

Tratado de Modelos Matemáticos de Predictores Estocásticos Universales (PEU)

Versión Extendida y Unificada

Consortio de Desarrollo de Meta-Predicción Adaptativa

18 de febrero de 2026

Índice

1	Fundamentación Teórica y Arquitectura	3
1.1	Espacios de Probabilidad y Filtraciones	3
1.2	El Meta-Estado del Sistema (Ξ_t)	3
1.3	Problema de Predicción Óptima	3
1.4	Arquitectura del Sistema Universal	4
2	Fase 1: Motor de Identificación de Sistemas (SIA)	5
2.1	Formalización del Vector de Estado Funcional	5
2.2	Análisis de Estacionariedad y Ergodicidad	5
2.2.1	Estacionariedad Fuerte	5
2.2.2	Integración Fraccionaria y Diferenciación	5
2.3	Análisis de Regularidad Hölderiana	5
2.3.1	Espectro de Singularidades Local	5
2.4	Descomposición de Semimartingalas	6
2.4.1	Variación Cuadrática	6
2.4.2	Teorema de Bichteler-Dellacherie	6
2.5	Operadores Espectrales y Flujo de Información	6
2.5.1	Operador de Koopman (\mathcal{K})	6
2.5.2	Engrosamiento de Filtraciones (Grossissement de Filtration)	6
3	Fase 2: Formalización de los Núcleos de Predicción	7
3.1	Rama A: Proyección en Espacios de Hilbert	7
3.1.1	Principio de Ortogonalidad y Wiener-Hopf	7
3.1.2	Condición de Paley-Wiener	7
3.2	Cálculo de Malliavin y Sensibilidad Estocástica	7
3.2.1	Teorema de Representación de Ocone-Haussmann	7
3.3	Rama B: Ecuaciones de Evolución y Viscosidad	8
3.3.1	Generador Infinitesimal	8
3.3.2	Soluciones de Viscosidad de Crandall-Lions	8
3.3.3	Principio de Entropía para Aproximación Neuronal (DGM)	8
3.4	Cálculo de Malliavin en Espacios de Poisson	9
3.4.1	Fórmula de Itô para Semimartingalas con Saltos	10
3.5	Rama D: Signature y Rough Paths (Invariancia Topológica)	10
3.5.1	El Espacio de Caminos Rugosos Geométricos	10
3.5.2	La Signatura (\mathcal{S}) y Álgebra de Hopf	10
3.5.3	Lema de Invariancia bajo Reparametrización	10
3.5.4	Predictor T-Linear	10

4	Fase 3: Orquestador de Pesaje Adaptativo	11
4.1	Dinámica de Transporte Óptimo y Geometría de Wasserstein	11
4.2	Grandes Desviaciones y Principio de Contracción	11
4.3	Acoplamiento Geométrico y Métrica Fisher-Rao	11
4.4	Funcional de Lyapunov Global	12
5	Fase 4: Convergencia y Estabilidad Global	13
5.1	Condiciones de Mezclado (Mixing)	13
5.2	Teorema de Sanov y Grandes Desviaciones	13
5.3	Estabilidad en Media L^p y Lyapunov	13
5.4	Complejidad Secuencial y Cotas de Generalización	13
5.5	Detección de Puntos de Cambio y Tiempos de Parada	14
5.5.1	Umbral Adaptativo para Colas Pesadas	14
6	Ecuación Diferencial Operacional Unificada	16
6.1	Dinámica del Meta-Estado	16
6.2	Teorema de Existencia y Unicidad Global	16
A	Postulado de Robustez ante Singularidades	17

Capítulo 1

Fundamentación Teórica y Arquitectura

Este tratado formaliza la construcción de un sistema de predicción estocástica capaz de operar sobre procesos dinámicos cuya ley de probabilidad subyacente es desconocida *a priori*.

1.1 Espacios de Probabilidad y Filtraciones

Definimos un espacio de probabilidad completo (Ω, \mathcal{F}, P) . La evolución de la información se modela mediante una filtración $\{\mathcal{F}_t\}_{t \geq 0}$ que satisface las *condiciones habituales* (usual conditions) de Dellacherie-Meyer:

1. **Complejitud:** \mathcal{F}_0 contiene todos los conjuntos P -nulos de \mathcal{F} .
2. **Continuidad por la derecha:** $\mathcal{F}_t = \bigcap_{s > t} \mathcal{F}_s$ para todo $t \geq 0$.

Esto garantiza que los tiempos de parada (como los definidos por el algoritmo CUSUM) sean medibles y el proceso admita modificaciones càdlàg.

1.2 El Meta-Estado del Sistema (Ξ_t)

Para unificar la dinámica de control y predicción, definimos el meta-estado en el tiempo t como la terna funcional en un espacio de Banach:

$$\Xi_t = (V_s(t), w_t, \mathcal{P}_h^\cup)$$

Donde V_s es el estado de identificación, w_t la distribución de pesos en la variedad estadística, y \mathcal{P}_h^\cup el operador de predicción activo.

1.3 Problema de Predicción Óptima

Definición 1.1 (Problema de Predicción Óptima). *Dado un proceso estocástico $X = \{X_t : t \in T\}$, busquemos el operador \mathcal{P}_h tal que $\hat{X}_{t+h} = \mathcal{P}_h(X_s, s \leq t)$ minimice la norma en $L^2(\Omega, \mathcal{F}, P)$:*

$$\hat{X}_{t+h} = \underset{Z \in L^2(\mathcal{F}_t)}{\operatorname{argmin}} E [\|X_{t+h} - Z\|^2] = E[X_{t+h} \mid \mathcal{F}_t]$$

1.4 Arquitectura del Sistema Universal

El sistema se estructura en tres fases operativas:

1. **Motor de Identificación (SIA):** Operador funcional $\Psi(X) \rightarrow \mathcal{C}$. Para garantizar la continuidad de los operadores de control en procesos multifractales, definimos el espacio de llegada \mathcal{C} como el espacio de Besov $B_{p,q}^s(\mathbb{R})$, que permite caracterizar singularidades locales a través de descomposiciones en wavelets.
2. **Núcleos de Predicción (\mathcal{P}_i):** Ramas A (Hilbert), B (Markov/Fokker-Planck), C (Itô/Lévy), D (Rough Paths/Signature).
3. **Orquestador Adaptativo (\mathcal{O}):** Dinámica de transporte óptimo en el espacio de medidas de probabilidad $\mathcal{P}_2(\Omega)$ dotado con la métrica de Wasserstein.

Capítulo 2

Fase 1: Motor de Identificación de Sistemas (SIA)

El SIA caracteriza la topología del proceso mediante un vector de estado funcional V_s .

2.1 Formalización del Vector de Estado Funcional

El vector $V_s(t)$ consolida las métricas estructurales del proceso:

$$V_s(t) = [d(t), \alpha(t), \sigma(\mathcal{K}), \mathcal{T}_{Y \rightarrow X}, [X]_t]^\top \in \mathcal{C}$$

2.2 Análisis de Estacionariedad y Ergodicidad

2.2.1 Estacionariedad Fuerte

El operador Ψ verifica la invariancia de la medida imagen bajo el grupo de traslaciones temporales $\{\theta_\tau\}_{\tau \in \mathbb{R}}$:

$$P \circ \theta_\tau^{-1} = P \quad \forall \tau$$

2.2.2 Integración Fraccionaria y Diferenciación

Para procesos con memoria larga, definimos el operador inverso del núcleo de Riesz unilateral I^α :

$$Y_t = D^\alpha X_t = \frac{1}{\Gamma(-\alpha)} \int_{-\infty}^t (t-s)^{-\alpha-1} X_s ds$$

Esto generaliza el operador $(1-L)^d$ al continuo.

2.3 Análisis de Regularidad Hölderiana

2.3.1 Espectro de Singularidades Local

La regularidad local se caracteriza por el exponente puntual de Hölder $\alpha(t_0)$, definido como el supremo de los α tales que::

$$\limsup_{\epsilon \rightarrow 0} \frac{|X(t_0 + \epsilon) - X(t_0)|}{|\epsilon|^\alpha} < \infty$$

La función $\alpha(t)$ induce una estratificación del dominio temporal $\bigcup_h \{t : \alpha(t) = h\}$.

2.4 Descomposición de Semimartingalas

2.4.1 Variación Cuadrática

Se define el proceso de variación cuadrática como el límite uniforme en probabilidad:

$$[X]_t = P - \lim_{\|\Pi\| \rightarrow 0} \sum_i (X_{t_i} - X_{t_{i-1}})^2$$

2.4.2 Teorema de Bichteler-Dellacherie

Si X_t es un integrador L^0 estocástico, admite la descomposición canónica:

$$X_t = X_0 + M_t + A_t$$

donde M_t es una martingala local y A_t es un proceso de variación finita predecible.

2.5 Operadores Espectrales y Flujo de Información

2.5.1 Operador de Koopman (\mathcal{K})

Definimos el operador de composición sobre el espacio de observables $L^\infty(\Omega)$:

$$\mathcal{K}^t g(\omega) = g(\theta_t \omega)$$

El espectro puntual $\sigma_p(\mathcal{K})$ caracteriza las invariantes ergódicas del sistema dinámico.

2.5.2 Engrosamiento de Filtraciones (Grossissement de Filtration)

Sea $\mathbb{G} = \{\mathcal{G}_t\}_{t \geq 0}$ una ampliación de la filtración original \mathbb{F} tal que $\mathcal{F}_t \subset \mathcal{G}_t$. Según el Teorema de Jeulin-Yor, si la hipótesis (H) se viola, la \mathbb{F} -martingala M_t se descompone en \mathbb{G} como:

$$M_t = \tilde{M}_t + \int_0^t \alpha_s ds$$

donde \tilde{M} es una \mathbb{G} -martingala y α_s es el proceso de deriva de información. Esto formaliza la asimilación de variables latentes exógenas.

Capítulo 3

Fase 2: Formalización de los Núcleos de Predicción

3.1 Rama A: Proyección en Espacios de Hilbert

El predictor se define en el espacio $\mathcal{H}_t = \overline{\text{span}}\{X_s : s \leq t\}$.

3.1.1 Principio de Ortogonalidad y Wiener-Hopf

El error de predicción debe ser ortogonal a la historia pasada:

$$\langle X_{t+h} - \hat{X}_{t+h}, X_s \rangle = 0 \quad \forall s \leq t$$

Esto conduce a la Ecuación Integral de Wiener-Hopf para el núcleo de impulso óptimo $h(\tau)$:

$$\gamma(t+h-s) = \int_0^\infty h(\tau)\gamma(s-\tau)d\tau, \quad s \geq 0$$

3.1.2 Condición de Paley-Wiener

Para garantizar la causalidad y existencia de la factorización espectral $f(\lambda) = |\Psi(i\lambda)|^2$, se requiere:

$$\int_{-\infty}^\infty \frac{|\log f(\lambda)|}{1+\lambda^2} d\lambda < \infty$$

3.2 Cálculo de Malliavin y Sensibilidad Estocástica

Consideramos el espacio de Wiener canónico (Ω, \mathcal{F}, P) y el subespacio de Cameron-Martin $H = L^2([0, T])$. Definimos el operador derivada de Malliavin $D : \mathbb{D}^{1,2} \rightarrow L^2(\Omega; H)$ sobre funcionales cilíndricos $F = f(W(h_1), \dots, W(h_n))$ como:

$$D_t F = \sum_{i=1}^n \partial_i f(W(h_1), \dots, W(h_n)) h_i(t)$$

3.2.1 Teorema de Representación de Ocone-Haussmann

Todo funcional $F \in \mathbb{D}^{1,2}$ admite la representación integral:

$$F = E[F] + \int_0^T E[D_t F | \mathcal{F}_t] dW_t$$

Esto permite caracterizar explícitamente el integrando en la descomposición de martingala del predictor óptimo como la proyección condicional de la derivada de Malliavin.

3.3 Rama B: Ecuaciones de Evolución y Viscosidad

3.3.1 Generador Infinitesimal

La evolución de la densidad de probabilidad $p(x, t)$ se rige por el operador adjunto \mathcal{L}^* . Consideramos la función valor $V(t, x)$ asociada al control estocástico óptimo, la cual satisface la ecuación de Hamilton-Jacobi-Bellman (HJB):

$$-\partial_t V + \inf_{u \in U} \{-\mathcal{L}^u V - f(x, u)\} = 0$$

3.3.2 Soluciones de Viscosidad de Crandall-Lions

Dado que V puede no ser diferenciable en C^2 , definimos la solución en el sentido de viscosidad. Una función semicontinua superior u es subsolución de viscosidad de $F(x, u, Du, D^2u) = 0$ si para toda $\phi \in C^2$ tal que $u - \phi$ tiene un máximo local en x_0 :

$$F(x_0, u(x_0), D\phi(x_0), D^2\phi(x_0)) \leq 0$$

Esta formulación garantiza la existencia y unicidad de la solución para hamiltonianos degenerados típicos en predicción robusta.

3.3.3 Principio de Entropía para Aproximación Neuronal (DGM)

En la implementación numérica de la Rama B mediante el Deep Galerkin Method (DGM), la red neuronal $V_\theta(t, x)$ aproxima la función valor. Para garantizar que la solución neuronal sea no degenerada y capture la estructura esencial de la PDE, establecemos un criterio de entropía.

Teorema 3.1 (Principio de Conservación de Entropía de Solución). *Sea $V_\theta : [0, T] \times \Omega \rightarrow \mathbb{R}$ la aproximación neuronal de la función valor que satisface la ecuación HJB, y sea $g : \Omega \rightarrow \mathbb{R}$ la condición terminal. Definimos la entropía diferencial de la solución en tiempo t como:*

$$H_t[V_\theta] = - \int_{\Omega} p_t(v) \log p_t(v) dv$$

donde $p_t(v)$ es la densidad de probabilidad empírica de los valores $\{V_\theta(t, x_i)\}_{i=1}^N$ sobre una malla de puntos $\{x_i\}$ que discretiza el dominio Ω .

Para que la solución neuronal sea admisible, debe satisfacer el criterio de conservación de entropía:

$$\frac{1}{T} \int_0^T H_t[V_\theta] dt \geq \gamma \cdot H[g]$$

donde:

- $H[g] = - \int p_g(v) \log p_g(v) dv$ es la entropía de la condición terminal
- $\gamma \in [0.5, 1.0]$ es el factor de retención de entropía

Este criterio previene el colapso de modo, donde la red neuronal converge a una solución constante o de varianza mínima que satisface trivialmente la PDE.

Demostración. La función valor $V(t, x)$ hereda estructura informativa de la condición terminal $g(x)$ mediante el principio de programación dinámica. Formalmente, para un problema de control óptimo con horizonte finito T :

$$V(t, x) = \inf_{u \in \mathcal{U}} E \left[\int_t^T f(X_s^{t,x,u}, u_s) ds + g(X_T^{t,x,u}) \mid \mathcal{F}_t \right]$$

La condicional expectativa es un operador de contracción en el espacio L^2 , pero preserva información en el siguiente sentido: si g tiene soporte disperso (alta entropía), entonces $V(t, \cdot)$ para $t < T$ también debe exhibir variabilidad espacial proporcional.

Más precisamente, consideremos la evolución backward de la entropía. Si $V(t, x)$ fuera constante en x para algún $t < T$, entonces:

$$\nabla_x V(t, x) \equiv 0 \implies u^*(t, x) = \arg \max_u \{b(x, u) \cdot 0 + \dots\}$$

lo cual implica que la política óptima u^* no depende del estado x . Esto contradice la no-trivialidad de $g(x)$ (que es no constante por hipótesis).

La desigualdad de entropía se obtiene aplicando la desigualdad de Jensen a la función cóncava $-x \log x$ combinada con el teorema de comparación para soluciones de viscosidad, que garantiza que si V_θ aproxima bien a V en norma L^∞ , entonces sus distribuciones inducidas son cercanas en el sentido de Kullback-Leibler, y por tanto sus entropías son comparables.

Formalmente, bajo la métrica de Wasserstein entre las medidas pushforward:

$$W_2(V_\theta(\cdot, t)_{\#}\mu, V(\cdot, t)_{\#}\mu) < \epsilon \implies |H[V_\theta(\cdot, t)] - H[V(\cdot, t)]| < C\epsilon$$

donde μ es la medida espacial en Ω y C es una constante que depende del dominio. \square

Nota 3.1. En la práctica, el criterio de entropía se monitorea durante el entrenamiento de la red DGM. Si $H_t[V_\theta]$ cae por debajo del umbral $\gamma H[g]$ de manera persistente, se recomienda:

1. Aumentar la capacidad de la red (más capas o neuronas)
2. Ajustar la tasa de aprendizaje para evitar convergencia prematura a mínimos locales triviales
3. Modificar la función de pérdida para incluir un término de regularización entrópica:

$$\mathcal{L}_{total} = \mathcal{L}_{PDE} + \mathcal{L}_{BC} - \lambda H_t[V_\theta]$$

donde $\lambda > 0$ penaliza soluciones de baja entropía

Corolario 3.1 (Relación con Varianza de Solución). En el caso de condiciones terminales con distribución aproximadamente Gaussiana, la entropía diferencial está relacionada con la varianza por:

$$H[X] = \frac{1}{2} \log(2\pi e \sigma^2)$$

Por tanto, el criterio de conservación de entropía implica la conservación de varianza:

$$\text{Var}_x[V_\theta(t, x)] \geq \gamma' \cdot \text{Var}[g(x)]$$

con $\gamma' = \exp(2\gamma - 1)$. Este es precisamente el test de colapso de modo descrito en el documento de pruebas.

3.4 Cálculo de Malliavin en Espacios de Poisson

Para procesos con componentes de salto (Rama C), extendemos el operador de derivada D_t al espacio canónico de Poisson $(\Omega, \mathcal{F}, P, N)$. Definimos el operador de diferencia $\mathcal{D}_{t,z}$ para funcionales $F \in \mathbb{D}^{1,2}(\mu)$:

$$\mathcal{D}_{t,z}F(\omega) = F(\omega + \delta_{(t,z)}) - F(\omega)$$

El integrando de la representación previsible para la martingala de saltos puros es dada por la proyección previsible de $\mathcal{D}_{t,z}F$.

3.4.1 Fórmula de Itô para Semimartingalas con Saltos

El proceso X_t se descompone según la estructura canónica de Lévy-Itô:

$$X_t = X_0 + \int_0^t b(X_{s-})ds + \int_0^t \sigma(X_{s-})dW_s + \int_0^t \int_{\mathbb{R}^n} z \tilde{N}(ds, dz)$$

La esperanza condicional $u(t, x)$ satisface la ecuación integro-diferencial parcial (PIDE) asociada al generador \mathcal{L}^ν :

$$\mathcal{L}^\nu \phi(x) = \frac{1}{2} \sigma^2 \Delta \phi + b \cdot \nabla \phi + \int_{\mathbb{R}^d} [\phi(x+z) - \phi(x) - z \cdot \nabla \phi \mathbb{1}_{|z| \leq 1}] \nu(dz)$$

3.5 Rama D: Signature y Rough Paths (Invariancia Topológica)

Para procesos donde la rugosidad de la trayectoria imposibilita el cálculo estocástico estándar (exponente de Hölder $H \leq 1/2$, variación $p \geq 2$), operamos en el marco de la Teoría de Caminos Rugosos de Lyons.

3.5.1 El Espacio de Caminos Rugosos Geométricos

Sea \mathbf{X} un proceso continuo con valores en el álgebra tensorial truncada $T^{(N)}(\mathbb{R}^d) = \bigoplus_{k=0}^N (\mathbb{R}^d)^{\otimes k}$. Definimos el espacio de caminos rugosos geométricos de p -variación finita $G\Omega_p(\mathbb{R}^d)$ como la clausura de los caminos suaves bajo la p -variación métrica:

$$d_p(\mathbf{X}, \mathbf{Y}) = \max_{k=1, \dots, [p]} \sup_{\mathcal{D}} \left(\sum_i |\mathbf{X}_{t_i, t_{i+1}}^k - \mathbf{Y}_{t_i, t_{i+1}}^k|^{p/k} \right)^{k/p}$$

3.5.2 La Signatura (\mathcal{S}) y Álgebra de Hopf

El mapa de signatura $\mathcal{S} : G\Omega_p([0, T], \mathbb{R}^d) \rightarrow T((\mathbb{R}^d))$ transforma la trayectoria en una serie formal de potencias no conmutativas (Serie de Chen):

$$\mathcal{S}(\mathbf{X})_{0,t} = 1 + \sum_{k=1}^{\infty} \int_{0 < u_1 < \dots < u_k < t} dX_{u_1} \otimes \dots \otimes dX_{u_k}$$

El espacio imagen es un grupo de Lie bajo la operación \otimes , y sus elementos satisfacen la propiedad del *Shuffle Product* para funcionales lineales duales $f, g \in T((\mathbb{R}^d))^*$:

$$\langle f, \mathbf{X} \rangle \langle g, \mathbf{X} \rangle = \langle f \amalg g, \mathbf{X} \rangle$$

Esto permite aproximar cualquier funcional continuo mediante combinaciones lineales de términos de la signatura (Teorema de Aproximación Universal).

3.5.3 Lema de Invariancia bajo Reparametrización

Lema 3.1. *La firma $\mathcal{S}(X)$ es invariante ante cualquier reparametrización temporal monótona $\psi(t)$:*

$$\mathcal{S}(X \circ \psi)_{0, T'} = \mathcal{S}(X)_{0, T}$$

Esto inmuniza a la Rama D contra el ruido de muestreo irregular, permitiendo una caracterización topológica pura.

3.5.4 Predictor T-Linear

El predictor se formaliza como un funcional lineal en el espacio de tensores:

$$\hat{X}_{t+h} = \langle W, \mathbf{X}_{0,t} \rangle$$

Capítulo 4

Fase 3: Orquestador de Pesaje Adaptativo

El Orquestador \mathcal{O} gestiona la mezcla convexa $\hat{X}_{t+h}^{PEU} = \sum w_i(t) \hat{X}_{t+h}^{(i)}$.

4.1 Dinámica de Transporte Óptimo y Geometría de Wasserstein

Consideramos la variedad Riemanniana de dimensión infinita $\mathcal{M} = (\mathcal{P}_{ac}(\Delta^n), g_W)$ dotada de la estructura métrica W_2 . El funcional de energía libre \mathcal{F} define un campo vectorial gradiente $\nabla_{W_2} \mathcal{F}$. La evolución sigue el flujo de JKO (Jordan-Kinderlehrer-Otto), que es el límite cuando $\tau \rightarrow 0$ del esquema variacional discreto:

$$\rho_{k+1} \in \operatorname{argmin}_{\rho} \left\{ \frac{1}{2\tau} W_2^2(\rho, \rho_k) + \mathcal{F}(\rho) \right\}$$

Esto converge a la Ecuación de Fokker-Planck no lineal:

$$\partial_t \rho = \nabla \cdot (\rho \nabla (\delta \mathcal{F} / \delta \rho))$$

4.2 Grandes Desviaciones y Principio de Contracción

La tasa de convergencia de la medida empírica L_n hacia la medida invariante μ^* se rige por el funcional de acción $I(\nu)$ (Entropía relativa o Divergencia de Kullback-Leibler):

$$I(\nu) = \sup_{f \in C_b} \{ \langle f, \nu \rangle - \Lambda(f) \}$$

Para procesos dependientes ϕ -mixing, el Principio de Grandes Desviaciones (LDP) se mantiene con una función de tasa "rate function" convexa y semicontinua inferiormente (good rate function).

4.3 Acoplamiento Geométrico y Métrica Fisher-Rao

Para incorporar la sensibilidad a la estructura de la variedad estadística, generalizamos la métrica a una estructura de tipo Hellinger-Kantorovich o Fisher-Rao deformada por el tensor de curvatura inducido por el operador Ψ :

$$G(\rho) = e^{-\beta \|\nabla \Psi\|} G_{FR}(\rho)$$

donde G_{FR} es la métrica de información de Fisher. Esto define una geodésica adaptativa en el simplex de probabilidad.

4.4 Funcional de Lyapunov Global

La estabilidad asintótica del orquestador se demuestra mediante la función de Lyapunov basada en la entropía relativa:

$$V(w) = \sum_{i \in \text{opt}} w_i^* \log \left(\frac{w_i^*}{w_i(t)} \right), \quad \frac{dV}{dt} \leq 0$$

Capítulo 5

Fase 4: Convergencia y Estabilidad Global

5.1 Condiciones de Mezclado (Mixing)

Se asumen condiciones de β -mezclado (regularidad absoluta) con decaimiento exponencial:

$$\beta(\tau) = E \left[\sup_{B \in \mathcal{F}_{t+\tau}^\infty} |P(B|\mathcal{F}_{-\infty}^t) - P(B)| \right] \sim e^{-\lambda\tau}$$

5.2 Teorema de Sanov y Grandes Desviaciones

La probabilidad de que la medida empírica de error se desvíe del conjunto óptimo decae exponencialmente:

$$P(\hat{L}_t \in \Gamma) \leq C \exp \left(-n \inf_{\nu \in \Gamma} I(\nu) \right)$$

5.3 Estabilidad en Media L^p y Lyapunov

La estabilidad exponencial del flujo estocástico $\{\Xi_t\}$ se demuestra mediante el criterio de deriva de Foster-Lyapunov para generadores Markovianos debilmente continuos. Sea $V : \mathcal{H} \rightarrow \mathbb{R}_+$ una función Lyapunov compacta. Si existe un conjunto compacto ("petite set") $K \subset \mathcal{H}$ y constantes $\gamma > 0, b < \infty$ tal que:

$$\mathcal{L}V(x) \leq -\gamma V(x) + b\mathbb{1}_K(x)$$

entonces el proceso es geoméricamente ergódico y admite una medida invariante única π .

5.4 Complejidad Secuencial y Cotas de Generalización

Para acotar el exceso de riesgo en procesos no i.i.d., utilizamos la complejidad de Rademacher condicional $\mathcal{R}_n(\mathcal{F}|\mathbf{x})$.

$$\mathcal{R}_n(\mathcal{F}|\mathbf{x}) = E_\sigma \left[\sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n \sigma_i f(x_i) \right]$$

Para procesos β -mixing, se aplica la técnica de bloqueo de Bernstein para descomponer la dependencia temporal y aplicar desigualdades de concentración de McDiarmid.

5.5 Detección de Puntos de Cambio y Tiempos de Parada

Definimos el tiempo de parada τ como el primer instante de cruce de barrera del proceso CUSUM generalizado:

$$\tau = \inf\{t > 0 : \max_{0 \leq k \leq t} |S_t - S_k| \geq h(\Psi_t)\}$$

donde S_t es la suma parcial de los residuos de innovación estandarizados. Bajo la hipótesis nula de estacionariedad, S_t converge débilmente al Puente Browniano. En el instante τ , la medida de probabilidad se reinicia al prior uniforme sobre el simplex: $\rho_\tau = \text{Unif}(\Delta^n)$ (Entropía máxima).

5.5.1 Umbral Adaptativo para Colas Pesadas

En regímenes de alta volatilidad con distribuciones no Gaussianas (fat tails), el umbral basado únicamente en varianza puede generar falsos positivos. Formalizamos un umbral adaptativo que incorpora el cuarto momento estadístico.

Lema 5.1 (Umbral Adaptativo con Curtosis). *Sea $\{Z_t\}$ el proceso de residuos estandarizados con momento de cuarto orden finito. Definimos el umbral adaptativo:*

$$h_t = k \cdot \sigma_t \cdot \left(1 + \beta \cdot \frac{\kappa_t - 3}{\kappa_0}\right)$$

donde:

- σ_t es la desviación estándar móvil de los residuos
- $\kappa_t = \frac{E[(Z_t - \mu_t)^4]}{\sigma_t^4}$ es la curtosis (kurtosis excess respecto a la distribución normal)
- $k \in [3, 5]$ es el factor de sensibilidad base
- $\beta \in [0.1, 0.3]$ es el coeficiente de ajuste por colas pesadas
- $\kappa_0 = 3$ es la curtosis de referencia Gaussiana

Para distribuciones con colas pesadas ($\kappa_t > 3$), el umbral se incrementa proporcionalmente, reduciendo la tasa de falsas alarmas mientras se mantiene la potencia de detección en cambios estructurales genuinos.

Demostración. Bajo el marco de detección secuencial de Lorden, la probabilidad de falsa alarma $P(FA)$ está acotada por:

$$P(FA) \leq e^{-\theta h}$$

donde θ depende de la tasa de cambio de la media. Para distribuciones sub-Gaussianas (exponencialmente acotadas), esta desigualdad se mantiene con constantes universales.

Para distribuciones con colas más pesadas que la Gaussiana, la varianza ya no controla adecuadamente la masa en las colas. El cuarto momento (curtosis) captura la frecuencia de eventos extremos. Formalmente:

$$P(|Z_t - \mu_t| > c\sigma_t) \leq \frac{\kappa_t}{c^4} \quad (\text{Desigualdad de Markov de cuarto orden})$$

Al ajustar el umbral h_t por el factor $(1 + \beta(\kappa_t - 3)/\kappa_0)$, garantizamos que la probabilidad condicional de superar el umbral bajo la hipótesis nula permanece acotada uniformemente:

$$P\left(\max_{0 \leq k \leq t} |S_t - S_k| > h_t \mid H_0\right) \leq \alpha$$

donde α es el nivel de significancia deseado, independientemente del régimen de curtosis. \square

Nota 5.1. *Este ajuste es particularmente relevante para procesos financieros que exhiben curtosis empírica $\kappa \in [5, 20]$ (distribuciones leptocúrticas). Sin corrección, el detector CUSUM estándar genera un exceso de señales durante períodos de alta volatilidad estacional sin cambio estructural subyacente en la deriva.*

Corolario 5.1 (Consistencia Asintótica). *Para una secuencia de umbrales $\{h_t\}$ calibrados por curtosis, el tiempo de parada τ satisface:*

$$\lim_{n \rightarrow \infty} P(\tau > n \mid H_1) = 0$$

donde H_1 denota la hipótesis alternativa de cambio de régimen. Es decir, el detector mantiene potencia asintótica completa independientemente de la estructura de colas.

Capítulo 6

Ecuación Diferencial Operacional Unificada

6.1 Dinámica del Meta-Estado

El sistema completo se describe mediante la EDO estocástica no lineal en el espacio de Hilbert funcional $H = \mathcal{C} \times L^2(\Delta^n) \times \mathcal{L}(\mathcal{H}, \mathcal{H})$ que gobierna el meta-estado Ξ_t :

$$d\Xi_t = (\Xi_t, X_t)dt + (\Xi_t, X_t)dW_t$$

El drift encapsula:

1. La identificación topológica del operador SIA (Ψ).
2. El flujo de gradiente Wasserstein $\text{grad}_{W_2}\mathcal{F}$ proyectado en el espacio tangente de medidas.
3. La evolución de los predictores locales.

6.2 Teorema de Existencia y Unicidad Global

Teorema 6.1 (Existencia y Unicidad Débil). *Asumiendo que los coeficientes α y β son medibles y satisfacen condiciones locales de Lipschitz y crecimiento lineal (o que α es Hölder continuo y acotado, invocando el criterio de Yamada-Watanabe en dimensión finita), existe una única solución débil $(\Omega, \mathcal{F}, P, W, \Xi)$ para la ecuación diferencial estocástica operacional, tal que:*

$$E \left[\sup_{0 \leq s \leq T} \|\Xi_s\|^2 \right] < C(T, \|\Xi_0\|)$$

Apéndice A

Postulado de Robustez ante Singularidades

Postulado A.1. *Si el SIA detecta una dimensión de Hausdorff $D > 1$ o $\alpha(t) \rightarrow 0$, el sistema prioriza la Rama D (Signature) y activa la regularización de Huber. Esto garantiza la operatividad del predictor en regímenes de rugosidad extrema donde el cálculo diferencial estocástico estándar falla.*