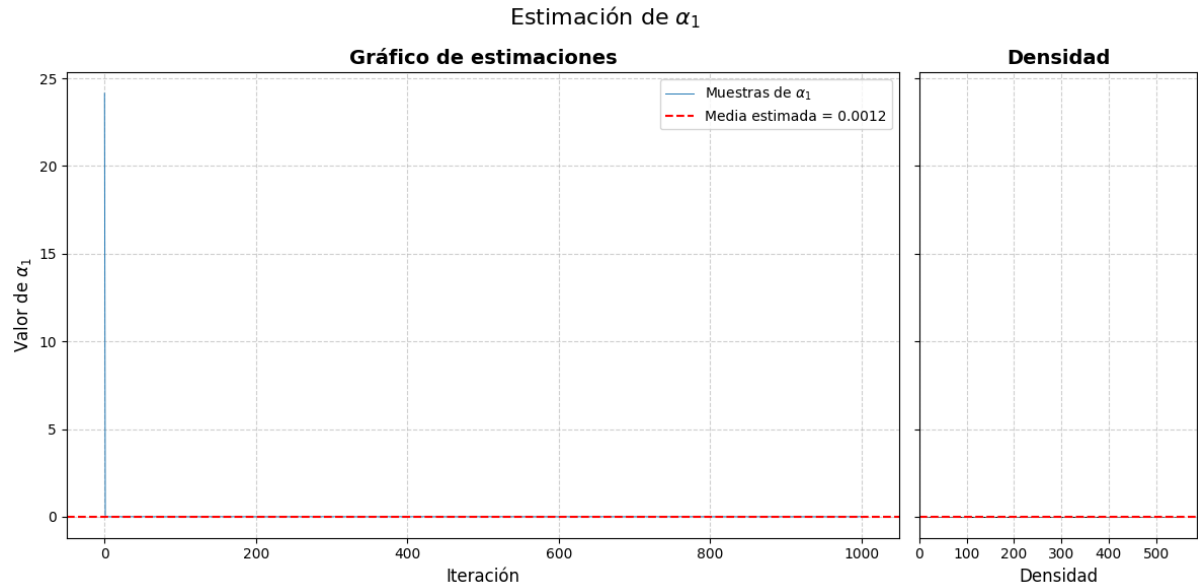


Figura 6: Estimación de α_1 para el indicador



Figura 7: Estimación de α_2 para el indicador

Parámetro	Gibbs Sample	IC Inferior	IC Superior
α_1	0.001220	0.000096	0.002905
α_2	0.001877	0.000475	0.003965
σ	0.016922	na	na

Cuadro 3: Estimaciones mediante muestreo de Gibbs y sus intervalos de confianza.

Dado que estos corresponden a la parametrización alternativa, se recuperan los parámetros originales del modelo de Jacobi mediante:

$$\theta = \alpha_2, \quad \mu = \frac{\alpha_1}{\alpha_2}$$

Tenemos que:

Estimador	α_1	α_2	σ	μ	θ
Gibbs Sample	0.00122	0.001877	0.016922	0.650007	0.001877
IC Inferior	0.000096	0.000475	na	0.201829	0.000475
IC Superior	0.002905	0.003965	na	0.732573	0.003965

Cuadro 4: Estimaciones de los parámetros del modelo original.

Con estos valores se calculó la vida media $t_{1/2} = \frac{\ln(2)}{\theta}$ que representa el tiempo necesario para que el indicador reduzca a la mitad su distancia al nivel de equilibrio en ausencia de difusión. El valor obtenido es de 369.25.

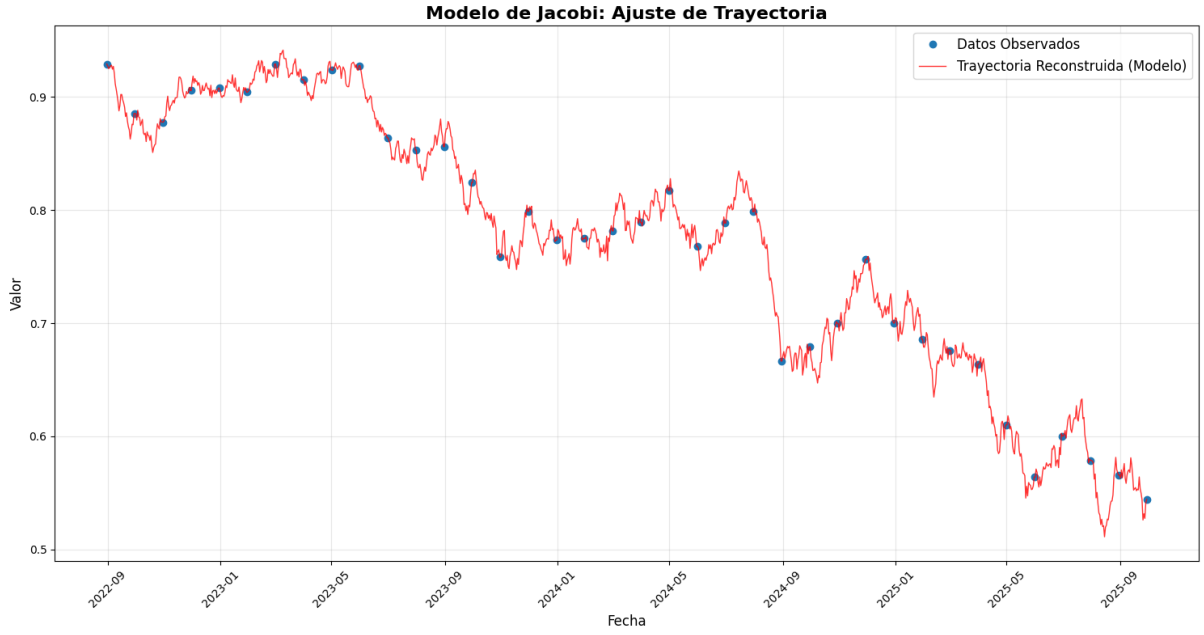


Figura 8: Trayectoria reconstruida

Para estimar la probabilidad de que el indicador presente alertas en los siguientes seis meses, se generaron 500 trayectorias simuladas con los parámetros estimados iniciando en el último valor registrado del indicador.

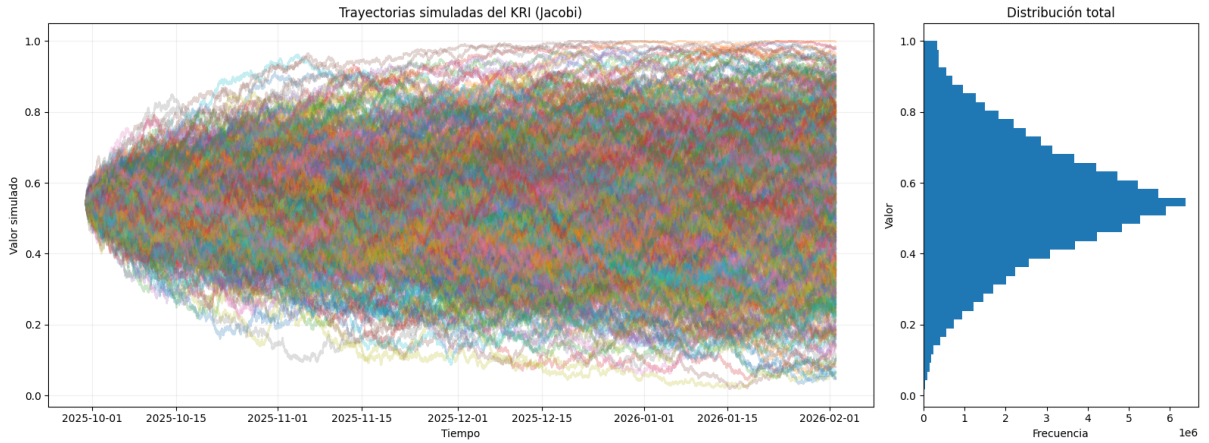


Figura 9: Predicciones de la trayectoria

A partir de estas simulaciones obtenemos que la probabilidad de alertamiento del indicador en amarillo es de , y de alertamiento en rojo es de .

Finalmente se verificó si los parámetros satisfacen las condiciones tipo Feller:

$$2\theta\mu \geq \sigma^2, \quad 2\theta(1-\mu) \geq \sigma^2.$$

Sustituyendo los valores estimados:

$$2\theta\mu \approx 0.0024, \quad \sigma^2 \approx 0.00028, \quad 2\theta(1-\mu) \approx 0.0013.$$

Dado que ambas desigualdades se cumplen, la difusión nunca domina al drift y el proceso permanece estrictamente dentro del intervalo $(0, 1)$.

Finalmente se analizó la distribución de $z = y_i - y_{i-1} - \alpha(y_{i-1})$ mediante la prueba Kolmogorov Smirnov Lilliefors. El resultado arrojó un p-value de 0.102, por lo que no se rechaza la hipótesis nula de normalidad al nivel de significancia del 5%. Este resultado nos dice que los residuos se comportan de manera consistente con los incrementos normales esperados en un proceso de difusión correctamente especificado.

A continuación se presentan las gráficas obtenidas:

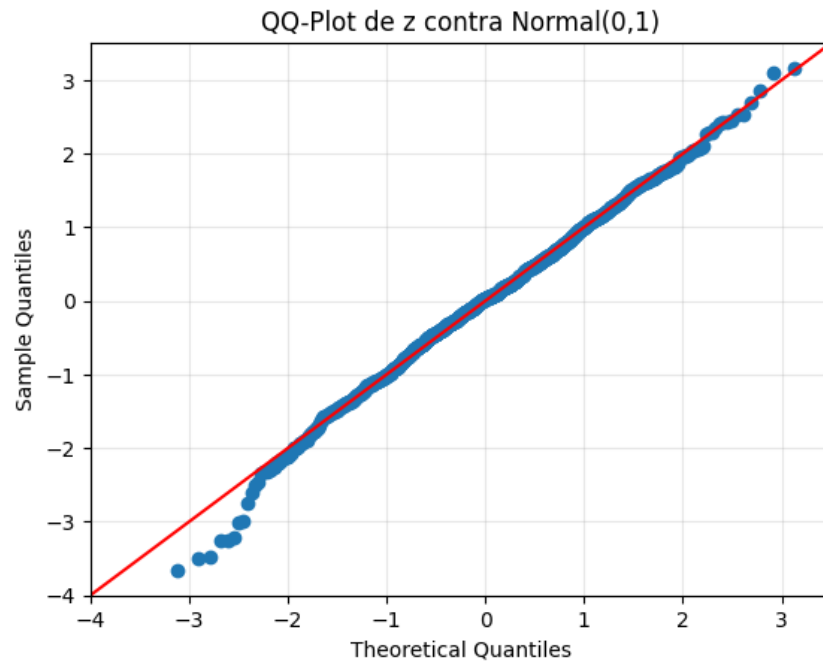


Figura 10: QQ-plot de residuos del modelo

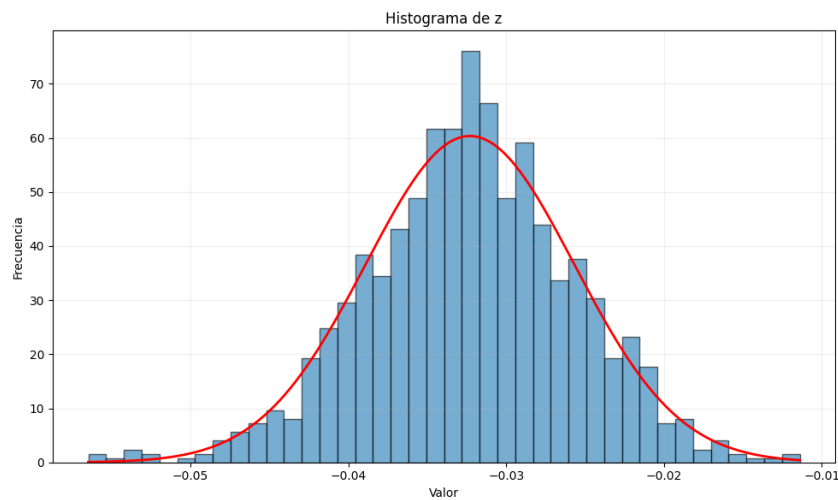


Figura 11: Histograma de residuos del modelo

6. Conclusiones

En este trabajo se analizó la aplicabilidad del proceso de Jacobi para modelar indicadores acotados en $[0, 1]$, considerando su estructura cuando la información es incompleta y aplicando esquemas de reconstrucción y simulación. Las trayectorias generadas mediante el método de Milstein reprodujeron adecuadamente la dinámica teórica del proceso, y la reconstrucción a través de puentes estocásticos nos dio una recuperación precisa de la información faltante. Además, la verificación estadística respalda que la especificación del modelo es adecuada, ya que los residuos presentan un comportamiento coherente con los incrementos gaussianos característicos de un proceso de difusión correctamente elegido.

La estimación de parámetros mediante regresión, variación cuadrática y muestreo de Gibbs arrojó valores coherentes con los datos sintéticos. En el caso del indicador, los parámetros estimados indican una reversión a la media lenta (369 días de vida media), lo cual sugiere una baja resiliencia operativa ante un evento. También presenta un nivel estacionario alrededor de 0.65 y una volatilidad moderada de 0.017, lo cual muestra que la infraestructura actual es adecuada y mantiene un colchón que evite llegar a una alerta amarilla (0.8) bajo condiciones normales. Además las condiciones tipo Feller se satisfacen en ambos extremos del intervalo, garantizando que el proceso no llegará a ninguna de las fronteras bajo los parámetros estimados.

Las simulaciones permitieron estimar la probabilidad de que el proceso alcance niveles elevados en un horizonte de seis meses, obteniendo una probabilidad aproximada de 7% para superar el umbral de alerta en amarillo (0.8) y de 1.9% para el umbral de alerta en rojo (0.9). En conjunto, los resultados muestran que bajo condiciones normales, la capacidad del sistema de backups permanecerá estable, aunque presentará dificultad para regresar a su nivel típico tras eventos que afecten la operación.

7. Bibliografía

Referencias

- [1] Baltazar-Larios, F. (2025). *Inferencia estadística para ecuaciones diferenciales estocásticas*. Facultad de Ciencias, UNAM.
- [2] Baltazar-Larios, F., & López Ortega, S. I. (2024). *Simulación estocástica*. Facultad de Ciencias, UNAM.
- [3] Basel Committee on Banking Supervision. (2021). *Principles for Operational Resilience*. Bank for International Settlements.
- [4] Moosa, I. (2007). *Operational Risk Management*. Palgrave Macmillan.

8. Apéndice. Repositorio de código

El código utilizado para la simulación de trayectorias, estimación de parámetros, métodos de reconstrucción y generación de gráficas se encuentra disponible en un repositorio público de GitHub. Este repositorio contiene:

- Scripts de simulación del proceso de Jacobi.
- Código para el cálculo de momentos, variación cuadrática y estimación bayesiana.
- Implementaciones de puentes estocásticos.
- Generación de figuras y experimentos reproducibles.
- Aplicación a datos del KRI de capacidad de backups

Los datos empleados fueron perturbados mediante ruido estocástico y censurados parcialmente para garantizar la protección de la información sensible de la organización.

El repositorio puede consultarse en el siguiente enlace, donde también se encuentra el modelo relacionado con la simulación de eventos de riesgo operacional:

<https://github.com/Yael01livaes17/Proyecto>