

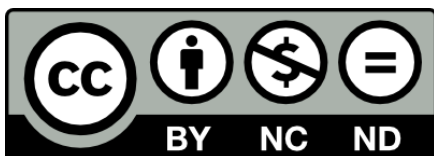
Stochastic Processes for Data Science

SAPIENZA
UNIVERSITÀ DI ROMA



Los Del DGIIM, losdeldgiim.github.io

Doble Grado en Ingeniería Informática y Matemáticas
Universidad de Granada



Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-SinDerivadas 4.0 Internacional (CC BY-NC-ND 4.0).

Eres libre de compartir y redistribuir el contenido de esta obra en cualquier medio o formato, siempre y cuando des el crédito adecuado a los autores originales y no persigas fines comerciales.

Stochastic Processes for Data Science

Los Del DGIIM, losdeldgiim.github.io

Joaquín Avilés de la Fuente

Granada, 2025-2026

Índice general

1. Introduction	7
1.1. Teorema de Recurrencia de Poincaré	7
1.2. Ergodicidad y el Lema de Kac	8
2. Stochastic Processes	9
2.1. Fundamentos y la Dualidad del Proceso Estocástico	9
2.2. Propiedades de las Distribuciones Finito-Dimensionales (FDD)	9
2.2.1. Desarrollo matemático de la Marginalización	10
2.3. Intercambiabilidad sin Independencia	10
2.4. Modelos de Asignación Combinatoria y F.O.F.	10
2.5. El Teorema de de Finetti (Versión Finita)	11
2.6. Modelos de urnas	11
2.6.1. Sin reemplazo (Hipergeométrica)	11
2.6.2. Con reforzamiento: Urna de Pólya	11
2.7. Teorema de Johnson y el Aprendizaje Inductivo	12
2.8. Resumen conceptual	12
3. Markov Chain	13
3.1. El Proceso de Simplificación Analítica	13
3.2. Time Homogeneous Markov Chains	14
3.3. Master Equation	15
3.3.1. Global Balance vs Detailed Balance	16
3.4. Ehrenfest Urn Model	17
3.4.1. Configuración y Espacio de Microestados (S_I)	17
3.4.2. Cadena de Macroestados o de Conteo (S_S)	17
3.4.3. Definición de Birth-Death Chains	17
3.4.4. La Solución: Versión Perezosa (α -retrasada)	17
3.5. Clasificación Topológica de Cadenas de Markov	18
3.6. Class Structures	18
3.6.1. Hitting Time & Absorption Probabilities	19
3.6.2. Stopping Times & Strong Markov Property	22
3.7. Estados Recurrentes y Transitorios	24
3.8. Positive recurrence vs Null recurrence	29
3.8.1. Convergencia a la Medida de Probabilidad Invariante	30
3.9. Reversible Markov Chain	37
3.9.1. Metropolis Algorithm	37
4. Procesos de Renovación y CTRW	39
4.1. Del paseo aleatorio discreto al proceso de renovación	39
4.2. El proceso CTRW: subordinando el espacio al tiempo	39
4.3. Markov, semi-Markov y el rol de la edad $\gamma(t)$	40
4.4. Distribución unipuntual y la ecuación de Montroll-Weiss	40
4.5. Caso particular: esperas exponenciales $\Rightarrow$ Proceso de Poisson	41
4.6. Límite difusivo: del CTRW a la ecuación del calor	41
4.7. Resumen conceptual	41

5. Procesos Gaussianos	43
5.1. Variable normal: lo básico	43
5.2. Vector Gaussiano: la idea importante	43
5.3. Proceso Gaussiano: lo que hay que saber	44
5.4. Movimiento Browniano: el resultado central del tema	44
5.5. Resumen en una frase por concepto	45
6. ¿Por qué funcionan los algoritmos MCMC?	47
6.1. Introducción	47
6.2. Motivación: el problema que resuelve MCMC	48
6.3. Construcción general: el Algoritmo de Metropolis-Hastings	48
6.4. Pilar 1: la construcción garantiza que π es estacionaria	50
6.5. Pilar 2: la cadena converge a π (sin importar el punto de partida)	50
6.6. Pilar 3: el Teorema Ergódico – por qué podemos <i>estimar</i> con MCMC	51
6.7. Resumen: los tres pilares	52

Índice general

1. Introduction

El tema aborda el problema de la predicción (forecasting) en sistemas complejos. Tradicionalmente, la física clásica asumía que conocer las variables y las ecuaciones (Escenario 1) garantizaba predecir el futuro de forma exacta. Sin embargo, el texto desmonta esta ilusión a través de tres barreras analíticas:

- **La Trampa del Caos:** En sistemas caóticos, los errores de medición iniciales (δ_0) crecen de forma exponencial en el tiempo. El horizonte de predicción (T) tiene un límite logarítmico: aunque gastes millones en mejorar la precisión de los sensores por un factor de 1000, el tiempo útil de predicción solo aumenta una minucia lineal. El determinismo no implica predictibilidad.
- **La Paradoja de la Alta Dimensión:** Si intentamos predecir usando el "Método de Análogos" (buscar en el pasado una situación idéntica a la actual para copiar su evolución), el espacio de fases se vuelve tan gigantesco debido al número de variables (N) que la probabilidad de que el sistema repita una configuración idéntica es prácticamente cero en tiempos humanos.
- **El Nacimiento del Paradigma Probabilístico:** Como es imposible realizar un seguimiento microscópico exacto, la Ciencia de Datos se ve obligada a abandonar las trayectorias rígidas y tratar el caos y la falta de información mediante Procesos Estocásticos.

1.1. Teorema de Recurrencia de Poincaré

Teorema 1.1 (Teorema de Recurrencia de Poincaré). *Sea un sistema dinámico determinista que preserva el volumen (como los sistemas hamiltonianos donde la energía total $T + V = E$ es constante) y cuya evolución ocurre en un dominio acotado o conjunto compacto. Para cualquier estado inicial $\underline{x}_0$ y cualquier entorno abierto σ alrededor de él (definido por una distancia máxima de tolerancia $\varepsilon > 0$), existen infinitos instantes de tiempo discretos k tales que el estado del sistema regresa al entorno original:*

$$\exists k \geq 1 : d(\underline{x}_k, \underline{x}_0) < \varepsilon \implies \underline{x}_k \in \sigma \quad (1.1)$$

¿Qué nos dice el resultado? Establece que si tienes un sistema dinámico determinista, que se mueve en un espacio acotado (cerrado y limitado) y que conserva su volumen (como la energía en un sistema físico), el sistema siempre terminará volviendo a pasar muy cerca de donde empezó. Analíticamente, si partes de un estado inicial $\underline{x}_0$ con un margen de error tolerado ε , existirán infinitos momentos en el futuro (k) donde el estado de la trayectoria regresará a estar a una distancia menor que ε de su origen:

$$\exists k \geq 1 : d(\underline{x}_k, \underline{x}_0) < \varepsilon$$

Imagínate una caja cerrada llena de gas donde todas las partículas se mueven de forma caótica. Si sacas una foto inicial, el teorema de Poincaré te garantiza que, si esperas lo suficiente, las partículas se moverán de tal manera que eventualmente volverán a colocarse casi exactamente en la misma posición de la foto inicial, infinitas veces.

1.2. Ergodicidad y el Lema de Kac

Para calcular cuánto tarda el sistema en regresar a ese estado inicial (el Tiempo de Retorno τ_σ), necesitamos introducir el concepto de Ergodicidad. Ergodicidad (Definición intuitiva): Un sistema es ergódico si seguir a un único individuo a lo largo de un tiempo infinito (Promedio Temporal) da exactamente el mismo resultado estadístico que congelar el tiempo y promediar a todos los individuos del sistema a la vez en ese instante (Promedio de Conjunto). El Lema de Kac Si asumimos que el sistema cumple la hipótesis de ergodicidad, el Lema de Kac formaliza matemáticamente la normalización del tiempo de retorno respecto a la medida del espacio. Nos da el siguiente resultado analítico:

$$\int_{\underline{x} \in \sigma} \tau_\sigma(x) d\mu(x) = 1$$

Al sustituir este lema en la fórmula clásica de la esperanza matemática, obtenemos una conclusión brillantemente simple:

$$\mathbb{E}[\tau_\sigma(x)] = \frac{1}{\mu(\sigma)}$$

El resultado fundamental: El tiempo promedio que tarda un sistema en regresar a una región de su espacio de estados es exactamente el inverso del volumen (medida μ) de esa región.

La gran conclusión (Unión de Poincaré, Kac y Ciencia de Datos) Aquí es donde los profesores conectan todo. Si un sistema tiene N variables independientes (grados de libertad), el volumen de la región objetivo se encoge exponencialmente: $\mu(\sigma) \propto \left(\frac{\varepsilon}{L}\right)^N$. Al aplicar el Lema de Kac, el tiempo esperado de retorno se dispara hacia arriba de forma exponencial:

$$\mathbb{E}[\tau_\sigma(x)] \propto \left(\frac{L}{\varepsilon}\right)^N$$

En Ciencia de Datos, cuando manejas problemas con cientos o miles de variables ($N \gg 1$), el tiempo promedio que debes esperar para que una trayectoria regrese a un estado pasado (paradoja de los sistemas de alta dimensión) es mayor que la edad del universo. ¿Por qué fracasa el método de análogos? Porque el tiempo de recurrencia de Kac nos demuestra que nunca encontrarás en tu base de datos histórica un día que sea .exactamente igual. al de hoy para poder predecir el de mañana. El pasado exacto no se repite en sistemas de alta dimensión.

2. Stochastic Processes

Antes de empezar a trabajar con procesos estocásticos vamos a recordar los **axiomas de Kolmogorov**, que son las tres condiciones fundamentales que definen matemáticamente una medida de probabilidad. Dado un espacio muestral Ω y una colección de sucesos $\mathcal{F}$, tenemos

- **Axioma de No Negatividad:** La probabilidad de cualquier suceso A es un número real mayor o igual a cero

$$P(A) \geq 0 \quad \forall A \in \mathcal{F}$$

- **Axioma de Certidumbre:** La probabilidad de que ocurra el espacio muestral completo Ω (el suceso seguro) es exactamente igual a 1

$$P(\Omega) = 1$$

- **Axioma de Aditividad Sigma (σ -aditividad):** Si tienes una secuencia infinita de sucesos mutuamente excluyentes o disjuntos entre sí ($A_i \cap A_j = \emptyset$ para todo $i \neq j$), la probabilidad de su unión es la suma de sus probabilidades individuales

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

2.1. Fundamentos y la Dualidad del Proceso Estocástico

Un **proceso estocástico** $X(\omega, t)$ es una colección de variables aleatorias indexadas en el tiempo, definidas sobre un mismo espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$. Su comprensión profunda radica en su **naturaleza dual**:

- **Realización o Trayectoria (ω fijo, t variable):** Al fijar el azar, el proceso se convierte en una función totalmente determinista del tiempo ($X(\bar{\omega}, t)$). Representa la evolución histórica concreta que ha seguido el fenómeno (por ejemplo, el gráfico de una acción de la bolsa ya cerrada).
- **Variable Aleatoria (t fijo, ω variable):** Al congelar el reloj en un instante preciso $\bar{t}$, el objeto colapsa en una variable aleatoria tradicional. Para poder asignarle probabilidades válidas en el mundo real, debe cumplir la condición analítica de **medibilidad** respecto a la σ -álgebra $\mathcal{F}$:

$$X^{-1}(B, \bar{t}) = \{\omega \in \Omega \mid X(\omega, \bar{t}) \in B\} \in \mathcal{F} \quad (2.1)$$

2.2. Propiedades de las Distribuciones Finito-Dimensionales (FDD)

Para caracterizar estadísticamente una sucesión ordenada de variables discretas $X_1, X_2, \dots, X_m$, recurrimos a las **Distribuciones Finito-Dimensionales**, las cuales representan la probabilidad de la intersección lógica de que cada variable tome un valor simultáneamente.

Para que el proceso sea consistente (garantizado por el *Teorema de Consistencia de Kolmogorov*), las FDD deben cumplir obligatoriamente dos propiedades esenciales:

1. **Simetría ante Permutaciones:** La distribución conjunta es invariante respecto al orden en el que se listen las componentes:

$$\mathbb{P}(\{X_1 = x_1\} \cap \cdots \cap \{X_m = x_m\}) = \mathbb{P}(\{X_{\sigma_1} = x_{\sigma_1}\} \cap \cdots \cap \{X_{\sigma_m} = x_{\sigma_m}\})$$

2. **Consistencia de Marginalización:** Permite colapsar la dimensión del problema eliminando las variables no deseadas mediante el Teorema de la Probabilidad Total.

2.2.1. Desarrollo matemático de la Marginalización

Para demostrar que una estructura conjunta es consistente, colapsamos un sistema tridimensional eliminando el tercer lanzamiento X_3 sobre su soporte elemental $x_3 \in \{0, 1\}$ en una distribución intercambiable de Bernoulli ($q = 1 - p$):

$$\begin{aligned} P(X_1 = 0, X_2 = 0) &= \sum_{x_3 \in \{0,1\}} P(X_1 = 0, X_2 = 0, X_3 = x_3) \\ &= P(0, 0, 1) + P(0, 0, 0) \\ &= (1-p)^2 p + (1-p)^3 \\ &= (1-p)^2 [p + (1-p)] \\ &= (1-p)^2 \cdot [1] = (1-p)^2 \end{aligned}$$

El resultado recupera con total fidelidad la probabilidad marginal del sistema de menor dimensión ($P(X_1 = 0) \cdot P(X_2 = 0)$), probando su consistencia.

2.3. Intercambiabilidad sin Independencia

Una **distribución intercambiable** es aquella en la que cualquier permutación de una secuencia de variables aleatorias comparte exactamente la misma probabilidad. Las secuencias que contienen *el mismo número de éxitos y fracasos* tienen idéntica probabilidad, sin importar el orden temporal.

- **Muestreo sin Reemplazo (Distribución Hipergeométrica):** La probabilidad de la secuencia (W, W, B) es $\frac{W}{N} \cdot \frac{W-1}{N-1} \cdot \frac{B}{N-2}$, mientras que la de (W, B, W) es $\frac{W}{N} \cdot \frac{B}{N-1} \cdot \frac{W-1}{N-2}$. Los valores numéricos son idénticos porque los numeradores y denominadores contienen los mismos factores en diferente orden.
- **Muestreo con Reforzamiento (Urna de Pólya):** Cada vez que se extrae una bola, se devuelve y se añade otra de su mismo color. Utilizando el **factorial ascendente** ($x^{(y)} = x(x+1) \cdots (x+y-1)$), la probabilidad de una trayectoria ordenada se reduce a:

$$P(\text{Trayectoria}) = \frac{W^{(w)} B^{(b)}}{N^{(n)}} \quad (2.2)$$

Como el orden de los factores en el numerador no altera el producto, el proceso es estrictamente intercambiable pero claramente dependiente.

2.4. Modelos de Asignación Combinatoria y F.O.F.

Para modelar la asignación de m objetos en g categorías de mayor a menor detalle:

1. **Descripción Individual (Microscópica):** Registra qué objeto cae en qué caja. Total: g^m .
2. **Descripción Estadística (Macroscópica):** Registra únicamente el conteo de ocupación de cada caja ($Y_i = m_i$). Total mediante combinaciones con repetición: $\binom{m+g-1}{g-1}$.
3. **Frecuencias de Frecuencias / F.O.F. (Máxima Compresión):** Variable $Z_i = j$, que significa *“existen exactamente j categorías que contienen i objetos”*.

Restricciones Algebraicas de las f.o.f.

Cualquier vector válido f.o.f. $\underline{Z}$ debe satisfacer obligatoriamente dos leyes de conservación:

- **Conservación de Categorías:** $\sum_{i=0}^m Z_i = g$
- **Conservación de Objetos:** $\sum_{i=0}^m i \cdot Z_i = m$

Para calcular cuántas descripciones estadísticas $\underline{Y}$ corresponden a un vector f.o.f. dado $\underline{z}$:

$$\text{Número de descripciones estadísticas} = \frac{g!}{Z_0! \cdot Z_1! \cdot Z_2! \cdot \dots \cdot Z_m!} \quad (2.3)$$

2.5. El Teorema de de Finetti (Versión Finita)

Teorema 2.1 (De Finetti - Versión Finita). *Si una distribución de probabilidad es estrictamente intercambiable sobre una secuencia finita de longitud n , existe una única distribución de probabilidad discreta a priori $\mu(h)$ tal que la probabilidad de cualquier trayectoria se puede expresar como una mezcla ponderada de densidades hipergeométricas uniformes:*

$$P(X_1 = x_1, \dots, X_n = x_n) = \sum_{h=0}^n \frac{1}{\binom{n}{h}} \mu(h) \quad (2.4)$$

Demostración. Definiendo la trayectoria microscópica como el suceso $A = \bigcap_{i=1}^n \{X_i = x_i\}$, aplicamos el Teorema de la Probabilidad Total condicionándolo sobre el número total de éxitos reales h :

$$P(A) = \sum_{h=0}^n P\left(A \mid \sum_{i=1}^n X_i = h\right) P\left(\sum_{i=1}^n X_i = h\right)$$

Por la hipótesis fundamental de **intercambiabilidad**, la distribución asigna un peso idéntico a cada una de las $\binom{n}{h}$ secuencias que comparten el mismo número de éxitos. Por tanto, la probabilidad condicional de observar la trayectoria exacta A colapsa a:

$$P\left(A \mid \sum_{i=1}^n X_i = h\right) = \frac{1}{\binom{n}{h}}$$

Sustituyendo esta equivalencia combinatoria, y definiendo la medida de ponderación discreta como la probabilidad marginal del conteo de éxitos ($\mu(h) = P(\sum X_i = h)$), obtenemos el resultado del teorema:

$$P(A) = \sum_{h=0}^n \frac{1}{\binom{n}{h}} \mu(h)$$

□

2.6. Modelos de urnas

2.6.1. Sin reemplazo (Hipergeométrica)

Al extraer sin reposición, las probabilidades dependen del historial (no hay independencia), pero el proceso sigue siendo intercambiable: el orden no altera la probabilidad de una trayectoria con w blancas y b negras. Esto da lugar a la **distribución hipergeométrica**.

2.6.2. Con reforzamiento: Urna de Pólya

Cada bola extraída se devuelve junto con una bola extra del mismo color (refuerzo positivo, genera correlación). Para g categorías con parámetros iniciales $\alpha_1, \dots, \alpha_g$ ($\alpha = \sum_i \alpha_i$), la **probabilidad predictiva** de observar la categoría 1 tras n extracciones es

$$P(X_{n+1} = 1 \mid X_1 = x_1, \dots, X_n = x_n) = \frac{\alpha_1 + n_1}{\alpha + n}.$$

Esta es la idea más importante del modelo: la predicción combina la "masa inicial" (creencia previa) con el conteo observado. La distribución conjunta resultante (agrupando categorías) es la **Distribución de Pólya Multivariante**, y en el límite $\alpha \rightarrow \infty$ (urna infinita) el efecto de refuerzo se diluye y el proceso converge a la **Multinomial** de ensayos independientes.

2.7. Teorema de Johnson y el Aprendizaje Inductivo

El teorema de W.E. Johnson demuestra que la probabilidad predictiva de observar una categoría en el futuro se reduce a una **combinación convexa** entre la información empírica real y las creencias teóricas iniciales:

$$P(X_{m+1} = j \mid m_j, m) = \left(\frac{m}{m + \alpha} \right) \left(\frac{m_j}{m} \right) + \left(\frac{\alpha}{m + \alpha} \right) \left(\frac{\alpha_j}{\alpha} \right) \quad (2.5)$$

Análisis matemático de los dos Regímenes Extremos

- **Régimen de Datos Escasos** ($m \rightarrow 0$): El peso del primer término se desvanece ($\frac{m}{m+\alpha} \rightarrow 0$), provocando que la predicción dependa exclusivamente de la estructura a priori:

$$\lim_{m \rightarrow 0} P(X_{m+1} = j \mid m_j, m) = \frac{\alpha_j}{\alpha}$$

- **Régimen de Datos Masivos** ($m \rightarrow \infty$): El factor de ponderación empírica tiende a la unidad ($\frac{m}{m+\alpha} \rightarrow 1$), mientras que el peso de la opinión a priori se diluye hacia cero. El sistema olvida su sesgo inicial y converge a la proporción empírica observada:

$$\lim_{m \rightarrow \infty} P(X_{m+1} = j \mid m_j, m) = \frac{m_j}{m}$$

2.8. Resumen conceptual

- Un proceso estocástico tiene una doble lectura: trayectoria (determinista) vs. variable aleatoria (en un instante).
- La consistencia de Kolmogorov (simetría + marginalización) garantiza que una familia de FDD define un proceso bien definido.
- La **intercambiabilidad** es más débil que la independencia, pero suficiente para que aparezcan regularidades fuertes (como en la urna de Pólya).
- La urna de Pólya y el Teorema de Johnson muestran cómo la predicción óptima es siempre una mezcla convexa entre prior y evidencia empírica.
- El **Teorema de de Finetti** es la conclusión final: toda secuencia infinita intercambiable es, necesariamente, una mixtura de procesos i.i.d. Bernoulli(θ) sobre un prior $Q(\theta)$ — el fundamento teórico del enfoque bayesiano.

3. Markov Chain

Definición 3.1. La **propiedad de markov local** establece que para cada nodo X_i de un DAG, la variable aleatoria X_i es condicionalmente independiente de sus antepasados remotos dado el conjunto de sus padres directos, es decir, se obtiene toda su información relevante para predecir X_i a partir de sus padres directos.

Definición 3.2 (Cadena de Márkov). Un proceso estocástico en tiempo discreto $\{X_m\}_{m \geq 0}$ definido sobre un **espacio de estados** S discreto (finito o infinito numerable, tal que $S = \{1, \dots, g\}$, $S = \mathbb{N}$ o $S = \mathbb{Z}$) se denomina **cadena de márkov** si cumple con la **Propiedad de Márkov**:

$$P(X_{m+1} = x_{m+1} \mid X_0 = x_0, X_1 = x_1, \dots, X_m = x_m) = P(X_{m+1} = x_{m+1} \mid X_m = x_m)$$

para todo $m \geq 0$ y para cualquier secuencia de estados $\{x_0, x_1, \dots, x_{m+1}\} \in S$ tales que $P(X_0 = x_0, \dots, X_m = x_m) > 0$.

3.1. El Proceso de Simplificación Analítica

Descomposición de la Distribución Conjunta

Por la regla de la cadena de la probabilidad elemental, la distribución finito-dimensional conjunta de una trayectoria de longitud m requiere almacenar toda la memoria histórica acumulada:

$$\begin{aligned} P(X_1 = x_1, \dots, X_m = x_m) &= P(X_m = x_m \mid X_1 = x_1, \dots, X_{m-1} = x_{m-1}) \\ &\quad \cdot P(X_{m-1} = x_{m-1} \mid X_1 = x_1, \dots, X_{m-2} = x_{m-2}) \\ &\quad \vdots \\ &\quad \cdot P(X_2 = x_2 \mid X_1 = x_1) \cdot P(X_1 = x_1) \end{aligned}$$

Al invocar formalmente la Propiedad de Márkov, el condicionamiento masivo colapsa hacia el eslabón cronológico inmediato anterior, reduciendo la probabilidad conjunta al producto de transiciones locales:

$$P(X_1 = x_1, \dots, X_m = x_m) = P(X_1 = x_1) \prod_{i=2}^m P(X_i = x_i \mid X_{i-1} = x_{i-1})$$

Homogeneidad Temporal

En una cadena de Márkov general, la regla de transición entre dos estados puede mutar según la edad del proceso, dependiendo intrínsecamente del paso del tiempo $m - 1$:

$$P(X_m = y \mid X_{m-1} = x) = P_{m-1}(x, y)$$

Para estabilizar dinámicamente el modelo, se introduce el supuesto de **Homogeneidad en el Tiempo**. Se asume que las leyes que rigen los saltos son estacionarias e invariantes respecto al eje temporal:

$$P(X_m = y \mid X_{m-1} = x) = P(x, y) \quad \forall m \geq 1$$

Dadas estas dos condiciones, la convergencia de la propiedad de Márkov y la homogeneidad temporal, una Cadena de Márkov queda unívocamente determinada y gobernada por la interacción de solo dos elementos matemáticos:

1. **Distribución Inicial** ($p(x)$): un vector de probabilidad que dicta el estado de arranque del sistema en el instante inicial:

$$P(X_1 = x) = p(x) \quad \forall x \in S$$

2. **Probabilidad de Transición** ($P(x, y)$): la probabilidad estática de migrar desde un estado origen x hacia un estado destino y en un solo paso temporal discreto:

$$P(X_{m+1} = y \mid X_m = x) = P(x, y) \quad \forall x, y \in S$$

3.2. Time Homogeneous Markov Chains

Definición 3.3. Un proceso estocástico en tiempo discreto $\{X_n\}_{n \geq 0}$ con espacio de estados finito o ajustable S es una **cadena de markov homogénea en el tiempo** si sus probabilidades de transición condicionales satisfacen que, para todo paso temporal m , la probabilidad no depende del índice global sino únicamente de la distancia entre los estados:

$$P(x, y) = P(X_{m+1} = y \mid X_m = x) = P(X_1 = y \mid X_0 = x) \quad \forall x, y \in S$$

Las probabilidades de transición deben preservar la medida de probabilidad unitaria en su espacio mapeado:

$$\sum_{y \in S} P(x, y) = 1 \quad \forall x \in S$$

Definición 3.4. Una **matriz estocástica (o matriz de Markov)** es una matriz cuadrada cuyas entradas son no negativas y cumplen con la condición de que la suma de los elementos de cada una de sus filas es estrictamente igual a la unidad, es decir,

$$\sum_{y \in S} P(x, y) = 1, \forall x \in S$$

Distribución Marginal y Transición en n Pasos

En esta sección, cambiaremos un poco la mentalidad, esta vez diremos “No me importa el camino intermedio; solo quiero saber la probabilidad de estar en el estado x_n en el paso n , habiendo empezado en algún lado”, lo que significa que vamos a marginalizar sobre todos los estados intermedios posibles $x_0, x_1, \dots, x_{n-1} \in S$.

Para recuperar la distribución marginal en cualquier etapa futura arbitraria n , denotada como $p_n(x_n) = P(X_n = x_n)$, aplicamos el Teorema de Probabilidad Total sumando sobre todas las historias o trayectorias internas posibles de los estados previos $x_0, x_1, \dots, x_{n-1} \in S$:

$$p_n(x_n) = \sum_{x_0 \in S} \sum_{x_1 \in S} \cdots \sum_{x_{n-1} \in S} P(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n)$$

Sustituyendo la ecuación estructural de la trayectoria conjunta de la sección anterior, el cálculo se reduce algebraicamente a:

$$\begin{aligned} p_n(x_n) &= \sum_{x_0 \in S} \left[\sum_{x_1, \dots, x_{n-1} \in S^{n-1}} P(x_{n-1}, x_n) \cdot P(x_{n-2}, x_{n-1}) \cdots P(x_0, x_1) \right] p_0(x_0) \\ &= \sum_{x_0 \in S} P^{(n)}(x_0, x_n) p_0(x_0) \end{aligned}$$

donde $P^{(n)}(x_0, x_n)$ representa la probabilidad de transición en n pasos desde el estado inicial x_0 hasta el estado final x_n , obtenida al sumar sobre todos los caminos intermedios posibles.

La sumatoria interna anidada representa algebraicamente la multiplicación iterada fila por columna de la matriz de transiciones P , dando lugar al elemento indexado de la matriz potencia P^n . Definimos matemáticamente dicho comportamiento como:

$$P^{(n)}(x_0, x_n) = P^n(x_0, x_n)$$

donde lo que se está haciendo es multiplicar la matriz de transición P consigo misma n veces, y luego tomar el elemento correspondiente a la fila x_0 y la columna x_n . Esto nos permite calcular la probabilidad de transición en n pasos de manera eficiente utilizando álgebra lineal.

Observación. Para ver la igualdad aplicada anteriormente haremos este desarrollo paso a paso. Existe una equivalencia exacta entre el **Teorema de la Probabilidad Total** (sumar las rutas intermedias de una cadena) y la **multiplicación de matrices** (producto escalar de fila por columna).

Sean dos matrices A y B , su multiplicación para una celda específica se define mediante un índice compartido k :

$$(A \times B)_{i,j} = \sum_k A(i, k) \cdot B(k, j)$$

Al calcular la probabilidad de transicionar de un estado inicial x_0 a un estado final x_2 en exactamente **2 pasos**, la teoría probabilística nos obliga a marginalizar sobre todos los estados intermedios posibles $x_1 \in S$:

$$P^{(2)}(x_0, x_2) = \sum_{x_1 \in S} \underbrace{P(x_0, x_1)}_{\text{Fila } x_0 \text{ de } P} \cdot \underbrace{P(x_1, x_2)}_{\text{Columna } x_2 \text{ de } P}$$

Como el índice de la etapa intermedia x_1 ocupa la posición de columna en la primera matriz y de fila en la segunda, la estructura matemática es idéntica. Por lo tanto:

- Un sumatorio indexado ($\sum_{x_1}$) equivale a operar **un producto matricial** ($P \times P$).
- Una sumatoria anidada de $n - 1$ variables temporales ($\sum_{x_1} \cdots \sum_{x_{n-1}}$) equivale a la **multiplicación iterada** de la matriz por sí misma, resultando directamente en la matriz potencia P^n .

3.3. Master Equation

Definición 3.5. Para una cadena de Markov homogénea en el tiempo, la probabilidad de transicionar desde un estado x a un estado y en exactamente dos pasos se calcula sumando sobre todos los estados intermedios posibles $z \in S$:

$$P^{(2)}(x, y) = \sum_{z \in S} P(x, z) \cdot P(z, y)$$

lo que se llama **Chapman-Kolmogorov equation** para $n = 2$. Para el caso general n , tenemos

$$P^{(n)}(x, y) = \sum_{i_1, \dots, i_{n-1} \in S} P(x, i_1) P(i_1, i_2) \dots P(i_{n-1}, y)$$

Teorema 3.1 (Chapman-Kolmogorov Equation). Sea $\{X_n\}_{n \geq 0}$ una cadena de Markov homogénea en el tiempo. Para cualesquiera pasos temporales $n, m \geq 1$ y estados $x, y \in S$, la probabilidad de transicionar de x a y en $n + m$ pasos satisface

$$P^{(n+m)}(x, y) = \sum_{z \in S} P^{(n)}(x, z) \cdot P^{(m)}(z, y)$$

Demostración. Para demostrarlo, debes apoyarnos firmemente en el Teorema de la Probabilidad Total condicionando en un “tiempo intermedio” exacto (el paso n).

Paso 1. Plantear el objetivo analítico. Queremos calcular $P(X_{n+m} = y \mid X_0 = x)$.

Paso 2. Introducir el estado intermedio. Para llegar a y en el paso $n + m$, el sistema obligatoriamente tuvo que pasar por algún estado z en el paso intermedio n . Aplicas el Teorema de la Probabilidad Total sumando sobre todo el espacio de estados S

$$P(X_{n+m} = y \mid X_0 = x) = \sum_{z \in S} P(X_{n+m} = y, X_n = z \mid X_0 = x)$$

Paso 3. Separar por la Regla de la Cadena, es decir, usando

$$P(A \cap B) = P(A \mid B) \cdot P(B)$$

entonces abrimos la probabilidad conjunta del sumatorio

$$\sum_{z \in S} P(X_{n+m} = y, X_n = z \mid X_0 = x) = \sum_{z \in S} P(X_{n+m} = y \mid X_n = z, X_0 = x) \cdot P(X_n = z \mid X_0 = x)$$

Paso 4. Por la Propiedad de Markov, el futuro ($n + m$) solo depende del presente (n), por lo que el pasado ($X_0 = x$) se olvida en el primer factor

$$P(X_{n+m} = y \mid X_n = z, X_0 = x) = P(X_{n+m} = y \mid X_n = z)$$

Por Homogeneidad temporal, avanzar m pasos desde el instante n al $n + m$ es exactamente lo mismo que avanzar m pasos desde el cero

$$P(X_{n+m} = y \mid X_n = z) = P^{(m)}(z, y)$$

Paso 5. Sustituyendo los términos simplificados, llegas directamente a la igualdad del teorema

$$P^{(n+m)}(x, y) = \sum_{z \in S} P^{(m)}(z, y) \cdot P^{(n)}(x, z)$$

□

Definición 3.6. Decimos que un vector de probabilidad $\pi = (\pi(x))_{x \in S}$ constituye una **distribución invariante** (o estacionaria) de la cadena si actúa como un punto fijo bajo el operador de transición, lo que significa que no sufre variaciones con el transcurso del tiempo:

$$\pi P = \pi \quad \equiv \quad \pi(y) = \sum_{x \in S} \pi(x) P(x, y) \quad \forall y \in S$$

donde P es la matriz de transición de la cadena de Markov en el caso discreto, y en el caso continuo sería Q .

3.3.1. Global Balance vs Detailed Balance

Definición 3.7. Para cualquier cadena de Markov, si existe una distribución invariante π , esta satisface de forma general la ecuación de **global balance**, lo que significa que el cambio neto de probabilidad en el equilibrio es nulo

$$0 = \sum_{x \in S} [P(x, j)\pi(x) - P(j, x)\pi(j)] \quad \forall j \in S$$

Una condición mucho más restrictiva (y sumamente útil en algoritmos de Data Science como MCMC) es el principio de **detailed balance**. Si existe un vector de probabilidad π tal que el flujo de probabilidad entre cada par de estados se equilibra de forma aislada y simétrica:

$$P(x, j)\pi(x) = P(j, x)\pi(j) \quad \forall x, j \in S$$

Definición 3.8. Una cadena de Markov se dice **reversible** si existe una distribución invariante π que satisface la condición de detailed balance, dada por la ecuación:

$$P(x, j)\pi(x) = P(j, x)\pi(j) \quad \forall x, j \in S$$

donde para cada par de estados $x, j \in S$, el flujo de probabilidad desde x hacia j es exactamente igual al flujo de probabilidad desde j hacia x cuando el sistema está en equilibrio.

3.4. Ehrenfest Urn Model

El Modelo de Ehrenfest es el ejemplo académico perfecto para evaluar si entiendes la diferencia entre **microestados** y **macroestados**, y cómo la **periodicidad** puede arruinar la convergencia de una cadena.

3.4.1. Configuración y Espacio de Microestados (S_I)

- **Concepto:** Describe la posición exacta e individual de cada una de las n partículas mediante un vector binario (e.g., 101 significa que las bolas 1 y 3 están a la izquierda y la 2 a la derecha).
- **Tamaño del espacio:** Tiene exactamente 2^n estados posibles (para $n = 3$, son 8 estados).
- **Distribución invariante π_I :** Es estrictamente **uniforme** debido a la simetría del problema:

$$\pi_I = \left(\frac{1}{2^n}, \dots, \frac{1}{2^n} \right)$$

3.4.2. Cadena de Macroestados o de Conteo (S_S)

- **Concepto:** Simplifica el sistema fijándose únicamente en una variable agregada: $Y_1 = \sum X_i$ (el número total de bolas en el compartimento izquierdo, L).
- **Tamaño del espacio:** Se reduce drásticamente a solo $n + 1$ estados (de 0 a n).
- **Leyes de Transición (¡Aprender de memoria!):** Estando en el estado k :
 - Probabilidad de Muerte ($k \rightarrow k - 1$): $\frac{k}{n}$ (*eliges una bola de la izquierda para moverla a la derecha*).
 - Probabilidad de Nacimiento ($k \rightarrow k + 1$): $\frac{n-k}{n}$ (*eliges una bola de la derecha para moverla a la izquierda*).
 - Probabilidad de Permanencia ($k \rightarrow k$): 0 (*en el modelo estándar obligatorio cambia una bola*).
- **Distribución estacionaria π_S :** Sigue una forma **Binomial**. Para $n = 3$ es $\pi_S = (\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8})$.
- **El Gran Problema Analítico (Pregunta de examen):** Esta cadena de conteo es irreducible pero **PERIÓDICA** (Periodo $d = 2$). Al no poder quedarse quieta en un paso ($P(x, x) = 0$), la cadena oscila eternamente entre estados pares e impares. Por lo tanto, **NO converge asintóticamente** (falla la hipótesis de aperiodicidad del Teorema de Convergencia Estacionaria).

3.4.3. Definición de Birth-Death Chains

- Es una cadena de Markov en un retículo entero donde las únicas transiciones permitidas con probabilidad mayor que cero en un paso son hacia sus vecinos inmediatos ($k - 1$ o $k + 1$) o quedarse en el mismo estado (k). Los saltos de largo alcance están prohibidos. El modelo de Ehrenfest pertenece a esta familia.

3.4.4. La Solución: Versión Perezosa (α -retrasada)

- **Objetivo fundamental:** Romper la periodicidad para forzar la convergencia al equilibrio.
- **Mecanismo:** Se introduce un parámetro de pereza $\alpha \in (0, 1)$. En cada paso, con probabilidad α la cadena se congela ($k \rightarrow k$), y con probabilidad $1 - \alpha$ se mueve una partícula al azar usando las leyes normales.
- **Consecuencia analítica (El argumento ganador):** Al añadir α en la diagonal principal, consigues que $P(x, x) = \alpha > 0$, lo que destruye los ciclos rígidos de oscilación par/impar y vuelve a la cadena **APERIÓDICA**.

- **Conclusión asintótica:** Al ser ahora simultáneamente **irreducible y aperiódica**, el Teorema Ergódico/Estacionario se cumple con rigor. Al elevar la matriz estocástica perezosa a una potencia $n \rightarrow \infty$, el sistema olvida por completo el estado inicial y todas sus filas convergen exactamente al vector de equilibrio binomial π_S .

3.5. Clasificación Topológica de Cadenas de Markov

La estructura analítica de un espacio de estados S se clasifica según la probabilidad de retorno a largo plazo:

- **Estado Transitorio:** existe una probabilidad estrictamente positiva de que, una vez abandonado el estado, el proceso jamás regrese a él.
- **Estado Recurrente:** el proceso regresa al estado con probabilidad igual a la unidad de forma casi segura.
- **Conjunto Absorbente / Cerrado:** un subconjunto de estados del cual la cadena es incapaz de escapar una vez que ha ingresado en él.

Definición 3.9. El **fenómeno de metaestabilidad** se refiere a la situación en la que un sistema estocástico parece estar en un estado estable durante un período prolongado, pero eventualmente transiciona a otro estado estable debido a perturbaciones pequeñas o aleatorias. Viene dado por una cadena con dos trampas absorbentes aisladas cuya matriz es la Identidad (I), donde introducimos una perturbación infinitesimal $\varepsilon > 0$ (donde $\varepsilon \ll 1$, por ejemplo $\varepsilon = 10^{-20}$), los muros puros se rompen:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \xrightarrow{\text{Perturbación}} \begin{pmatrix} 1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon \end{pmatrix}$$

Observación. Este fenómeno recibe el nombre de metaestabilidad, pues el sistema parece estable y congelado localmente, pero macroscópicamente evoluciona de manera lenta hacia una distribución uniforme.

El comportamiento en este fenómeno se ve de dos formas

- **Comportamiento a corto plazo:** Si el proceso inicia en A , se mantendrá atrapado en A durante un tiempo extremadamente largo ($1 - \varepsilon \approx 1$). Localmente parece una cadena absorbente estática.
- **Comportamiento asintótico (Long-run):** Al ser una matriz simétrica perturbada, los estados se comunican y la cadena se vuelve **irreducible**. A largo plazo ($n \rightarrow \infty$), el sistema alcanzará un equilibrio perfecto y balanceado:

$$\pi_A = \pi_B = \frac{1}{2}$$

3.6. Class Structures

Definición 3.10 (Accesibilidad o Conducción). Se dice que el estado i **conduce al** estado j (notado como $i \rightarrow j$) si existe una probabilidad estrictamente positiva de visitar el estado j en algún momento, dado que el sistema comenzó en el estado i :

$$P(\exists n \geq 0 : X_n = j \mid X_0 = i) > 0$$

Definición 3.11 (Comunicación). Se dice que el estado i **se comunica con** el estado j (notado como $i \longleftrightarrow j$) si se cumplen simultáneamente que $i \rightarrow j$ y $j \rightarrow i$.

Teorema 3.2. Para dos estados distintos $i \neq j$, las siguientes afirmaciones son equivalentes:

1. $i \rightarrow j$.
2. Existen estados intermedios $i_1, i_2, \dots, i_{n-1}$ tales que $P_{ii_1} P_{i_1 i_2} \dots P_{i_{n-1} j} > 0$.

3. $P_{ij}^{(n)} > 0$ para algún $n \geq 1$.

Proposición 3.3. La relación de comunicación " $\longleftrightarrow$ " es una **relación de equivalencia** sobre el espacio de estados S , de cadenas de Márkov.

Demostración. Debemos verificar tres propiedades fundamentales:

1. **Reflexiva** ($i \longleftrightarrow i$): Por definición, en el paso $n = 0$,

$$P_{ii}^{(0)} = P(X_0 = i \mid X_0 = i) = 1 > 0$$

por lo que $i \rightarrow i$.

2. **Simétrica** ($i \longleftrightarrow j \implies j \longleftrightarrow i$): Se deriva directamente de la conmutatividad de la conjunción lógica en la definición del operador ($\longleftrightarrow$), veámoslo

$$i \longleftrightarrow j \iff (i \rightarrow j) \wedge (j \rightarrow i) \iff (j \rightarrow i) \wedge (i \rightarrow j) \iff j \longleftrightarrow i$$

3. **Transitiva** ($i \longleftrightarrow j \wedge j \longleftrightarrow k \implies i \longleftrightarrow k$): Si $i \rightarrow j$ y $j \rightarrow k$, existen $n, m \geq 0$ tales que $P_{ij}^{(n)} > 0$ y $P_{jk}^{(m)} > 0$. Por Chapman-Kolmogorov:

$$P_{ik}^{(n+m)} = \sum_{l \in S} P_{il}^{(n)} P_{lk}^{(m)} \geq P_{ij}^{(n)} P_{jk}^{(m)} > 0 \implies i \rightarrow k$$

Siguiendo el mismo argumento de forma inversa para $k \rightarrow j \rightarrow i$, se obtiene $k \rightarrow i$. Por lo tanto, $i \longleftrightarrow k$.

□

Observación. Como toda relación de equivalencia, la comunicación divide al conjunto S en clases de equivalencia mutuamente disjuntas llamadas **clases de comunicación**. Si todos los estados pertenecen a una única clase, la cadena se denomina **irreducible**.

Definición 3.12 (Clase Cerrada). Una clase de comunicación C se dice que es **cerrada** si para todo estado $i \in C$, la condición $i \rightarrow j$ implica obligatoriamente que $j \in C$. Es decir, una vez que el proceso entra en una clase cerrada, no puede salir de ella.

Definición 3.13 (Estado Absorbente). Un estado $i \in S$ es **absorbente** si el subconjunto unipersonal $\{i\}$ forma por sí mismo una clase cerrada. Esto equivale a decir que $P_{ii} = 1$.

Definición 3.14 (Irreducibilidad). Una cadena de Markov se denomina **irreducible** si el espacio de estados total S está constituido por una única clase de comunicación común. En otras palabras, todos los estados de la cadena se comunican mutuamente entre sí.

3.6.1. Hitting Time & Absorption Probabilities

Consideremos una cadena de Markov en tiempo discreto $(X_n)_{n \geq 0}$ con matriz de probabilidad de transición P que opera sobre el espacio de estados S .

Definición 3.15 (Tiempo de Llegada / Hitting Time). El **tiempo de primera llegada** a un subconjunto de estados $A \subseteq S$ es una variable aleatoria $H^A : \Omega \rightarrow \{0, 1, 2, \dots\} \cup \{\infty\}$ que cuantifica el número mínimo de pasos necesarios para intersectar la región A :

$$H^A(\omega) = \inf\{n \geq 0 : X_n(\omega) \in A\}$$

Bajo el convenio matemático estándar, si el conjunto es vacío, adoptamos que $\inf \emptyset = \infty$.

Definición 3.16 (Probabilidad de Llegada). La **probabilidad de llegada** o **hitting probability** es la probabilidad de alcanzar el conjunto de estados A en tiempo finito partiendo inicialmente desde un estado específico i se define como:

$$h_i^A = P_i(H^A < \infty) = P(H^A < \infty \mid X_0 = i)$$

Observación. Si la región de destino A se corresponde con una clase cerrada de la cadena de Markov, la métrica h_i^A adopta el nombre específico de **probabilidad de absorción**.

Definición 3.17 (Distribución Impropia). La variable aleatoria H^A puede presentar una **distribución de probabilidad de carácter impropio** si existe una probabilidad residual no nula de no alcanzar nunca el conjunto objetivo A . En tal escenario, la suma sobre la trayectoria finita cumple:

$$P_i(H^A < \infty) = \sum_{n=0}^{\infty} P_i(H^A = n) = \eta \leq 1$$

De lo cual se deduce de forma inmediata la **probabilidad complementaria** para el horizonte infinito:

$$P_i(H^A = \infty) = 1 - \eta$$

Definición 3.18 (Tiempo Medio de Llegada). El **tiempo esperado de llegada** al subconjunto A partiendo desde el estado inicial i viene dado por el valor esperado estadístico $E_i[H^A]$:

$$k_i^A = E_i[H^A] = \sum_{n=0}^{\infty} n P_i(H^A = n) + \infty \cdot P_i(H^A = \infty)$$

donde si la probabilidad de no alcanzar el conjunto objetivo es nula ($P_i(H^A = \infty) = 0$), el término asintótico indeterminado desaparece de la ecuación de forma directa.

Interpretación Intuitiva de las Variables

- h_i^A : Cuantifica la probabilidad pura de lograr alcanzar el conjunto A en algún punto de la historia futura.
- k_i^A : Modela el número esperado de pasos temporales que el sistema consumirá antes de tocar por primera vez la región A .

Teorema 3.4. El vector de probabilidades de primera llegada $h^A = (h_i^A)_{i \in S}$ es la solución no negativa mínima para el sistema de ecuaciones lineales:

$$\begin{cases} h_i^A = 1 & \text{para } i \in A \\ h_i^A = \sum_{j \in S} P_{ij} h_j^A & \text{para } i \notin A \end{cases} \quad (1)$$

Minimalidad significa que si $x = (x_i)_{i \in S}$ es otra solución con $x_i \geq 0 \forall i$, entonces $x_i \geq h_i^A \forall i$.

Demostración. Analizaremos el comportamiento dinámico del sistema dividiéndolo según la naturaleza del estado inicial respecto al conjunto objetivo A :

- Si $X_0 = i$ con $i \in A$, entonces por definición el tiempo de llegada es inmediato, $H^A = 0$, por lo que

$$h_i^A = P_i(H^A < \infty) = 1$$

- Si $X_0 = i$ con $i \notin A$, entonces necesariamente el tiempo de primera llegada cumple que $H^A \geq 1$. Utilizando la propiedad de Markov espacial (pérdida de memoria) y la homogeneidad en el tiempo (la probabilidad pasar de i a j es la misma independientemente del instante n), evaluamos el condicionamiento al primer paso de la trayectoria:

$$P(H^A < \infty \mid X_0 = i, X_1 = j) = P(H^A < \infty \mid X_1 = j) = h_j^A$$

Aplicando el teorema de la probabilidad total mediante el condicionamiento exhaustivo sobre el primer estado visitado X_1 , obtenemos:

$$\begin{aligned} h_i^A &= P_i(H^A < \infty) = \sum_{j \in S} P_i(H^A < \infty, X_1 = j) \\ &\stackrel{(*)}{=} \sum_{j \in S} P(H^A < \infty \mid X_0 = i, X_1 = j) P(X_1 = j \mid X_0 = i) \\ &= \sum_{j \in S} (P_{ij} h_j^A) = \sum_{j \in S} P_{ij} h_j^A \end{aligned}$$

donde en $(*)$ se ha aplicado $P(A \cap B) = P(A \mid B)P(B)$.

Esto demuestra que el vector real de probabilidades de llegada h^A satisface formalmente el sistema lineal (1).

Demostraremos ahora la minimalidad. Sea $x = (x_i)_{i \in S}$ un vector que representa cualquier solución no negativa arbitraria del sistema (1). Evaluamos de igual forma por casos:

- Si $i \in A$, por diseño del sistema lineal se cumple de forma directa que $x_i = 1 = h_i^A$.
- Si $i \notin A$, la ecuación correspondiente nos permite descomponer el sumatorio sobre el espacio de estados separando la región objetivo A de su complemento:

$$x_i = \sum_{j \in S} P_{ij} x_j = \sum_{j \in A} P_{ij} x_j + \sum_{j \notin A} P_{ij} x_j$$

Sustituyendo el valor fijo $x_j = 1$ para $j \in A$, y reconociendo que $\sum_{j \in A} P_{ij} = P_i(X_1 \in A)$, el término se reescribe como:

$$x_i = P_i(X_1 \in A) + \sum_{j \notin A} P_{ij} x_j$$

Procedemos a expandir la ecuación de forma iterativa sustituyendo la estructura analógica para el término no acotado x_j dentro del sumatorio complementario:

$$x_i = P_i(X_1 \in A) + \sum_{j \notin A} P_{ij} \left(P_j(X_1 \in A) + \sum_{k \notin A} P_{jk} x_k \right)$$

Expandiendo los productos e identificando las probabilidades de transición conjuntas mediante la propiedad de Markov, obtenemos:

$$x_i = P_i(X_1 \in A) + P_i(X_1 \notin A, X_2 \in A) + \sum_{j \notin A} \sum_{k \notin A} P_{ij} P_{jk} x_k$$

donde para el término sumatorio del medio, hemos usado que

$$P_j(X_1 \in A) = P(X_2 \in A \mid X_1 = j)$$

gracias a aplicar la homogeneidad del tiempo, es decir, es lo mismo partir de j (instante 0) y tener en el instante 1 un estado en A , a estar en el instante 1 con estado j y que el instante siguiente 2 caiga en un estado de A .

Al reiterar este procedimiento inductivo un número finito m de pasos, la expresión adopta la siguiente forma generalizada:

$$\begin{aligned} x_i = & P_i(X_1 \in A) + P_i(X_1 \notin A, X_2 \in A) + \dots \\ & + P_i(X_1 \notin A, \dots, X_{m-1} \notin A, X_m \in A) + \sum_{j_1 \notin A} \dots \sum_{j_m \notin A} P_{ij_1} \dots P_{j_{m-1}j_m} x_{j_m} \end{aligned}$$

Reconociendo que la suma finita de probabilidades de trayectorias truncadas modela exactamente la probabilidad acumulada de alcanzar el conjunto A en el paso m o antes, simplificamos la expresión como:

$$x_i = P_i(H^A \leq m) + \sum_{j_1 \notin A} \dots \sum_{j_m \notin A} P_{ij_1} P_{j_1 j_2} \dots P_{j_{m-1} j_m} x_{j_m}$$

Dado que el vector de soluciones alternativo es estrictamente no negativo por hipótesis ($x \geq 0$) y las probabilidades de transición P_{kl} son no negativas, el término del sumatorio múltiple residual es mayor o igual a cero. Por lo tanto, podemos establecer la siguiente desigualdad válida para cualquier entero m :

$$x_i \geq P_i(H^A \leq m) \quad \forall m \geq 1$$

Tomando el límite matemático cuando el horizonte de pasos tiende a infinito en ambos lados de la desigualdad, y aplicando la propiedad de continuidad de la medida de probabilidad para eventos monótonos crecientes, concluimos formalmente que:

$$x_i \geq \lim_{m \rightarrow \infty} P_i(H^A \leq m) = P_i(H^A < \infty) = h_i^A$$

Por lo tanto, queda demostrado que $x_i \geq h_i^A$ para todo estado $i \in S$, lo que confirma que el vector real h^A constituye la solución mínima no negativa del sistema. $\square$

Teorema 3.5. *El vector de tiempos esperados de primera llegada $k^A = (k_i^A)_{i \in S}$ es la solución no negativa mínima para el sistema de ecuaciones lineales:*

$$\begin{cases} k_i^A = 0 & \text{para } i \in A \\ k_i^A = 1 + \sum_{j \notin A} P_{ij} k_j^A & \text{para } i \notin A \end{cases} \quad (2)$$

Minimalidad significa que si $x = (x_i)_{i \in S}$ es otra solución con $x_i \geq 0 \forall i$, then $x_i \geq k_i^A \forall i$.

3.6.2. Stopping Times & Strong Markov Property

Definición 3.19 (Tiempo de Parada / Stopping Time). Una variable aleatoria

$$T : \Omega \rightarrow \{0, 1, 2, \dots\} \cup \{\infty\}$$

se denomina **tiempo de parada** si para cada instante finito $n \in \{0, 1, 2, \dots\}$, la ocurrencia o no del suceso $\{T = n\}$ está determinada única y exclusivamente por la trayectoria histórica acumulada del proceso hasta ese momento, es decir, depende únicamente de los valores de las variables aleatorias $X_0, X_1, \dots, X_n$.

Teorema 3.6 (Strong Markov Property). *Sea $(X_n)_{n \geq 0}$ una cadena de Markov homogénea con vector de probabilidad inicial λ y matriz de probabilidad de transición P . Sea T un tiempo de parada respecto a la filtración natural del proceso $(X_n)_{n \geq 0}$.*

Entonces, condicionado al suceso $\{T < \infty\}$ y $\{X_T = i\}$, el proceso reiniciado $(X_{T+n})_{n \geq 0}$ es una cadena de Markov autónoma con vector de probabilidad inicial $\delta_i = (0, \dots, 0, 1, 0, \dots, 0)$ y matriz de probabilidad de transición P . Además, el proceso futuro $(X_{T+n})_{n \geq 0}$ es estadísticamente independiente de la trayectoria histórica pasada $X_0, X_1, \dots, X_T$.

Observación. Si nos fijamos, lo que nos dice el teorema es que si el proceso alcanza un estado i en un tiempo de parada T , entonces el proceso reiniciado a partir de ese instante T se comporta exactamente como si hubiéramos comenzado la cadena de Markov desde cero con estado inicial i . Es decir, el proceso “olvida” su historia pasada y se reinicia con las mismas probabilidades de transición que al inicio, lo que refleja la esencia de la propiedad de Markov fuerte.

Esto realmente parece muy parecido a la propiedad de Markov normal, que simplemente indica un instante de tiempo n y dice que el futuro no está determinado por su pasado, pero la diferencia clave con la propiedad de Markov fuerte es que entra en juego cuando el instante T es una variable aleatoria, es decir, un instante que tú no sabes qué número será hasta que ves evolucionar la cadena, porque depende del propio comportamiento del proceso.

En resumen, la propiedad normal solo sabe congelar el tiempo en números fijos ($n = 5$). No sabe lidiar con un tiempo que “flitrea” y cambia de valor según la suerte de la cadena. El Teorema Fuerte de Markov es el que viene al rescate para demostrar matemáticamente que, incluso si el instante T es una sorpresa aleatoria, la cadena también se reseteará en ese momento.

Definición 3.20 (Estados Recurrentes y Transitorios). Sea $(X_n)_{n \geq 0}$ una cadena de Markov que opera con vector de probabilidad inicial λ y matriz de probabilidad de transición P . Se define la clasificación fundamental de un estado $i \in S$ según las siguientes condiciones asintóticas:

- Se dice que el estado i es **recurrente** si la probabilidad de regresar a él una cantidad infinita de veces es exactamente igual a la certeza:

$$P_i(X_n = i \text{ infinitas veces}) = 1$$

- Se dice que el estado i es **transitorio** si la probabilidad de regresar a él una cantidad infinita de veces es nula:

$$P_i(X_n = i \text{ infinitas veces}) = 0$$

Recordemos que el **tiempo de primer paso** o llegada a un estado específico i viene gobernado por la variable aleatoria:

$$T_i(\omega) = \inf\{n \geq 1 : X_n(\omega) = i\}, \quad \text{bajo el convenio } \inf \emptyset = \infty$$

Definición 3.21 (n -ésimo Tiempo de Paso e Intervalo de Recurrencia). Para caracterizar las visitas sucesivas de la cadena, definimos el r -ésimo **tiempo de paso a un estado** i , denotado por $T_i^{(r)}$, de forma recursiva mediante las siguientes identidades:

$$\begin{aligned} T_i^{(0)}(\omega) &= 0 \\ T_i^{(1)}(\omega) &= T_i(\omega) \\ \text{Para todo } r \geq 1 : \quad T_i^{(r+1)}(\omega) &= \inf\{n > T_i^{(r)} : X_n(\omega) = i\} \end{aligned}$$

De manera complementaria, el tiempo transcurrido entre dos visitas consecutivas al estado i se define como el n -ésimo **intervalo de recurrencia** $S_i^{(n)}$, el cual adopta la estructura:

$$S_i^{(n)} = \begin{cases} T_i^{(n)} - T_i^{(n-1)} & \text{si } T_i^{(n-1)} < \infty \\ 0 & \text{en otro caso} \end{cases}$$

Definición 3.22 (Número de Visitas a un Estado). Sea $(X_n)_{n \geq 0}$ una cadena de Markov. Se define la variable aleatoria discreta V_i , que cuantifica el **número total de visitas al estado** i , mediante la suma de funciones indicadoras a lo largo de toda la trayectoria infinita del proceso:

$$V_i = \sum_{n=0}^{\infty} I_{\{X_n=i\}}$$

Observación. Haciendo uso de la linealidad del operador del valor esperado, el número esperado de visitas que realiza la cadena al estado i equivale a la suma de las probabilidades marginales de ocupación:

$$E(V_i) = E\left(\sum_{n=0}^{\infty} I_{\{X_n=i\}}\right) = \sum_{n=0}^{\infty} E[I_{\{X_n=i\}}] = \sum_{n=0}^{\infty} P(X_n = i)$$

Cuando asumimos que el proceso inicia con total certeza en dicho estado de referencia ($X_0 = i$), esta identidad se reduce de forma directa a la suma del término diagonal de las matrices de transición en n pasos:

$$E_i(V_i) = \sum_{n=0}^{\infty} P_i(X_n = i) = \sum_{n=0}^{\infty} P_{ii}^{(n)}$$

Definición 3.23 (Probabilidad de Retorno). La **probabilidad de retorno** h_i se define como la medida de probabilidad de que la cadena regrese al estado de origen i en algún instante de tiempo estrictamente positivo y finito:

$$h_i = P_i(T_i < \infty)$$

donde $T_i = \inf\{n \geq 1 : X_n = i\}$ representa el tiempo de primer paso al estado i .

3.7. Estados Recurrentes y Transitorios

Lema 3.7. Para todo índice entero de visitas $r = 0, 1, 2, \dots$, la probabilidad de que la cadena de Markov visite el estado i estrictamente más de r veces, dado que inició en i , sigue una distribución geométrica de supervivencia gobernada por la potencia de la probabilidad de retorno:

$$P_i(V_i > r) = h_i^r$$

Nota aclaratoria estructural: Recuerda que el $(r+1)$ -ésimo tiempo absoluto de paso se construye de forma aditiva mediante $T_i^{(r+1)} = T_i^{(r)} + S_i^{(r+1)}$, donde $S_i^{(r+1)}$ representa el correspondiente intervalo de recurrencia relativo.

Teorema 3.8 (Dicotomía Fundamental de los Estados). Para todo estado $i \in S$ de una cadena de Markov, se satisface de forma estricta y excluyente una de las siguientes dos condiciones analíticas asintóticas:

1. Si $P_i(T_i < \infty) = 1$, entonces el estado i es **recurrente** y la serie infinita de sus probabilidades de transición diagonal **DIVERGE**:

$$\sum_{m=0}^{\infty} P_{ii}^{(m)} = \infty$$

2. Si $P_i(T_i < \infty) < 1$, entonces el estado i es **transitorio** y la serie infinita de sus probabilidades de transición diagonal **CONVERGE**:

$$\sum_{m=0}^{\infty} P_{ii}^{(m)} < \infty$$

Demostración. Evaluamos de forma separada los dos comportamientos mutuamente excluyentes de la probabilidad de retorno $h_i = P_i(T_i < \infty)$:

Caso i) Supongamos que la probabilidad de primer retorno es unitaria: $P_i(T_i < \infty) = 1$, lo que implica que $h_i = 1$. Evaluando el límite de la probabilidad acumulada de colas para el número de visitas obtenemos:

$$P_i(V_i = \infty) = \lim_{r \rightarrow \infty} P_i(V_i > r) = \lim_{r \rightarrow \infty} h_i^r = \lim_{r \rightarrow \infty} 1^r = 1$$

Esto confirma que el estado i es recurrente. Adicionalmente, aplicando la expansión lineal para el valor esperado de la variable, la serie de transiciones diagonales diverge:

$$\sum_{m=0}^{\infty} P_{ii}^{(m)} = E_i(V_i) \stackrel{(*)}{=} \sum_{r=0}^{\infty} P_i(V_i > r) = \sum_{r=0}^{\infty} 1^r = \infty$$

donde en $(*)$ hemos usado lo obtenido a continuación

$$\begin{aligned} E(X) &= \sum_{k=1}^{\infty} k \cdot P(X = k) \implies E(X) \stackrel{(**)}{=} \sum_{k=1}^{\infty} \left(\sum_{r=0}^{k-1} 1 \right) P(X = k) = \sum_{k=1}^{\infty} \sum_{r=0}^{k-1} P(X = k) \implies \\ &\implies E(X) = \sum_{r=0}^{\infty} \sum_{k=r+1}^{\infty} P(X = k) = \sum_{r=0}^{\infty} P(X > r) \end{aligned}$$

donde en $(**)$ hemos utilizado que $k = \sum_{r=0}^{k-1} 1$.

Caso ii) Supongamos ahora que la probabilidad de primer retorno es estrictamente fraccionaria: $P_i(T_i < \infty) < 1$, de manera que $h_i < 1$. Al calcular de igual forma el comportamiento límite de la sucesión geométrica se deduce que:

$$P_i(V_i = \infty) = \lim_{r \rightarrow \infty} P_i(V_i > r) = \lim_{r \rightarrow \infty} h_i^r = 0$$

Lo que define al estado i como transitorio. Evaluando la esperanza matemática del número de visitas a través del sumatorio de su función de supervivencia, recuperamos de forma exacta la suma de una serie geométrica convergente de razón h_i :

$$\sum_{m=0}^{\infty} P_{ii}^{(m)} = E_i(V_i) = \sum_{r=0}^{\infty} P_i(V_i > r) = \sum_{r=0}^{\infty} h_i^r = \frac{1}{1 - h_i} < \infty$$

Al ser $h_i < 1$, el denominador es estrictamente positivo, asegurando la convergencia finita de la serie de probabilidades, completando formalmente la demostración del teorema.

□

Teorema 3.9. Toda clase de comunicación recurrente es matemáticamente cerrada.

Teorema 3.10. Toda clase de comunicación que sea cerrada y finita es recurrente.

Definición 3.24 (Variables de Ocupación antes del Retorno). Sea $(X_n)_{n \geq 0}$ una cadena de Markov que evoluciona sobre un espacio de estados arbitrario E . Fijado un estado de referencia base $k \in E$, definimos el tiempo de primera llegada como la variable aleatoria:

$$T_k = \inf\{n \geq 1 : X_n = k\} \in \mathbb{N} \cup \{\infty\}$$

A partir de este hito temporal, se define la variable estadística $\gamma_i^{(k)}$ como el **número esperado de visitas realizadas al estado i antes de que la cadena regrese por primera vez al estado original k** :

$$\gamma_i = \gamma_i^{(k)} = E_k \left(\sum_{n=0}^{T_k-1} I_{\{X_n=i\}} \right) \in [0, +\infty]$$

donde hemos usado que $E_k(\cdot) \equiv E(\cdot \mid X_0 = k)$.

Teorema 3.11 (Teorema de existencia general de medida invariante). *Para una cadena de Markov irreducible y recurrente que comienza su trayectoria en el estado de referencia $k \in E$, la medida del número esperado de visitas intermedias $\gamma = (\gamma_i)_{i \in E}$ cumple rigurosamente las siguientes tres propiedades estructurales:*

1. $\gamma_k = \gamma_k^{(k)} = 1$
2. Para todo estado $i \in E$, el valor está estrictamente acotado: $0 < \gamma_i < \infty$.
3. El vector $\gamma = (\gamma_i)_{i \in E}$ constituye una **medida invariante** del sistema, es decir

$$\gamma P = \gamma$$

Demostración. Desglosamos la verificación analítica para cada uno de los tres apartados del teorema:

- (1) Por definición constructiva de la variable aleatoria de ocupación, evaluamos la esperanza de la suma indicadora partiendo del estado k :

$$\gamma_k = E_k \left(\sum_{m=0}^{T_k-1} I_{\{X_m=k\}} \right)$$

Dado que el proceso arranca con certeza en k ($X_0 = k$), la primera indicadora toma el valor de 1. Para cualquier instante de tiempo posterior $1 \leq m \leq T_k - 1$, por la propia definición matemática del tiempo de primer retorno ($T_k = \inf\{n \geq 1 : X_n = k\}$), la cadena tiene estrictamente prohibido regresar a k . Por lo tanto, el sumatorio se reduce únicamente a su término inicial:

$$\gamma_k = E_k(I_{\{X_0=k\}}) = E_k(1) = 1$$

- (3) Para demostrar analíticamente que γ es una medida invariante bajo la acción de la matriz de transición, debemos verificar el cumplimiento exacto de la ecuación algebraico-lineal de balance para cualquier estado de destino $j \in E$:

$$\gamma_j = \sum_{i \in E} \gamma_i P_{ij}$$

Dado que la hipótesis del teorema nos asegura que la cadena es recurrente, el tiempo de primer retorno ocurre de forma finita casi con total certeza ($T_k < \infty$ c.s.), lo que implica la identidad de fronteras $X_0 = k = X_{T_k}$. Mediante un desplazamiento cíclico del índice temporal, alteramos los extremos del sumatorio interior:

$$\gamma_j = E_k \left(\sum_{m=0}^{T_k-1} I_{\{X_m=j\}} \right) = E_k \left(\sum_{m=1}^{T_k} I_{\{X_m=j\}} \right)$$

Extendiendo el sumatorio formalmente hacia el horizonte infinito mediante la introducción de la condición indicadora $\{m \leq T_k\}$ y explotando la linealidad del operador valor esperado, desarrollamos:

$$= E_k \left(\sum_{m=1}^{\infty} I_{\{X_m=j, m \leq T_k\}} \right) = \sum_{m=1}^{\infty} E_k (I_{\{X_m=j, m \leq T_k\}}) = \sum_{m=1}^{\infty} P_k(X_m = j, m \leq T_k)$$

Aplicamos el Teorema de la Probabilidad Total introduciendo una partición exhaustiva sobre el espacio de estados intermedios visitados en el paso temporal inmediatamente anterior $X_{m-1} = i$, pues si sumamos la probabilidad de transición desde todos los estados posibles $i \in E$, obtenemos la probabilidad total de alcanzar el estado j en el paso m

$$= \sum_{i \in E} \sum_{m=1}^{\infty} P_k(X_m = j, X_{m-1} = i, T_k \geq m)$$

Invocando la **Propiedad de Markov** ordinaria, el comportamiento dinámico del paso m se desvincula de las restricciones históricas del pasado acumuladas en el evento $\{T_k \geq m\}$, es decir, aplicando esto en (**) y la probabilidad condicionada en (*) tenemos

$$\begin{aligned} P_k(X_m = j, X_{m-1} = i, T_k \geq m) &\stackrel{(*)}{=} P_k(X_m = j \mid X_{m-1} = i, T_k \geq m) P_k(X_{m-1} = i, T_k \geq m) = \\ &\stackrel{(**)}{=} P_k(X_m = j \mid X_{m-1} = i) P_k(X_{m-1} = i, T_k \geq m) \end{aligned}$$

y aplicando la homogeneidad temporal tenemos

$$\gamma_j = \sum_{i \in E} \sum_{m=1}^{\infty} P_{ij} P_k(X_{m-1} = i, T_k \geq m)$$

Efectuamos un cambio de variable elemental sobre el índice temporal de la serie interior fijando $m' = m - 1$:

$$= \sum_{i \in E} \sum_{m'=0}^{\infty} P_{ij} P_k(X_{m'} = i, T_k \geq m' + 1)$$

Restaurando la variable indicadora dentro de la esperanza matemática y traduciendo la condición de cota superior $\{T_k \geq m' + 1\}$ como $\{T_k - 1 \geq m'\}$, la igualdad de balance se consolida al reagrupar los términos:

$$= \sum_{i \in E} \sum_{m=0}^{\infty} P_{ij} E_k (I_{\{X_m=i, T_k \geq m+1\}}) = \sum_{i \in E} P_{ij} \underbrace{E_k \left(\sum_{m=0}^{T_k-1} I_{\{X_m=i\}} \right)}_{\gamma_i} = \sum_{i \in E} \gamma_i P_{ij}$$

Por lo tanto, queda formalmente demostrado que el vector γ constituye una medida invariante de la cadena.

- (2) Debido a la hipótesis de irreducibilidad del sistema, sabemos que todos los estados se comunican libremente entre sí. Por consiguiente, fijado un estado arbitrario $i \in E$, se garantiza de forma deductiva la existencia de dos horizontes de pasos finitos $m, n \geq 0$ tales que las probabilidades de transición en múltiples pasos satisfacen que $P_{ki}^{(m)} > 0$ y $P_{ik}^{(n)} > 0$.

Haciendo uso de la propiedad de invarianza de la medida demostrada en el apartado anterior ($\gamma = \gamma P^m$), desglosamos el sumatorio matricial para aislar la contribución del estado de referencia k :

$$\gamma_i = \sum_{j \in E} \gamma_j P_{ji}^{(m)} \geq \gamma_k P_{ki}^{(m)}$$

Sustituyendo el valor unitario de la frontera obtenido en el primer apartado ($\gamma_k = 1$), establecemos la acotación inferior estricta:

$$\gamma_i \geq (1) \cdot P_{ki}^{(m)} = P_{ki}^{(m)} > 0$$

Para verificar la cota superior, evaluamos de igual forma la invarianza de la medida sobre el estado k tras un horizonte de n pasos ($\gamma = \gamma P^n$), aislando en esta ocasión el aporte del estado intermedio i :

$$1 = \gamma_k = \sum_{j \in E} \gamma_j P_{jk}^{(n)} \geq \gamma_i P_{ik}^{(n)}$$

Despejando el término de interés, dado que la intensidad de transición es estrictamente positiva ($P_{ik}^{(n)} > 0$), la componente queda confinada superiormente por un valor real real finito:

$$\gamma_i \leq \frac{1}{P_{ik}^{(n)}} < \infty$$

Al haber demostrado que la componente es estrictamente positiva y finita para cualquier estado del espacio, se completa la demostración de todas las tesis del teorema.

□

Teorema 3.12 (Teorema del núcleo fundamental de medida invariante). *Sea $(X_n)_{n \geq 0}$ una cadena de Markov irreducible y recurrente que opera sobre un espacio de estados discreto E . Fijemos un estado arbitrario de referencia $k \in E$. Entonces, cualquier medida invariante $\lambda = (\lambda_j)_{j \in E}$ asociada al sistema es única salvo por un factor de proporcionalidad constante. Es decir, se puede escribir de forma unívoca como:*

$$\lambda_j = c \gamma_j^{(k)} \quad \forall j \in E$$

donde la constante multiplicativa viene dada por la masa del estado de referencia, $c = \lambda_k$, la cual es estrictamente independiente del estado de destino j .

Observación. Como consecuencia directa de este teorema, se establece que una cadena de Markov que es simultáneamente irreducible y recurrente posee una **única medida invariante** salvo por un factor de escala constante multiplicativo. Dependiendo de las propiedades de sumabilidad de la trayectoria, esta medida invariante única puede adoptar una masa total finita ($\sum \lambda_i < \infty$, caso recurrente positivo) o bien una masa total infinita ($\sum \lambda_i = \infty$, caso recurrente nulo).

Corolario 3.12.1. *Toda cadena de Markov que sea irreducible y opere sobre un espacio de estados finito posee una **única medida de probabilidad invariante**.*

Demostración. Desglosamos la verificación analítica en dos fases fundamentales: la construcción de una medida normalizada (de probabilidad) y la demostración formal de su unicidad.

Fase 1. Por resultados previos del temario, sabemos que toda cadena de Markov que es simultáneamente cerrada (por ser irreducible) y finita cumple de forma obligatoria con la condición de ser **recurrente**.

Al haber demostrado la recurrencia, podemos invocar el teorema de existencia general, el cual nos garantiza que existe una medida invariante $\lambda = (\lambda_i)_{i \in E}$ que es única salvo por una constante multiplicativa escalar. Además, sabemos que todos sus componentes son estrictamente positivos y reales ($\lambda_i > 0$).

1. **Finitud de la masa total:** Dado que el espacio de estados E es estrictamente **finito**, podemos asegurar con total rigor matemático que la suma de todos los componentes de la medida λ no diverge. Definimos esta masa total como la constante M :

$$M = \sum_{i \in E} \lambda_i < \infty$$

Si el espacio E fuese infinito, esta suma podría divergir a infinito, impidiendo el proceso de normalización.

2. **Construcción de la medida de probabilidad (λ'):** Para transformar nuestra medida general λ en una medida de probabilidad (cuya suma total sea exactamente 1), procedemos a dividir cada componente individual por la masa total M . Definimos algebraicamente este nuevo vector como λ' ("lambda prima"):

$$\lambda'_i = \frac{\lambda_i}{M} \quad \forall i \in E$$

3. **Verificación del axioma de probabilidad:** Haciendo uso de la linealidad del operador sumatorio, verificamos analíticamente que λ' distribuye una masa de probabilidad unitaria sobre todo el espacio:

$$\sum_{i \in E} \lambda'_i = \sum_{i \in E} \frac{\lambda_i}{M} = \frac{1}{M} \sum_{i \in E} \lambda_i = \frac{1}{M} \cdot M = 1$$

Dado que $\lambda P = \lambda$, por linealidad se cumple inmediatamente que $\lambda' P = \lambda'$, consolidando a λ' como una medida de probabilidad invariante legítima.

Fase 2. Para demostrar que no puede existir otra medida de probabilidad invariante distinta, utilizaremos el método directo de reducción analítica por igualación de constantes.

Supongamos hipotéticamente que existen dos medidas de probabilidad invariantes del sistema, a las que denominaremos ν' y ν'' . Fijemos un estado arbitrario de referencia $k \in E$.

1. **Relación con el vector de visitas esperadas ($\gamma^{(k)}$):** Por el teorema del núcleo fundamental, cualquier medida invariante del sistema debe ser un múltiplo escalar del vector fundamental de visitas intermedias esperado $\gamma^{(k)} = (\gamma_i^{(k)})_{i \in E}$. Consiguientemente, deben existir dos constantes reales estrictamente positivas c' y c'' tales que:

$$\nu'_i = c' \gamma_i^{(k)} \quad \text{y} \quad \nu''_i = c'' \gamma_i^{(k)} \quad \forall i \in E$$

2. **Aplicación de la condición de normalización:** Como por hipótesis tanto ν' como ν'' son medidas de probabilidad, la suma de sus componentes sobre el espacio exhaustivo E debe ser idéntica a la unidad. Expandiendo ambas series mediante sustitución, obtenemos:

$$\begin{aligned} 1 &= \sum_{i \in E} \nu'_i = \sum_{i \in E} c' \gamma_i^{(k)} = c' \sum_{i \in E} \gamma_i^{(k)} \\ 1 &= \sum_{i \in E} \nu''_i = \sum_{i \in E} c'' \gamma_i^{(k)} = c'' \sum_{i \in E} \gamma_i^{(k)} \end{aligned}$$

3. **Colapso e igualdad de constantes:** Despejando algebraicamente de forma directa los escalares c' y c'' de las ecuaciones de balance anteriores, observamos que quedan unívocamente determinados por el mismo valor numérico real:

$$c' = \frac{1}{\sum_{i \in E} \gamma_i^{(k)}} \quad \text{y} \quad c'' = \frac{1}{\sum_{i \in E} \gamma_i^{(k)}}$$

Por transitividad algebraica, se deduce de forma inmediata que:

$$c' = c''$$

4. **Conclusión:** Al ser los escalares idénticos ($c' = c''$), los vectores constituyentes se igualan componente a componente para todo el espacio:

$$\nu'_i = c' \gamma_i^{(k)} = c'' \gamma_i^{(k)} = \nu''_i \implies \nu' = \nu''$$

Queda formalmente demostrado que cualquier medida de probabilidad invariante alternativa colapsa necesariamente en la misma estructura vector-espacio, confirmando la estricta unicidad de la solución.

□

Observación. La discrepancia fundamental entre **medida invariante** y **medida invariante de probabilidad** radica única y exclusivamente en la **masa total** del sistema (la suma de sus componentes), lo cual determina su interpretación matemática y su aplicabilidad en Data Science:

- **Medida Invariante (γ):** Es un vector de pesos no negativos que satisface la ecuación de balance algebraico $\gamma P = \gamma$. Su masa total es arbitraria y puede ser finita o infinita, es decir:

$$M = \sum_{i \in E} \gamma_i \in (0, \infty]$$

No representa probabilidades directas, sino proporciones relativas de tiempo o visitas esperadas entre los estados.

- **Medida de Probabilidad Invariante (π):** Es un caso particular de medida invariante ($\pi P = \pi$) que cumple estrictamente con los axiomas de Kolmogorov. Exige obligatoriamente que la masa total esté normalizada a la unidad:

$$\sum_{i \in E} \pi_i = 1$$

Permite una interpretación estadística asintótica directa (probabilidad de hallar a la cadena en el estado i en un horizonte temporal lejano, $n \rightarrow \infty$).

Observación. Si el espacio de estados E es **finito**, toda medida invariante γ posee una masa total finita ($M < \infty$), lo que garantiza que siempre se puede construir su correspondiente medida de probabilidad mediante el operador de normalización lineal $\pi_i = \frac{\gamma_i}{M}$. Si el espacio E es infinito, la suma puede divergir ($M = \infty$), imposibilitando dicha normalización (cadenas recurrentes nulas).

Observación. Este corolario es brillante, porque aparte de sustituir la recurrencia que nombrábamos en el teorema anterior por la dimensión finita del espacio de estados, nos da una medida de probabilidad única, contando incluso multiplicaciones por constantes, es decir, es realmente única.

Observación. Es fundamental advertir que para el escenario complementario de una cadena de Markov que sea **transitoria** e irreducible, el comportamiento asintótico se desestabiliza y **todo es posible**, pudiendo clasificar los sistemas en tres categorías excluyentes:

1. El proceso no admite ninguna medida invariante no trivial en absoluto.
2. El proceso admite una única medida invariante salvo constante.
3. El proceso admite al menos dos medidas invariantes linealmente independientes que no constituyen múltiplos multiplicativos una de la otra.

3.8. Positive recurrence vs Null recurrence

Sea $(X_n)_{n \geq 0}$ una cadena de Markov que evoluciona sobre un espacio de estados discreto S . Tomemos un estado genérico $i \in S$, asumamos la condición de inicio fija $X_0 = i$ y definimos el tiempo de primer paso o retorno al estado de origen mediante la variable aleatoria:

$$T_i = \inf\{m \geq 1 : X_m = i\} \in \{1, 2, \dots\} \cup \{\infty\}$$

Definimos el **tiempo medio de primer retorno** al estado i como el primer momento teórico de dicha variable aleatoria, denotado por $m_i = E_i[T_i]$.

Bajo este marco, se desprenden los siguientes comportamientos asintóticos:

- Si el estado i es **transitorio**, la variable toma el valor infinito con probabilidad estrictamente positiva, es decir

$$P_i(T_i = \infty) = 1 - h_i > 0$$

lo que fuerza a que el tiempo medio de retorno diverja de manera automática: $m_i := \infty$, veámoslo fácilmente así

$$m_i = \sum_{n=1}^{\infty} n \cdot P_i(T_i = n) + \underbrace{\infty \cdot P_i(T_i = \infty)}_{\text{Miramos este término}} = (\text{Algo positivo}) + \infty(1 - h_i) \stackrel{(*)}{=} \infty$$

donde en $(*)$ hemos usado que $1 - h_i > 0$.

- Si el estado i es **recurrente**, el sistema regresa con total certeza en tiempo finito, es decir

$$P_i(T_i < \infty) = h_i = 1 \quad \text{y} \quad P_i(T_i = \infty) = 0$$

No obstante, la serie del valor esperado puede converger o divergir, pues la serie queda así definida

$$m_i = \sum_{n=1}^{\infty} n \cdot P_i(T_i = n) + \infty \cdot P_i(T_i = \infty) = \sum_{n=1}^{\infty} n \cdot P_i(T_i = n) + \infty(0) = \sum_{n=1}^{\infty} n \cdot P_i(T_i = n)$$

dando lugar a una subdivisión fundamental de los estados recurrentes

Definición 3.25 (Estado Positivo Recurrente). Se dice que un estado $i \in S$ es **positivo recurrente** si su tiempo medio de primer retorno es estrictamente finito:

$$m_i = E_i[T_i] < \infty$$

Observación. Todo estado positivo recurrente es, por implicación directa, un estado recurrente.

Definición 3.26 (Estado Nulo Recurrente). Se dice que un estado $i \in S$ es **nulo recurrente** si el proceso regresa a él con probabilidad unitaria pero el valor esperado del tiempo necesario para que se efectúe dicho retorno es infinito:

$$m_i = E_i[T_i] = \infty \quad \text{e} \quad i \text{ es recurrente}$$

3.8.1. Convergencia a la Medida de Probabilidad Invariante

Teorema 3.13 (Ley Fuerte de los Grandes Números - Caso Clásico). Sea $\{Y_i\}_{i \geq 1}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas (i.i.d.) definidas sobre un espacio de probabilidad común $(\Omega, \mathcal{F}, P)$, cuyo primer momento absoluto es estrictamente finito ($E[|Y_1|] < \infty$). Entonces, el promedio muestral converge con certeza casi segura (c.s.) hacia su esperanza matemática teórica:

$$\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m Y_i \stackrel{\text{c.s.}}{=} E[Y_1]$$

Expresado formalmente mediante la medida de probabilidad de los sucesos del espacio muestral, la identidad se escribe como:

$$P \left(\omega \in \Omega \text{ tal que } \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m Y_i(\omega) = E[Y_1] \right) = 1$$

Observación (Temario de Repaso Metodológico). Para la correcta interpretación de los teoremas ergódicos de convergencia asintótica, es imperativo realizar un repaso riguroso de los cuatro modos fundamentales de convergencia de sucesiones de variables aleatorias:

1. **Convergencia en Distribución (o en Ley):** enfocada en el límite de las funciones de distribución acumulada, dado como sigue

$$X_n \xrightarrow{d} X \iff \lim_{n \rightarrow \infty} F_n(x) = F(x) \quad \forall x \in \mathbb{R} \text{ donde } F \text{ es continua}$$

donde $F_n(x) = P(X_n \leq x)$ y $F(x) = P(X \leq x)$.

2. **Convergencia en Probabilidad:** Gobernada por el colapso de la medida de las desviaciones epsilon, dado como sigue

$$X_n \xrightarrow{P} X \iff \forall \epsilon > 0, \quad \lim_{n \rightarrow \infty} P(|X_n - X| \geq \epsilon) = 0$$

3. **Convergencia Casi Seguro (c.s.):** Concentrada en la certeza unitaria del límite puntual de las trayectorias ω , dado como sigue

$$X_n \xrightarrow{c.s.} X \iff P\left(\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right\}\right) = 1$$

4. **Convergencia en Media r -ésima (o en espacios L^r):** Basada en la norma del límite de los momentos del error, dado como sigue

$$X_n \xrightarrow{L^r} X \iff \lim_{n \rightarrow \infty} E[|X_n - X|^r] = 0$$

Esquema de implicaciones fundamentales de repaso:

Casi Seguro $\implies$ En Probabilidad $\implies$ En Distribución

Teorema 3.14 (Teorema Ergódico de Ocupación). *Considere una cadena de Markov irreducible $(X_n)_{n \geq 0}$ que opera sobre un espacio de estados S con una distribución inicial arbitraria $\alpha = (\alpha_i)_{i \in S}$. Se verifican los siguientes comportamientos asintóticos para las frecuencias de visita:*

1. *Si la cadena de Markov es transitoria o nulo recurrente, entonces para todo estado $i \in S$ se tiene que la proporción temporal de permanencia colapsa a cero casi con total certeza (c.s.):*

$$\lim_{m \rightarrow \infty} \frac{V_i(m)}{m} = 0 \quad c.s.$$

donde la variable aleatoria $V_i(m)$ cuantifica el número acumulado de visitas realizadas al estado i dentro del horizonte temporal m :

$$V_i(m) = \sum_{k=0}^{m-1} I_{\{X_k=i\}}$$

2. *Si la cadena de Markov es positivo recurrente con medida de probabilidad invariante $\lambda = (\lambda_i)_{i \in S}$, entonces para todo estado $i \in S$ la proporción temporal de visitas converge con certeza casi segura hacia dicha masa estacionaria:*

$$\lim_{m \rightarrow \infty} \frac{V_i(m)}{m} = \lambda_i \quad c.s.$$

Corolario 3.14.1. *Consideremos una cadena de Markov positivo recurrente $(X_n)_{n \geq 0}$ dotada de una probabilidad invariante $\pi = (\pi_i)_{i \in S}$, y sea $f : S \rightarrow \mathbb{R}$ una función acotada real discreta definida sobre el espacio de estados. Entonces, el promedio muestral de la función a lo largo de la trayectoria converge casi seguro hacia su media espacial teórica de equilibrio:*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) = \sum_{i \in S} \pi_i f_i = E_\pi[f] \quad c.s.$$

donde denotamos de forma abreviada las evaluaciones locales como $f_i = f(i)$.

Demostración. Para la demostración tendremos en cuenta estos puntos:

1. Denotamos la esperanza espacial límite como la constante escalar

$$E_\pi[f] = \sum_{i \in S} f_i \pi_i$$

2. Sin pérdida de generalidad, asumimos que la función está acotada en norma por la unidad, es decir

$$\|f\| = \max_{i \in S} |f_i| \leq 1$$

Cualquier función acotada general se puede reducir a este caso dividiendo toda la expresión por su norma máxima, sin alterar la convergencia.

3. Definimos

$$V_i(n) = \sum_{k=0}^{n-1} I_{\{X_k=i\}}$$

como la variable aleatoria que cuenta el número de visitas totales de la trayectoria al estado i hasta el horizonte temporal $n - 1$.

Reagrupando los términos del promedio temporal según los estados visitados, se cumple la identidad exacta:

$$\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) = \frac{1}{n} \sum_{i \in S} V_i(n) f_i = \sum_{i \in S} \frac{V_i(n)}{n} f_i$$

donde el término $\frac{V_i(n)}{n}$ representa la proporción de tiempo de estancia en el estado i , y $f(X_k)$ es la evaluación de la función en el estado j visitado en el paso k .

Fase 1. Fijamos un subconjunto arbitrario y finito de estados $\mathcal{J} \subseteq S$. Estudiamos el módulo de la diferencia entre el promedio temporal y la media espacial límite mediante la propiedad del valor absoluto y la desigualdad triangular:

$$\left| \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) - E_\pi[f] \right| = \left| \sum_{i \in S} \frac{V_i(n)}{n} f_i - \sum_{i \in S} \pi_i f_i \right| = \left| \sum_{i \in S} \left(\frac{V_i(n)}{n} - \pi_i \right) f_i \right|$$

Aplicando la propiedad distributiva del módulo sobre el sumatorio, y sabiendo que por hipótesis $|f_i| \leq 1$:

$$\left| \sum_{i \in S} \left(\frac{V_i(n)}{n} - \pi_i \right) f_i \right| \leq \sum_{i \in S} \left| \frac{V_i(n)}{n} - \pi_i \right| \cdot |f_i| \leq \sum_{i \in S} \left| \frac{V_i(n)}{n} - \pi_i \right|$$

Para gestionar un espacio S potencialmente infinito, descomponemos analíticamente el sumatorio en dos regiones complementarias usando nuestro conjunto finito $\mathcal{J}$ y su complemento $\mathcal{J}^c$ (los estados que no pertenecen a $\mathcal{J}$):

$$= \sum_{i \in \mathcal{J}} \left| \frac{V_i(n)}{n} - \pi_i \right| + \sum_{i \notin \mathcal{J}} \left| \frac{V_i(n)}{n} - \pi_i \right|$$

En el segundo sumatorio (el de la cola infinita del espacio, $i \notin \mathcal{J}$), aplicamos de forma directa la desigualdad triangular standard $|a-b| \leq |a|+|b|$. Como tanto $\frac{V_i(n)}{n}$ como π_i son magnitudes no negativas, sus valores absolutos se eliminan:

$$\left| \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) - E_\pi[f] \right| \leq \sum_{i \in \mathcal{J}} \left| \frac{V_i(n)}{n} - \pi_i \right| + \sum_{i \notin \mathcal{J}} \frac{V_i(n)}{n} + \sum_{i \notin \mathcal{J}} \pi_i$$

Antes de seguir con el proceso hay que tener claro estas dos igualdades

$$\begin{aligned}\sum_{i \in S} \frac{V_i(n)}{n} = 1 &\implies \sum_{i \notin \mathcal{J}} \frac{V_i(n)}{n} = 1 - \sum_{i \in \mathcal{J}} \frac{V_i(n)}{n} \\ \sum_{i \in S} \pi_i = 1 &\implies 1 = \sum_{i \in \mathcal{J}} \pi_i + \sum_{i \notin \mathcal{J}} \pi_i\end{aligned}$$

Para eliminar la dependencia muestral de la variable aleatoria $V_i(n)$ en el término central de la cola, usaremos las igualdades indicadas antes así

$$\begin{aligned}\sum_{i \notin \mathcal{J}} \frac{V_i(n)}{n} &= \left(\sum_{i \in \mathcal{J}} \pi_i + \sum_{i \notin \mathcal{J}} \pi_i \right) - \sum_{i \in \mathcal{J}} \frac{V_i(n)}{n} = \\ &= \sum_{i \notin \mathcal{J}} \pi_i + \sum_{i \in \mathcal{J}} \left(\pi_i - \frac{V_i(n)}{n} \right) \leq \\ &\leq \sum_{i \notin \mathcal{J}} \pi_i + \sum_{i \in \mathcal{J}} \left| \frac{V_i(n)}{n} - \pi_i \right|\end{aligned}$$

Aplicando esta última desigualdad obtenida tenemos finalmente que

$$\begin{aligned}\left| \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) - E_\pi[f] \right| &\leq \sum_{i \in \mathcal{J}} \left| \frac{V_i(n)}{n} - \pi_i \right| + \left(\sum_{i \in \mathcal{J}} \left| \frac{V_i(n)}{n} - \pi_i \right| + \sum_{i \notin \mathcal{J}} \pi_i \right) + \sum_{i \notin \mathcal{J}} \pi_i = \\ &= 2 \sum_{i \in \mathcal{J}} \left| \frac{V_i(n)}{n} - \pi_i \right| + 2 \sum_{i \notin \mathcal{J}} \pi_i\end{aligned}$$

Fase 2. Por el Teorema Ergódico de Ocupación a largo plazo en cadenas positivo recurrentes, sabemos que la proporción del tiempo de ocupación converge casi con certeza a la probabilidad invariante del estado:

$$P \left(\lim_{m \rightarrow \infty} \frac{V_i(m)}{m} = \pi_i \right) = 1 \quad \forall i \in S$$

Dado un umbral de tolerancia arbitrario $\varepsilon > 0$, procedemos a controlar los dos bloques de la cota mediante elecciones de diseño secuenciales:

1. **Control de la cola infinita ($\mathcal{J}^c$):** Dado que

$$\sum_{i \in S} \pi_i = 1$$

es una serie convergente, la masa de probabilidad residual de las colas tiende a cero. Por tanto, existe y podemos elegir un subconjunto finito $\mathcal{J}$ lo suficientemente grande tal que la masa de probabilidad fuera de él sea inferior a $\frac{\varepsilon}{4}$:

$$\sum_{i \notin \mathcal{J}} \pi_i < \frac{\varepsilon}{4}$$

2. **Control de la región finita ($\mathcal{J}$):** Una vez fijado y estabilizado el conjunto finito $\mathcal{J}$ en el paso anterior, invocamos la convergencia puntual de las variables de ocupación. Al ser una suma de un número finito de términos convergiendo a cero, se garantiza que existe un horizonte temporal determinista $N(\omega)$ tal que, para todo instante $m > N(\omega)$, la desviación acumulada es menor a $\frac{\varepsilon}{4}$:

$$\sum_{i \in \mathcal{J}} \left| \frac{V_i(m)}{m} - \pi_i \right| < \frac{\varepsilon}{4}$$

de aquí se hereda la convergencia casi segura.

Sustituyendo ambas acotaciones de control en la desigualdad obtenida al final de la Fase 1, la expresión colapsa directamente en el umbral ε :

$$2 \sum_{i \in \mathcal{J}} \left| \frac{V_i(m)}{m} - \pi_i \right| + 2 \sum_{i \notin \mathcal{J}} \pi_i \leq 2 \left(\frac{\varepsilon}{4} \right) + 2 \left(\frac{\varepsilon}{4} \right) = \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

Al cumplirse la cota superior para cualquier valor real positivo de ε , queda formalmente demostrado que el límite de la diferencia es cero, consolidando la convergencia del teorema ergódico. □

Definición 3.27 (Estado Aperiódico). Se dice que un estado $i \in S$ de una cadena de Markov es **aperiódico** si existe un horizonte de pasos finito a partir del cual las probabilidades de retorno diagonal son estrictamente positivas de forma ininterrumpida, es decir, $P_{ii}^{(m)} > 0$ para todo m suficientemente grande.

Lema 3.15 (Solidaridad de la Aperiodicidad). *Supongamos que la cadena de Markov $(X_n)_{n \geq 0}$ es irreducible y cuenta con al menos un estado aperiódico $i \in S$. Entonces, para cada par de estados del espacio $j, k \in S$, se verifica que $P_{jk}^{(m)} > 0$ para todo paso m suficientemente grande. En particular, la aperiodicidad es una propiedad de clase, lo que implica que todos los estados de la cadena son aperiódicos.*

Demostración. Debido a que la cadena de Markov es irreducible, todos los estados del espacio se comunican entre sí. Por lo tanto, para cualquier par de estados arbitrarios $j, k \in S$:

- Existe un camino para viajar de j a i en un número finito de pasos $r \geq 0$, tal que $p_{ji}^{(r)} > 0$.
- Existe un camino para viajar de i a k en un número finito de pasos $s \geq 0$, tal que $p_{ik}^{(s)} > 0$.

Evaluamos la probabilidad de transición de j a k en un horizonte temporal de $r + m + s$ pasos. Utilizando las ecuaciones de Chapman-Kolmogorov, forzamos al sistema a pasar por el estado intermedio i y acotamos inferiormente:

$$p_{jk}^{(r+m+s)} = \sum_{x \in S} \sum_{y \in S} p_{jx}^{(r)} p_{xy}^{(m)} p_{yk}^{(s)} \geq p_{ji}^{(r)} p_{ii}^{(m)} p_{ik}^{(s)}$$

Analizando los tres términos del producto del extremo derecho:

- Por ser irreducible dijimos que $p_{ji}^{(r)} > 0$ y $p_{ik}^{(s)} > 0$.
- Por hipótesis, sabemos que $p_{ii}^{(m)} > 0$ para todo $m \geq M_i$.

Dado que el producto de términos estrictamente positivos es positivo, se cumple analíticamente que:

$$p_{jk}^{(r+m+s)} \geq p_{ji}^{(r)} p_{ii}^{(m)} p_{ik}^{(s)} > 0 \quad \forall m \geq M_i$$

Haciendo un cambio de variable para el tiempo total $n = r + m + s$, tenemos que para n suficientemente grande (m era suficientemente grande), tenemos que

$$p_{jk}^{(n)} > 0$$

Para la segunda parte del enunciado que dice que todos los estados son aperiódicos, es tan fácil como realizar el mismo desarrollo para $k = j$ y lo tendríamos. □

Teorema 3.16 (Teorema de Convergencia Estacionaria). *Considere una cadena de Markov $(X_n)_{n \geq 0}$ que sea simultáneamente aperiódica, irreducible y positivo recurrente, gobernada por una matriz de probabilidad de transición P y dotada de una medida de probabilidad invariante $\lambda = (\lambda_i)_{i \in S}$.*

Entonces, bajo cualquier distribución inicial arbitraria del sistema, la distribución marginal de ocupación de cada estado en el paso m converge de forma asintótica hacia su correspondiente masa de probabilidad estacionaria pura:

$$\lim_{m \rightarrow \infty} P[X_m = j] = \lambda_j \quad \forall j \in S$$

Demostración. Para demostrar este resultado en espacios de estados generales, utilizaremos una técnica probabilística avanzada conocida como **método de acoplamiento** (*coupling*). La estrategia consiste en construir un espacio de probabilidad ampliado donde coexistan dos realizaciones independientes de la cadena y estudiar el instante probabilístico en el que sus trayectorias se intersectan geoméricamente.

Step 1. Definimos dos procesos estocásticos de tiempo discreto, denotados como $(X_m)_{m \geq 0}$ y $(Y_m)_{m \geq 0}$, gobernados bajo el mismo espacio de probabilidad subyacente y caracterizados por las siguientes condiciones estructurales:

- 1) **Cadena Original:** El proceso $(X_m)_{m \geq 0}$ es la cadena de Markov objeto de nuestro estudio, cuya evolución temporal está determinada por la matriz de transición P y arranca con la distribución inicial arbitraria α , es decir

$$P[X_0 = i] = \alpha_i$$

- 2) **Cadena Estacionaria de Referencia:** El proceso $(Y_m)_{m \geq 0}$ es una réplica exacta en su dinámica, lo que significa que posee la **misma** matriz de probabilidad de transición P . Sin embargo, se inicializa de forma idealizada utilizando la medida invariante λ como su distribución inicial, es decir

$$P[Y_0 = k] = \lambda_k$$

Por las propiedades de la estacionariedad, sabemos analíticamente que Y_m mantiene dicha distribución para todo instante de tiempo:

$$P[Y_m = j] = \lambda_j \quad \forall m \geq 0, \forall j \in S$$

esto lo vemos fácilmente con inducción como sigue

- Para el instante inicial ($m = 0$)

$$P[Y_0 = j] = \lambda_j \quad \forall j \in S$$

- Para el paso 1 ($m = 1$)

$$P[Y_1 = j] = \sum_{i \in S} P[Y_0 = i] \cdot P[Y_1 = j \mid Y_0 = i] = \sum_{i \in S} \lambda_i p_{ij} \stackrel{(*)}{=} \lambda_j$$

donde en $(*)$ hemos usado la propiedad de que λ es estacionaria ($\lambda P = \lambda$).

- Por inducción para cualquier paso m ($m \geq 0$):

$$P[Y_m = j] = \sum_{i \in S} P[Y_{m-1} = i] \cdot P[Y_m = j \mid Y_{m-1} = i] \stackrel{(*)}{=} \sum_{i \in S} \lambda_i p_{ij} = \lambda_j$$

donde en $(*)$ hemos usado la hipótesis de inducción que dice que $P[Y_{m-1} = i] = \lambda_i$.

- 3) **Independencia Estructural:** Los dos procesos son completamente independientes entre sí desde el origen del tiempo, lo que formalmente denotamos mediante el símbolo de ortogonalidad estocástica:

$$(X_m)_{m \geq 0} \perp\!\!\!\perp (Y_m)_{m \geq 0}$$

Fijamos un estado testigo cualquiera del espacio de estados, al que denotaremos como $b \in S$. Definimos la variable aleatoria T como el primer instante del reloj en el cual ambas cadenas coinciden físicamente sobre dicho estado de referencia b :

$$T = \inf \{m \geq 0 : X_m = Y_m = b\} \in \mathbb{N} \cup \{\infty\}$$

Esta variable mapea el primer choque.^o cruce de trayectorias. Si las dos cadenas de Markov evolucionaran de forma paralela sin intersectarse jamás en el punto b , el conjunto evaluado sería vacío y, por convención, el tiempo de acoplamiento tomaría el valor infinito ($T = \infty$).

Step 2. Para garantizar que las cadenas se van a cruzar con total certeza, necesitamos demostrar rigurosamente que el evento de colisión ocurre en tiempo finito casi seguro, es decir:

$$P(T < \infty) = 1$$

Para analizar matemáticamente este comportamiento, fusionamos los dos procesos individuales en un único sistema vectorial dinámico. Definimos el proceso bidimensional agrupado $W_m = (X_m, Y_m)$, el cual toma valores sobre el espacio de estados producto cartesiano denotado como $\tilde{S} = S \times S$.

Debido a la independencia mutua de las componentes de partida decretada en el Paso 1, el proceso conjunto W_m hereda de forma limpia la estructura de una cadena de Markov homogénea en el tiempo. Sus parámetros analíticos quedan determinados por las siguientes componentes:

- **Probabilidades de Transición Conjuntas:** la probabilidad de pasar del estado compuesto (i, k) al estado compuesto (j, l) en un paso se calcula mediante el producto directo de las probabilidades marginales individuales, dado que las cadenas se mueven sin afectarse entre sí:

$$\tilde{p}_{(i,k),(j,l)} = P[X_{m+1} = j, Y_{m+1} = l \mid X_m = i, Y_m = k] = p_{ij} \cdot p_{kl}$$

- **Distribución Inicial del Espacio Producto:** la ley de probabilidad inicial de la cadena conjunta en el origen del tiempo $W_0 = (X_0, Y_0)$ se construye multiplicando las respectivas distribuciones de arranque de los procesos marginales:

$$\mu_{(i,k)} = P[W_0 = (i, k)] = P[X_0 = i, Y_0 = k] = \alpha_i \cdot \lambda_k$$

Por aperiodicidad e irreducibilidad, tenemos que $\exists N = N(i, j, k, l)$ tal que para $m > N$ tenemos:

$$\tilde{p}_{(i,k),(j,l)}^{(m)} = p_{ij}^{(m)} \cdot p_{kl}^{(m)} > 0$$

Por lo tanto, W_m es irreducible, y $\tilde{\lambda} = \lambda_i \cdot \lambda_k$ es su distribución de probabilidad invariante, entonces

$$W_m \text{ es positivo recurrente} \implies P(T < \infty) = 1$$

donde sabemos la implicación final, porque al ser positivo recurrente, irreducible y aperiódico, tenemos que el estado (b, b) es visitado infinitas veces casi seguro, y por lo tanto, la primera visita a (b, b) ocurre en un tiempo finito con probabilidad 1.

Step 3. Definamos el proceso:

$$Z_m = \begin{cases} X_m & m < T \\ Y_m & m \geq T \end{cases}$$

donde Z_m sigue la trayectoria de la cadena original X_m al principio. En el momento exacto T en el que se cruza con Y_m , Z_m se “salta” a la trayectoria de Y_m y la sigue para siempre.

Este proceso es una cadena de Markov con distribución de probabilidad inicial α y las mismas probabilidades de transición P que X_m, Y_m .

Como Y_m comienza en la distribución estacionaria λ , se cumple que

$$P(Y_m = j) = \lambda_j \quad \forall j \in S$$

Por construcción, X_m y Z_m tienen la misma distribución en cada paso de tiempo m

$$P(X_m = j) = P(Z_m = j) \quad \forall j \in S$$

pues antes de T son idénticas y después de T se comporta como Y_m , pero como ya se tiene un punto de partida se usa la matriz P (la misma para Y_m y para X_m) y no la medida λ . Tenemos por tanto el siguiente desarrollo

$$\begin{aligned}
 0 \leq |P(X_m = j) - \lambda_j| &= |P(X_m = j) - P(Y_m = j)| = |P(Z_m = j) - P(Y_m = j)| = \\
 &= |P(X_m = j, m < T) + P(Y_m = j, m \geq T) - P(Y_m = j)| = \\
 &= |P(X_m = j, m < T) - P(Y_m = j, m < T)| \leq \\
 &\leq |P(X_m = j, m < T)| + |P(Y_m = j, m < T)| \leq \\
 &\stackrel{(*)}{\leq} P(m < T) + P(m < T) = 2P(T > m)
 \end{aligned}$$

donde en $(*)$ hemos usado que $P(A \cap B) \leq P(B)$. Tenemos finalmente

$$\begin{aligned}
 \lim_{m \rightarrow \infty} 0 &\leq \lim_{m \rightarrow \infty} |P(X_m = j) - \lambda_j| \leq 2 \lim_{m \rightarrow \infty} P(T > m) \implies \\
 \implies 0 &\leq \lim_{m \rightarrow \infty} |P(X_m = j) - \lambda_j| \leq P(T = \infty) = 0 \implies \\
 \implies \lim_{m \rightarrow \infty} |P(X_m = j) - \lambda_j| &= 0
 \end{aligned}$$

□

3.9. Reversible Markov Chain

Recordemos la definición de cadena de Márkov reversible.

Definición 3.28 (Balance Detallado). Se dice que una cadena de Markov discreta con matriz de transición $P = (p_{ij})$ es **reversible** respecto a una distribución de probabilidad macroscópica $\pi = (\pi_i)_{i \in \mathcal{S}}$ si satisface de manera unívoca el sistema de ecuaciones algebraicas lineales denominado **Condición de Balance Detallado**:

$$p_{ij}\pi_j = p_{ji}\pi_i \quad \forall (i, j) \in \mathcal{S} \times \mathcal{S}$$

Físicamente, esta condición impone una simetría estadística temporal absoluta: el flujo neto de probabilidad que transita desde el estado j hacia el estado i en equilibrio es idéntico al flujo en sentido inverso. Una consecuencia analítica directa de este principio es que cualquier distribución π que verifique el balance detallado constituye automáticamente una **distribución estacionaria e invariante** del proceso ($\pi = \pi P$).

3.9.1. Metropolis Algorithm

Para garantizar de manera unívoca el cumplimiento de la relación cinemática

$$\frac{P(\underline{s}' | \underline{s})}{P(\underline{s} | \underline{s}')} = e^{-\beta \Delta H}$$

sobre un computador, Metropolis y su equipo dividieron la probabilidad de transición en dos fases consecutivas independientes: la probabilidad de proponer un estado $q(\underline{s}' | \underline{s})$ (simétrica) y la probabilidad de aceptarlo $\alpha(\underline{s}' | \underline{s})$. Esto formaliza el siguiente procedimiento iterativo de muestreo estocástico:

Teorema 3.17 (Metropolis Algorithm). *Dada una configuración de partida y una temperatura inversa fija β , el proceso evoluciona de acuerdo con las siguientes reglas estocásticas paso a paso:*

- 1) Inicialización:** Comenzar la simulación desde un estado de espín macroscópico inicial $\underline{s}$.
- 2) Mutación Local:** Seleccionar un único espín de la red cristalina de forma aleatoria uniforme e invertir su orientación ($S_i \rightarrow -S_i$) para construir una configuración de prueba tentativa denotada como $\underline{s}'$.

- 3) **Evaluación Energética:** Computar la barrera de energía potencial entre ambos estados calculando el gradiente local del Hamiltoniano:

$$\Delta H = H(\underline{s}') - H(\underline{s})$$

- 4) **Criterio de Descenso Energético:** Si $\Delta H \leq 0$, el movimiento propuesto conduce a una configuración energéticamente más estable o degenerada. El sistema **acepta el movimiento con probabilidad unitaria**: se almacena $\underline{s}'$ como el nuevo estado actual de la cadena y se reinicia el ciclo desde el Paso 2.

- 5) **Criterio de Activación Térmica** ($\Delta H > 0$): Si el movimiento propuesto impone un sobre coste energético, la transición se ve condicionada por las fluctuaciones del baño térmico. Se selecciona un número pseudoaleatorio auxiliar α muestreado bajo una distribución uniforme continua en el intervalo abierto estándar:

$$\alpha \sim U(0, 1)$$

- 6) **Aceptación Estocástica:** Si el umbral aleatorio satisface la desigualdad:

$$\alpha < e^{-\beta \Delta H}$$

Se acepta la transición a pesar del incremento energético. Se guarda $\underline{s}'$ como el siguiente eslabón de la trayectoria y se reinicia el ciclo.

- 7) **Rechazo y Conservación de Masa:** Si no se cumple la desigualdad anterior, el movimiento de prueba es denegado. El sistema permanece estático haciendo $\underline{s} = \underline{s}$. Se vuelve a registrar la configuración original $\underline{s}$ como el estado consecutivo de la serie cronológica (acumulando peso estadístico en el histograma de ocupación) y se reinicia el proceso.

4. Procesos de Renovación y CTRW

4.1. Del paseo aleatorio discreto al proceso de renovación

Definición 4.1. Un **paseo aleatorio discreto** es una sucesión $\{X_i\}_{i=1}^{\infty}$ de variables i.i.d. en $\mathbb{R}$ que modelan la magnitud de cada salto espacial. La posición tras m saltos es

$$Y_m = \sum_{i=1}^m X_i.$$

Si $X_i \sim N(\mu, \sigma)$, por estabilidad Gaussiana Y_m tiene densidad $f_{Y_m}(u) = f_{X_1}^{*m}(u)$ (convolución de orden m), con $\mathbb{E}[Y_m] = m\mu$ y $\text{Var}[Y_m] = m\sigma^2$. La idea importante a retener no es el caso Gaussiano en sí, sino que **la densidad de una suma de variables i.i.d. es siempre la convolución iterada de la densidad base**, sea o no Gaussiana.

Tiempos de espera y proceso de renovación

Para pasar a tiempo continuo se desacopla el "número de saltos" del "tiempo real" introduciendo tiempos de espera.

Definición 4.2. Sea $\{J_i\}$ una sucesión i.i.d. de variables **estrictamente positivas** (tiempos de espera entre saltos). El instante del m -ésimo salto es

$$T_m = \sum_{i=1}^m J_i, \quad T_0 = 0.$$

$\{T_m\}$ es un **proceso de renovación**: tras cada salto el "cronómetro" se reinicia con la misma distribución (el sistema "olvida" cuánto llevaba esperando).

Definición 4.3. El **proceso de conteo** asociado es

$$N(t) = \max\{n \in \mathbb{N}_0 : T_n \leq t\},$$

es decir, cuántos saltos han ocurrido hasta el instante t . Sus trayectorias son **càdlàg** (continuas por la derecha, con límite por la izquierda).

Ejemplo. Si $J_1 \sim \text{Exp}(\lambda)$, entonces T_m sigue una **distribución de Erlang** (caso particular de la Gamma con parámetro de forma entero):

$$f_{T_m}(u) = \lambda^m u^{m-1} \frac{e^{-\lambda u}}{(m-1)!}, \quad u \geq 0.$$

4.2. El proceso CTRW: subordinando el espacio al tiempo

Definición 4.4 (CTRW). El **paseo aleatorio en tiempo continuo** (Continuous Time Random Walk) es el proceso compuesto

$$Y(t) = Y_{N(t)} = \sum_{i=1}^{N(t)} X_i,$$

es decir, la cadena de saltos espaciales Y_m "subordinada" al reloj aleatorio $N(t)$.

Observación. Aunque (Y_m) y (T_m) son, cada una por separado, cadenas de Markov en su índice de pasos m , y aunque el par (T_m, Y_m) es Markoviano si $(Y_m) \perp (T_m)$, el proceso final $(Y(t))_{t \geq 0}$ **no es Markoviano en general**: la probabilidad de saltar depende de cuánto tiempo lleva ya esperando la partícula (su “edad”), no solo de su posición actual.

4.3. Markov, semi-Markov y el rol de la edad $\gamma(t)$

Para formalizar esto se usa un **espacio filtrado** $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, P)$, donde $\mathcal{F}_t$ recoge toda la historia conocida hasta t . La propiedad de Markov clásica exige

$$P(X_s \in B \mid \mathcal{F}_t) = P(X_s \in B \mid X_t = x), \quad s > t,$$

es decir, que el futuro solo dependa del presente, no de toda la historia.

Para analizar el CTRW se definen, en torno al instante t (entre el último salto T_3 y el próximo T_4 , digamos):

- **Edad** $\gamma(t) = t - T_3$: tiempo transcurrido desde el último salto. Es conocida (medible respecto a $\mathcal{F}_t$).
- **Vida residual** $\sigma(t) = T_4 - t$: tiempo que falta hasta el próximo salto. Es desconocida y depende de $\gamma(t)$.

Como la vida residual depende de la edad acumulada, el sistema es un **proceso semi-Markoviano**.

Proposición 4.1 (Vectorización de Markov). *Aunque $Y(t)$ por sí solo no es Markoviano, el par $(Y(t), \gamma(t))$ sí es un proceso de Markov en tiempo continuo: basta añadir la edad al espacio de estados para recuperar la falta de memoria.*

Proposición 4.2 (Condición de Markovianidad pura).

$$Y(t) \text{ es Markoviano} \iff J_i \sim \text{Exp}(\lambda) \quad \forall i.$$

*La exponencial es la **única** distribución de espera sin memoria, así que es la única que evita tener que aumentar el espacio de estados con $\gamma(t)$.*

4.4. Distribución unipuntual y la ecuación de Montroll-Weiss

Definición 4.5. $q(t, u)$ es la densidad de probabilidad de encontrar la partícula en la posición u en el instante t (**f.d.p. unipuntual** de $Y(t)$).

Calcular $q(t, u)$ directamente en (t, u) es muy complicado por las infinitas convoluciones que involucra. La solución práctica es pasar al **espacio dual** aplicando Fourier en el espacio ($u \rightarrow k$) y Laplace en el tiempo ($t \rightarrow s$):

- $f_{X_1}(u) \rightarrow \hat{f}_{X_1}(k)$ (función característica de los saltos)
- $f_{J_1}(t) \rightarrow \tilde{f}_{J_1}(s)$ (Laplace de los tiempos de espera)

Teorema 4.3 (Ecuación de Montroll-Weiss). *Bajo desacoplamiento (saltos y esperas independientes), la solución en el espacio dual es*

$$\tilde{q}(s, k) = \frac{\tilde{F}_{J_1}(s)}{1 - \hat{f}_{X_1}(k) \tilde{f}_{J_1}(s)},$$

donde $\tilde{F}_{J_1}(s) = \frac{1 - \tilde{f}_{J_1}(s)}{s}$ es la transformada de la función de supervivencia de J_1 .

Protocolo de uso: (1) calcular $\hat{f}_{X_1}(k)$ y $\tilde{f}_{J_1}(s)$, (2) sustituir en la fórmula, (3) invertir Fourier-Laplace para volver al espacio real. Esta es la idea central de toda la sección: el acoplamiento espacio-temporal, intratable en el espacio real, se vuelve un simple cociente algebraico en el espacio dual.

4.5. Caso particular: esperas exponenciales $\Rightarrow$ Proceso de Poisson

Ejemplo. Si $J_1 \sim \text{Exp}(\lambda)$, sustituyendo en Montroll-Weiss (sin la parte espacial) se obtiene tras invertir Laplace

$$\boxed{P(N(t) = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}} \implies N(t) \sim \text{Poi}(\lambda t).$$

Observación. Esto no es casualidad: la exponencial es la única ley sin memoria, así que es la única que convierte a $N(t)$ en un **proceso de Poisson**, caracterizado por tener **incrementos independientes y estacionarios**:

$$N(t_2) - N(t_1) \stackrel{d}{=} N(t_2 - t_1), \quad N(t_2) - N(t_1) \perp N(t_1).$$

Esta propiedad es la que define a los **procesos de Lévy** en general.

4.6. Límite difusivo: del CTRW a la ecuación del calor

Existe una forma equivalente de $q(t, u)$ escrita directamente en (t, x) , la **ecuación de renovación semi-Markoviana**, que combina "la partícula aún no ha saltado" con "la partícula ya saltó y siguió moviéndose". Aplicando Fourier-Laplace a esa ecuación se recupera exactamente Montroll-Weiss (es la misma información, vista de otra forma).

Idea clave del límite a gran escala: si los saltos tienen varianza finita y las esperas tienen media finita, cuando se observa el sistema a escalas grandes ($k, s \rightarrow 0$) las transformadas se aproximan por Taylor ($\hat{f}_{X_1}(k) \approx 1 - k^2$, $\tilde{f}_{J_1}(s) \approx 1 - s$), y reescalando convenientemente se obtiene

$$\tilde{\mu}(s, k) \xrightarrow{c \rightarrow \infty} \frac{1}{k^2 + s}.$$

Teorema 4.4 (Límite de difusión). *Esta expresión, al deshacer las transformadas, equivale al **problema de Cauchy de la ecuación de difusión (del calor)**:*

$$\left(\frac{\partial}{\partial t} - \frac{\partial^2}{\partial x^2} \right) u(t, x) = \delta(x) \delta(t), \quad u(0, x) = \delta(x),$$

cuya solución es el núcleo del calor

$$u(t, x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}.$$

Observación. Este resultado es la versión "macroscópica" del Teorema Central del Límite: a escalas grandes, **cualquier** CTRW desacoplado con saltos de varianza finita y esperas de media finita converge a un **Movimiento Browniano**, sin importar la forma exacta de las distribuciones base — solo importan sus primeros momentos.

4.7. Resumen conceptual

- El **CTRW** combina un paseo aleatorio espacial discreto (Y_m) con un reloj aleatorio dado por un **proceso de renovación** ($T_m, N(t)$).
- El resultado, $Y(t) = Y_{N(t)}$, en general **no es Markoviano**: hace falta conocer la edad $\gamma(t)$ desde el último salto para predecir el futuro (proceso semi-Markoviano). Añadiendo $\gamma(t)$ al estado, se recupera Markov.
- Solo cuando los tiempos de espera son **exponenciales** el proceso es Markoviano puro, y $N(t)$ se convierte en un **proceso de Poisson** (caso particular de proceso de Lévy).
- La herramienta analítica central es pasar al **espacio dual de Fourier-Laplace**, donde la f.d.p. unipuntual se resuelve algebraicamente (**ecuación de Montroll-Weiss**).
- A escalas grandes, todo CTRW desacoplado razonable (varianza de saltos y media de esperas finitas) converge a la **ecuación de difusión**, es decir, al **Movimiento Browniano**: es el análogo macroscópico del Teorema Central del Límite.

5. Procesos Gaussianos

5.1. Variable normal: lo básico

Definición 5.1. $X \sim N(\mu, \sigma^2)$ si tiene densidad

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Aquí μ es la media y σ^2 la varianza. Estandarizando, $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$.

Propiedades clave a recordar (sin entrar en el detalle de cada fórmula):

- Las colas de la normal decaen muy rápido (tipo $e^{-x^2/2}$): es muy poco probable alejarse de la media.
- Si X, Y son normales **independientes**, su suma vuelve a ser normal: $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.
- Un límite de normales (independientes) vuelve a ser normal.

5.2. Vector Gaussiano: la idea importante

Definición 5.2. $X = (X_1, \dots, X_n)$ es un **vector Gaussiano** si **cualquier** combinación lineal de sus componentes, $\sum_i c_i X_i$, es una variable normal.

Observación. **Ojo:** que cada componente X_i sea normal por separado **no** significa que el vector sea Gaussiano (hay contraejemplos). Lo que define al vector Gaussiano es que se comporten bien *en conjunto*, no solo de forma individual.

Un vector Gaussiano queda totalmente determinado por:

- su **vector de medias** $m = E[X]$,
- su **matriz de covarianzas** $C(X) = (\text{Cov}(X_i, X_j))_{i,j}$.

Proposición 5.1 (La propiedad más útil). Si X, Y son **conjuntamente Gaussianas**, entonces

$$X \perp\!\!\!\perp Y \iff \text{Cov}(X, Y) = 0.$$

Es decir: en el mundo Gaussiano, “no correlacionadas” e “independientes” son lo mismo (en general, fuera del caso Gaussiano, covarianza cero NO implica independencia).

Si la matriz C es invertible, el vector tiene densidad multivariante (la generalización en n dimensiones de la fórmula de la normal):

$$f_X(x) = \frac{1}{(2\pi)^{n/2}(\det C)^{1/2}} \exp\left(-\frac{1}{2}\langle x - m, C^{-1}(x - m) \rangle\right).$$

(No es necesario memorizar la deducción, solo saber que existe y que generaliza la fórmula 1D.)

5.3. Proceso Gaussiano: lo que hay que saber

Definición 5.3. $(X_t)_{t \geq 0}$ es un **proceso Gaussiano** si para cualquier elección finita de instantes $t_1, \dots, t_n$, el vector $(X_{t_1}, \dots, X_{t_n})$ es un vector Gaussiano. Queda caracterizado por:

- la **función de media** $m(t) = E[X_t]$,
- la **función de covarianza** $K(s, t) = \text{Cov}(X_s, X_t)$.

Observación. Idea clave: si me dan cualquier función de media $m(t)$ y cualquier función $K(s, t)$ que sea simétrica y "positiva definida" (una condición técnica de consistencia), el **Teorema de Existencia (vía Kolmogorov)** garantiza que existe de verdad un proceso Gaussiano con esa media y esa covarianza. Esto es justo lo que permite construir el Movimiento Browniano en la sección siguiente.

Estacionariedad (idea básica)

Un proceso es **estacionario** si su comportamiento estadístico no cambia al desplazarlo en el tiempo. Hay una versión fuerte (toda la distribución es invariante) y una débil (solo la media y la covarianza son invariantes). **Dato importante:** para procesos Gaussianos, ambas nociones **coinciden** (porque la Gaussiana queda totalmente determinada por media y covarianza).

Continuidad de las trayectorias (idea básica)

Existen resultados técnicos (Kolmogorov-Chentsov) que garantizan que, si los incrementos del proceso no crecen demasiado rápido, las trayectorias pueden tomarse **continuas** (de hecho Hölder-continuas). No hace falta el detalle, solo saber que esta propiedad es la que hace razonable dibujar trayectorias continuas de un proceso Gaussiano como el Movimiento Browniano.

5.4. Movimiento Browniano: el resultado central del tema

Definición 5.4 (Movimiento Browniano Estándar, MBE). Es un proceso Gaussiano $(B_t)_{t \geq 0}$ con

$$E[B_t] = 0, \quad \text{Cov}(B_s, B_t) = \min(s, t).$$

Observación. De dónde viene: si tomamos un CTRW con saltos cada vez más pequeños pero cada vez más frecuentes (límite difusivo visto en el tema anterior), el proceso converge exactamente a este Movimiento Browniano. Es, en cierto sentido, el ruido aleatorio más simple posible.

Propiedades que hay que recordar (las importantes):

- $B_0 = 0$ y las trayectorias son continuas.
- Para $s < t$: el incremento $B_t - B_s \sim N(0, t - s)$ — la varianza solo depende de **cuánto tiempo** ha pasado, no de en qué momento empieza.
- Los incrementos en intervalos disjuntos son **independientes** (incrementos independientes y estacionarios).
- Es un **proceso de Markov**: el futuro solo depende del valor actual, no de cómo se llegó hasta ahí.
- Tiene propiedades de simetría útiles: se puede trasladar, reflejar ($-B_t$), reescalar ($\sqrt{c} B_{t/c}$) o invertir en el tiempo, y en todos los casos se obtiene de nuevo un Movimiento Browniano.

Martingala (idea básica)

Una **martingala** es un proceso cuya mejor predicción del futuro, sabiendo todo el pasado, es simplemente su valor actual: $E[Y_t | \mathcal{F}_s] = Y_s$ para $s < t$.

Teorema 5.2. (B_t) y $(B_t^2 - t)$ son ambos martingalas.

Teorema 5.3 (Caracterización de Lévy — muy útil en la práctica). *Si un proceso (X_t) cumple: empieza en 0, tiene trayectorias continuas, y tanto (X_t) como $(X_t^2 - t)$ son martingalas, **entonces** (X_t) es un **Movimiento Browniano**.*

Observación. Esto es muy práctico: en vez de comprobar la definición Gaussiana completa, basta verificar estas tres condiciones (más fáciles de chequear) para saber que un proceso dado es un Movimiento Browniano.

5.5. Resumen en una frase por concepto

- **Vector Gaussiano:** todas sus combinaciones lineales son normales (no basta con que cada componente lo sea).
- **Independencia = covarianza cero**, pero solo si las variables son conjuntamente Gaussianas.
- **Proceso Gaussiano:** cualquier conjunto finito de instantes da un vector Gaussiano; queda definido por su media y su covarianza, y siempre existe si la covarianza es razonable” (positiva definida).
- **Estacionariedad fuerte = débil** solo en el caso Gaussiano.
- **Movimiento Browniano:** el proceso Gaussiano más simple (media 0, covarianza $\min(s, t)$); tiene incrementos independientes y normales, es Markoviano, y es una martingala. La caracterización de Lévy permite reconocerlo fácilmente sin pasar por la definición completa.

6. ¿Por qué funcionan los algoritmos MCMC?

6.1. Introducción

Buenos días a todos. Hoy vamos a responder a una pregunta fundamental en la computación y la física moderna: ¿Por qué funcionan los algoritmos MCMC?

Antes de entrar en las ecuaciones, descifremos qué significan estas siglas. MCMC significa Markov Chain Monte Carlo (o Monte Carlo por Cadenas de Markov en español). Este nombre une dos mundos de la probabilidad en un solo atajo genial:

- Monte Carlo: Es la filosofía de resolver problemas matemáticos imposibles mediante el azar, usando una computadora para generar miles de muestras aleatorias y quedarnos con su promedio.
- Cadenas de Markov: Es un mecanismo dinámico, un caminante aleatorio que se mueve paso a paso por un mapa de estados, donde su siguiente posición depende únicamente de dónde está parado justo ahora (falta de memoria).

¿Qué es lo más importante de MCMC y por qué lo necesitamos? En el mundo real —como en el Modelo de Ising que vimos en el Tema 3 o en la Estadística Bayesiana— a menudo queremos analizar una distribución de probabilidad que llamaremos $\pi(x)$. El gran drama es que esta función tiene un denominador, una constante de normalización (como la función de partición $Z(\beta)$ en física, o la evidencia en Bayes), que requiere sumar o integrar sobre espacios de estados gigantescos (¡del orden de $2^{10^{23}}$ configuraciones!). Enumerar ese espacio es físicamente imposible para cualquier ordenador.

La genialidad de MCMC es la siguiente: En lugar de intentar calcular esa constante imposible desde arriba, programamos a un caminante aleatorio inteligente (la Cadena de Markov) para que deambule por el espacio de estados. Al diseñar sus reglas de movimiento de forma astuta, el caminante visitará más seguido las zonas de alta probabilidad y menos seguido las de baja probabilidad. Al final, la trayectoria de sus pasos “dibujará” la forma exacta de la distribución $\pi(x)$.

Gracias a esto, podemos estimar esperanzas y generar muestras de π sin necesidad de conocer jamás la constante de normalización. En otras palabras, MCMC nos permite **simular** la distribución que nos interesa sin tener que calcularla explícitamente, con el objetivo de obtener resultados prácticos en problemas de física, estadística y aprendizaje automático, que suelen ser los valores promedio de ciertas funciones bajo π , por ejemplo

- ¿Cuál es la energía interna promedio ($\langle H \rangle$) del material a esta temperatura?
- ¿Cuál es la magnetización promedio ($\langle M \rangle$) para saber si el material es magnético?
- ¿Cuál es el valor medio esperado ($\langle \theta \rangle$) de este parámetro?

Matemáticamente, cualquiera de estas preguntas se traduce exactamente en la definición de esperanza

$$\mathbb{E}_{\pi}[f] = \sum_{i \in S} \pi(i) f(i)$$

donde $f(i)$ es la propiedad que mides (la energía, la magnetización, etc.) en la configuración i , multiplicada por su peso estadístico $\pi(i)$.

Hoy vamos a demostrar formalmente por qué este “atajo” de usar pasos correlacionados de un caminante funciona exactamente igual de bien que tomar muestras independientes de la distribución real, y todo ello sin necesidad de conocer jamás esa constante de normalización.

6.2. Motivación: el problema que resuelve MCMC

Recordemos el contexto (**Tema 1** y el **Modelo de Ising** del Tema 3): muchas veces queremos calcular esperanzas, o simplemente **generar muestras**, de una distribución de probabilidad π definida sobre un espacio de estados S **enorme**. Dos ejemplos que ya conocemos:

- **Modelo de Ising:** $\pi(\underline{s}) = \frac{e^{-\beta H(\underline{s})}}{Z(\beta)}$, con $|S| = 2^N$ configuraciones de espín. Para $N \sim 10^{23}$ (como en el Tema 1), enumerar el espacio es **imposible**: ni siquiera podemos calcular $Z(\beta) = \sum_{\underline{s}} e^{-\beta H(\underline{s})}$.
- **Estadística Bayesiana:** $\pi(\theta) \propto \text{verosimilitud}(\theta) \cdot \text{prior}(\theta)$, donde la constante de normalización (la evidencia) es una integral intratable.

Observación (La idea central de MCMC). En vez de calcular π directamente, **construimos una cadena de Markov** $(X_m)_{m \geq 0}$ cuya **distribución estacionaria sea exactamente** π . Si simulamos esa cadena durante suficientes pasos, las muestras X_m se comportan (aproximadamente) como muestras de π , y los promedios temporales de la trayectoria nos dan las esperanzas que buscábamos. **Toda la pregunta de “por qué funciona MCMC” se reduce a justificar rigurosamente esta frase.**

6.3. Construcción general: el Algoritmo de Metropolis-Hastings

En el Tema 3 vimos el algoritmo de Metropolis aplicado al modelo de Ising (propuesta = voltear un espín al azar, aceptar con $\min(1, e^{-\beta \Delta H})$). Generalicemos esa idea a **cualquier** distribución objetivo π sobre un espacio de estados S .

Observación. Cuando hablamos de una distribución objetivo sobre la que trabajar π **no necesitamos conocer su constante de normalización**, es decir, si

$$\pi(x) = \frac{f(x)}{Z}$$

basta con conocer $f(x)$ (la parte “salvo constante multiplicativa”) para poder construir la cadena de Markov. Esto es lo que hace que MCMC sea útil en la práctica: nunca necesitamos calcular Z ni la evidencia bayesiana.

Definición 6.1 (Metropolis-Hastings). Dada una distribución objetivo π (que solo necesitamos conocer **salvo constante multiplicativa**) y una **distribución de propuesta** $q(x, \cdot)$ (fácil de simular), el algoritmo construye una cadena de Markov (X_m) así:

1. Estando en el estado x , proponemos un estado candidato $y \sim q(x, \cdot)$.
2. Calculamos la **probabilidad de aceptación**:

$$\alpha(x, y) = \min \left(1, \frac{\pi(y) q(y, x)}{\pi(x) q(x, y)} \right)$$

3. Con probabilidad $\alpha(x, y)$ aceptamos: $X_{m+1} = y$. En caso contrario, rechazamos y nos quedamos: $X_{m+1} = x$.

Observación. La distribución de propuesta

$$q(x, \cdot)$$

es simplemente la regla de exploración del ordenador. Si el caminante está en el estado x , $q(x, y)$ es la probabilidad de que al ordenador se le ocurra sugerir el estado y como próximo destino. Es una distribución completamente arbitraria que elegimos nosotros porque es sumamente fácil y rápida de simular (generar números aleatorios con ella toma microsegundos).

Observación. para entender con detalle el segundo paso, conviene escribir el cociente de aceptación como

$$\frac{\pi(y) q(y, x)}{\pi(x) q(x, y)} = \underbrace{\left(\frac{\pi(y)}{\pi(x)} \right)}_{\text{Filtro Espacial}} \times \underbrace{\left(\frac{q(y, x)}{q(x, y)} \right)}_{\text{Filtro Temporal}}$$

donde explicaremos el concepto de ambos términos como sigue

- **Primer componente:** este bloque compara la altura o el peso real de los dos estados en nuestro “mapa objetivo”
 - Imagina el escenario: El caminante está en la casilla actual x y el ordenador le propone viajar a la casilla y .
 - Si el destino es mejor ($\pi(y) > \pi(x)$): Este cociente será un número mayor que 1. Al multiplicarlo por el segundo bloque, tenderá a hacer que la aceptación α sea alta. El algoritmo premia el movimiento hacia las cumbres de probabilidad.
 - Si el destino es peor ($\pi(y) < \pi(x)$): El cociente será un número menor que 1. Esto reduce el valor de α , obligando al caminante a “pensárselo dos veces” (lanzando la moneda estocástica) antes de descender a un valle.
- **Segundo componente:** este bloque compara la facilidad de ir de x a y con la facilidad de volver de y a x según la distribución de propuesta. Se llama **Factor de Corrección de Hastings** y sirve para compensar los sesgos del propio software de exploración.

Es de vital importancia, pues si diseñamos un algoritmo de exploración que, por culpa de cómo está programado, tiene una enorme tendencia natural a proponer pasos hacia la derecha y casi nunca propone pasos hacia la izquierda

$$q(\text{izquierda}, \text{derecha}) \gg q(\text{derecha}, \text{izquierda})$$

entonces nuestro caminante pasará la vida en el lado derecho de la pizarra, por lo que el factor de corrección de Hastings compensará este sesgo, de la siguiente manera.

Este bloque evalúa la proporción inversa

$$\frac{\text{Probabilidad de regresar } (y \rightarrow x)}{\text{Probabilidad de ir } (x \rightarrow y)}$$

si ir de x a y era “demasiado fácil” para el ordenador, el denominador $q(x, y)$ será muy grande, lo que hará que este factor de corrección se vuelva muy pequeño. Al multiplicar, penaliza y frena el paso. Castiga los caminos que son “autopistas fáciles” de propuesta y premia los caminos que son “difíciles” de proponer, garantizando que el caminante se mueva guiado por la física del problema (π) y no por los caprichos del programador (q)

Observación. El algoritmo de Metropolis del Tema 3 es el caso particular donde q es **simétrica** ($q(x, y) = q(y, x)$, p. ej. “elige un espín al azar y vuélcalo”), por lo que el cociente se reduce a

$$\frac{\pi(y) q(y, x)}{\pi(x) q(x, y)} = \frac{\pi(y)}{\pi(x)} = e^{-\beta \Delta H}$$

que es exactamente la fórmula que ya conocíamos.

Punto clave (escribir en la pizarra): en el cociente

$$\frac{\pi(y)}{\pi(x)}$$

la **constante de normalización se cancela**. Esto es lo que hace que MCMC sea útil en la práctica: nunca necesitamos calcular $Z(\beta)$ ni la evidencia bayesiana.

6.4. Pilar 1: la construcción garantiza que π es estacionaria

La probabilidad de transición resultante es

$$P(x, y) = q(x, y) \alpha(x, y), \quad \forall x \neq y \in S$$

Es importante tener claro desde el primer momento que la probabilidad de transición usada aquí, la matriz enorme P , no sabemos si realmente va a hacer el correcto trabajo a largo plazo en función de π . Para ello viendo que se cumple la condición de balance detallado (lo demostramos más adelante), tenemos que queda demostrado matemáticamente que, sin importar lo caprichosa que fuera tu q original, el filtro α corrige la trayectoria de tal manera que el caminante se ve obligado a mantener un equilibrio perfecto respecto a la distribución real π .

Proposición 6.1 (La cadena de Metropolis-Hastings es reversible respecto a π). *La matriz P construida arriba satisface la condición de **balance detallado**:*

$$\pi(x)P(x, y) = \pi(y)P(y, x) \quad \forall x, y \in S.$$

Demostración. Sin pérdida de generalidad supongamos $\pi(x)q(x, y) \leq \pi(y)q(y, x)$, con $x \neq y$, puesto que el caso $x = y$ es trivial. Entonces:

$$\alpha(x, y) = 1, \quad \alpha(y, x) = \frac{\pi(x)q(x, y)}{\pi(y)q(y, x)}.$$

Por tanto:

$$\begin{aligned} \pi(x)P(x, y) &= \pi(x)q(x, y) \cdot 1 = \pi(x)q(x, y), \\ \pi(y)P(y, x) &= \pi(y)q(y, x) \cdot \frac{\pi(x)q(x, y)}{\pi(y)q(y, x)} = \pi(x)q(x, y). \end{aligned}$$

como ambos lados coinciden, entonces se cumple el balance detallado. $\square$

Corolario 6.1.1. Como ya vimos en el Tema 3 (Cadenas de Markov Reversibles), **el balance detallado implica estacionariedad**: sumando sobre x la igualdad anterior,

$$\sum_{x \in S} \pi(x)P(x, y) = \sum_{x \in S} \pi(y)P(y, x) = \pi(y) \sum_{x \in S} P(y, x) = \pi(y), \quad \forall y \in S$$

es decir, $\pi P = \pi$. π **es, por construcción, la distribución invariante de la cadena**, sin importar lo complicada que sea π ni si conocemos su constante de normalización.

6.5. Pilar 2: la cadena converge a π (sin importar el punto de partida)

Que π sea **una** distribución estacionaria no basta: necesitamos que la cadena realmente **llegue** a ella partiendo de cualquier estado inicial. Aquí es donde entra el **Teorema de Convergencia Estacionaria** que ya demostramos en el Tema 3:

Teorema 6.2 (Recordatorio: Teorema de Convergencia Estacionaria). *Si una cadena de Markov es **irreducible**, **aperiódica** y **positivo recurrente**, con distribución invariante única π , entonces para cualquier distribución inicial α :*

$$\lim_{m \rightarrow \infty} P(X_m = j) = \pi(j) \quad \forall j \in S.$$

Este límite realmente significa que, después de un número suficientemente grande de pasos, la distribución de probabilidad del caminante se vuelve indistinguible de π , sin importar dónde empezó. En otras palabras, la cadena “olvida” su punto de partida y se estabiliza en la distribución objetivo, lo que significa que en la práctica sabemos que si descartamos los primeros K pasos de nuestra simulación, el resto de la trayectoria ya estará “limpia” del sesgo del punto de partida X_0 .

Aplicado a Metropolis-Hastings, necesitamos verificar dos condiciones **sobre el diseño del algoritmo** (no son automáticas, ¡hay que elegir bien q !):

- **Irreducibilidad:** la propuesta q debe permitir, en un número finito de pasos, llegar desde cualquier estado a cualquier otro (con probabilidad positiva). Por ejemplo, en Ising: si solo pudiéramos voltear espines en bloques fijos, podríamos quedar atrapados en una sub-región del espacio de estados.
- **Aperiodicidad:** significa que la cadena no se mueva en ciclos fijos de longitud estructurada, lo cual se obtiene con que exista **algún** estado donde la cadena pueda quedarse parada en un paso, es decir $P(x, x) > 0$. Esto ocurre automáticamente en Metropolis-Hastings en cuanto hay **algún rechazo posible** ($\alpha(x, y) < 1$ para algún y), lo cual es casi siempre el caso salvo en ejemplos degenerados.

Observación. Si el espacio de estados es **finito** (como en el Ising con $|S| = 2^N$), recordemos que “Finito + Irreducibilidad implica automáticamente positivo recurrencia (Tema 3). Así que en la práctica solo hay que preocuparse de diseñar q para que la cadena sea irreducible y aperiódica, y la convergencia hacia π queda **garantizada matemáticamente**.

El problema que tenemos actualmente, es que el teorema de convergencia nos da la distribución π para un estado $j \in S$, que es el estado final de nuestra cadena de Markov después de m pasos, es decir, $P(X_m = j)$, por lo que para poder hacer el promedio (esperanza), necesitaríamos ejecutar miles de cadenas de Markov independientes y tomar el promedio de sus estados finales. Esto es lo que nos lleva al siguiente pilar: el Teorema Ergódico, que nos permite usar **una sola trayectoria larga** para estimar esperanzas bajo π .

6.6. Pilar 3: el Teorema Ergódico – por qué podemos *estimar* con MCMC

Hasta aquí hemos justificado que X_m se parece a una muestra de π cuando m es grande. Pero en la práctica **no relanzamos la cadena miles de veces**: simulamos **una única trayectoria larga** y usamos sus valores para estimar esperanzas. Esto se justifica con el resultado que también demostramos en el Tema 3:

Teorema 6.3 (Recordatorio: Teorema Ergódico / Corolario de promedios temporales). *Si la cadena es positivo recurrente con invariante π , entonces para toda función acotada $f : S \rightarrow \mathbb{R}$:*

$$\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{k=0}^{m-1} f(X_k) = \sum_{i \in S} \pi(i) f(i) = E_{\pi}[f] \quad c.s.$$

Observación (Por qué esto es la pieza que faltaba). Este teorema es la versión markoviana de la **Ley Fuerte de los Grandes Números**: aunque los X_k de una misma trayectoria están **correlacionados** (no son i.i.d.), su promedio temporal converge igualmente al promedio espacial bajo π . Es exactamente lo que se necesita para que el “Monte Carlo” en MCMC tenga sentido: **una sola simulación larga es suficiente** para estimar $E_{\pi}[f]$ (por ejemplo, la energía interna $\langle H \rangle$ del Ising, o la media a posteriori de un parámetro bayesiano θ).

6.7. Resumen: los tres pilares

Para la pizarra, esquema final:

1. Balance detallado (diseño)	$\implies \pi$ es invariante: $\pi P = \pi$
2. Irreducible + aperiódica	$\implies P(X_m = \cdot) \xrightarrow{m \rightarrow \infty} \pi$ (converge sea cual sea X_0)
3. Teorema Ergódico	$\implies \frac{1}{m} \sum_{k=0}^{m-1} f(X_k) \xrightarrow[m \rightarrow \infty]{\text{c.s.}} E_\pi[f]$

Observación (Conclusión, en una frase). MCMC funciona porque **(1)** construimos la cadena a propósito para que π sea su distribución de equilibrio, **(2)** la teoría de cadenas de Markov garantiza que ese equilibrio se alcanza sin importar dónde empezamos, y **(3)** el Teorema Ergódico convierte una única trayectoria simulada en un estimador válido de cualquier esperanza bajo π – todo ello **sin necesitar jamás conocer la constante de normalización de π** .

Notas para la exposición oral:

- Si sobra tiempo, se puede mencionar (sin demostrar) que la **velocidad** de convergencia hacia π está gobernada por el segundo autovalor más grande en módulo de P (Tema 3, Teorema Espectral para Cadenas Reversibles): cuanto más cerca de 1 esté $|\lambda_2|$, más lento es el "mixing" de la cadena.
- Conviene dibujar en la pizarra el esquema circular: **Distribución objetivo π → Cadena de Markov diseñada (Metropolis-Hastings) → Simulación larga → Muestras / Estimaciones.**
- Buen cierre: enlazar con el Tema 1 – el Teorema de Recurrencia de Poincaré mostraba que esperar a que un sistema determinista "visite" sus estados de interés es inviable en alta dimensión; MCMC es la respuesta probabilística a ese mismo problema: en vez de esperar recurrencias deterministas, **diseñamos** una dinámica aleatoria que converge rápidamente al lugar que nos interesa.