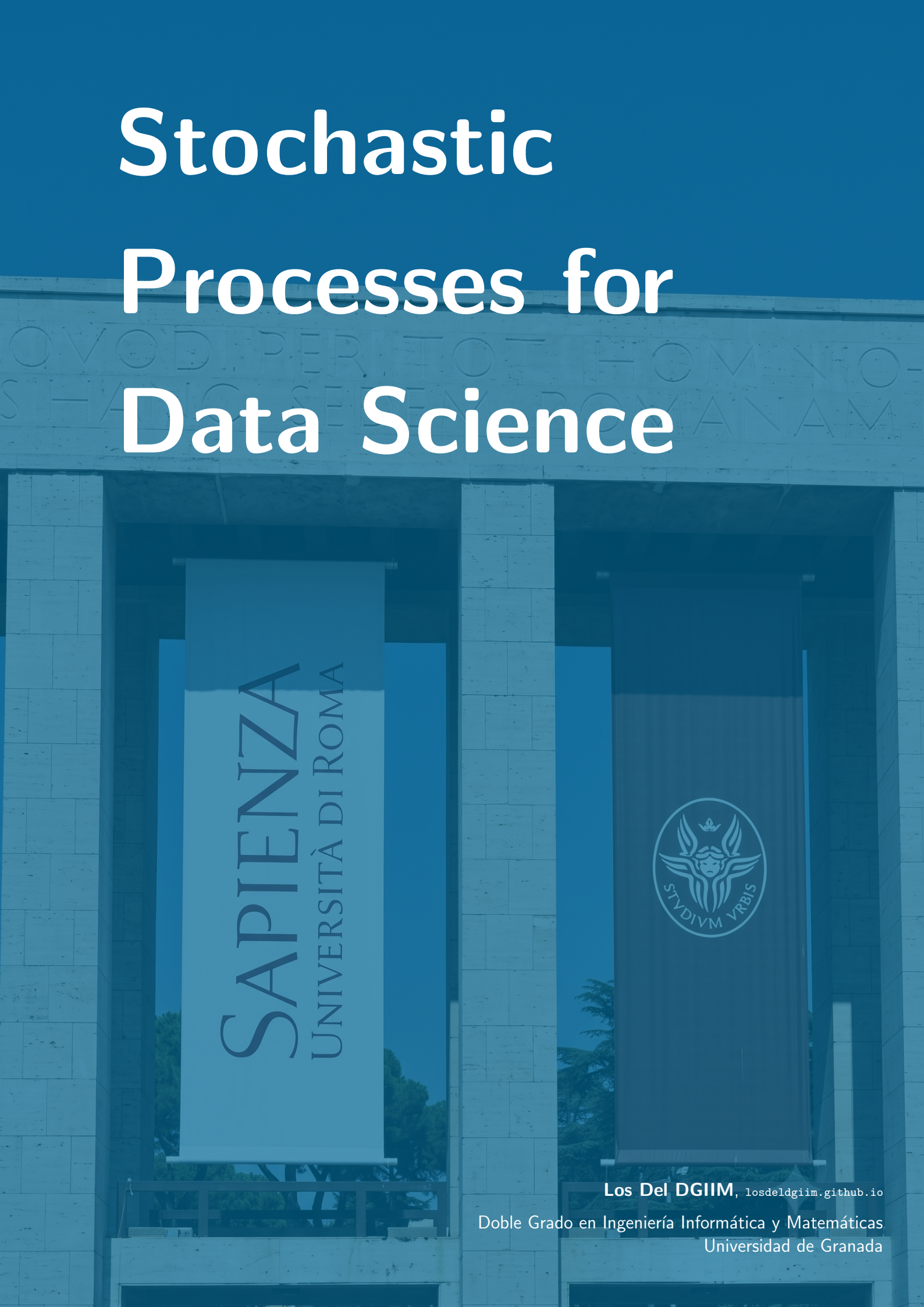
Stochastic Processes for Data Science



Los Del DGIIM, losdeldgiim.github.io

Doble Grado en Ingeniería Informática y Matemáticas
Universidad de Granada

Stochastic Processes for Data Science

Los Del DGIIM, losdeldgiim.github.io

Joaquín Avilés de la Fuente

Granada, 2025-2026

Índice general

1. Introduction

A stochastic process can be motivated by the fundamental problem of **forecasting** (prediction). Given a sequence of observations up to the current time m:

$$x_0, x_1, \ldots, x_m$$

Our objective is to determine the next state via an evolution function G characterized by a set of parameters θ:

$$x_{m+1} = G(x_0, \ldots, x_m; \theta)$$

Depending on our priori knowledge of the system, we can classify the forecasting problem into three distinct scenarios:

1. **Scenario 1:** We know the relevant variables and the exact time-evolution equations.
2. **Scenario 2:** We know the relevant variables, *but* we do not know the governing equations.
3. **Scenario 3:** We know neither the relevant variables nor the underlying equations.

Analysis of Scenario 1: Classical Determinism and Chaos

As an illustrative example, consider **Newtonian dynamics** governing the state of a particle, where positions x_m and velocities v_m are tracked over discrete time increments Δt:

$$\begin{cases} x_{m+1} = x_m + v_m \Delta t \\ v_{m+1} = v_m + a_m \Delta t \end{cases} \qquad \text{where } a_m = \frac{F}{m}$$

Let δ_0 denote the initial measurement error at $t = 0$. In chaotic systems characterized by a maximum positive **Lyapunov exponent** $\lambda_1 > 0$, the error at time t diverges exponentially:

$$\delta_t \simeq \delta_0 e^{\lambda_1 t}$$

If we establish Δ as the maximum acceptable error threshold ($\delta_T = \Delta$), the absolute **prediction horizon** T is strictly bounded by:

$$\Delta = \delta_0 e^{\lambda_1 T} \implies T(\delta_0, \Delta) \approx \frac{1}{\lambda_1} \log\left(\frac{\Delta}{\delta_0}\right)$$

Definición 1. The **prediction horizon** T es, sencillamente, el tiempo exacto en el que el error acumulado alcanza tu límite de tolerancia (Δ). Es decir, es el momento en el que el error se vuelve tan grande que la predicción deja de ser úti

Implicación para la Ciencia de Datos

La presencia del operador logarítmico revela la "trampa del caos": incluso si invertimos recursos masivos en mejorar los instrumentos de medición reduciendo el error inicial δ_0 por un factor de 1000, el horizonte temporal de predicción T solo aumentará de forma lineal y marginal. Por lo tanto, el determinismo físico no garantiza la predictibilidad.

1.1. El Teorema de Recurrencia de Poincaré y Ergodicidad

Frente a la pérdida de predictibilidad individual, cabe preguntarse si un sistema determinista, al menos macroscópicamente, regresa a sus configuraciones pasadas. Esto nos conduce al análisis del espacio de fases acotado.

Teorema 1.1.1 (**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:*

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

Para estudiar este fenómeno, definimos formalmente el **Tiempo de Retorno** o primer tiempo de entrada $\tau_\sigma(\underline{x}_0)$ como el ínfimo de los pasos temporales en los que el sistema vuelve a la región de interés:

$$\tau_\sigma(\underline{x}_0) = \text{ínf}\left\{k \geqslant 1 \;\middle|\; \underline{x}_0 \in \sigma \;\cap\; \underline{x}_k \in \sigma\right\} \tag{1.2}$$

Dado que evaluar el tiempo de retorno exacto para una sola trayectoria es impracticable, recurrimos a su valor esperado o promedio estadístico $\mathbb{E}[\tau_\sigma(\underline{x}_0)]$. Introduciendo una medida de probabilidad invariante μ en el espacio de fases (asociada al hipervolumen de la región), la esperanza se modela como:

$$\mathbb{E}[\tau_\sigma(\underline{x}_0)] = \frac{1}{\mu(\sigma)} \int_{\underline{x}\in\sigma} \tau_\sigma(x)\, d\mu(x) \tag{1.3}$$

1.1.1. La Hipótesis de Ergodicidad y el Lema de Kac

Definición 1.1.2. La hipótesis de **ergodicidad** establece que, para un sistema dinámico ergódico, el promedio temporal de una trayectoria individual es equivalente al promedio sobre todo el conjunto (*ensemble*) del espacio de fases. En otras palabras, si se espera un tiempo suficientemente largo ($t \to \infty$), ambos promedios convergen al mismo valor.

Ejemplo. Para entender a que nos referimos con la **hipótesis de ergodicidad**, imagínate que quieres estudiar el comportamiento de los clientes en un supermercado. Tienes dos formas de calcular un promedio:

- **Promedio de Conjunto (Espacio de fases):** Congelas el tiempo en un instante exacto, sacas una "fotografía" de todo el supermercado y calculas el promedio usando a todos los clientes a la vez en ese momento.
- **Promedio Temporal:** eliges a un único cliente y lo sigues con una cámara durante 10 años, registrando su comportamiento a lo largo del tiempo para luego hacer el promedio.

La hipótesis de ergodicidad afirma que, si el sistema es ergódico y esperas el tiempo suficiente ($t \to \infty$), ambos promedios darán exactamente el mismo resultado.

Si asumimos que el sistema es ergódico, podemos enunciar el **Lema de Kac**, que establece una relación fundamental entre el tiempo de retorno y la medida del conjunto de estados admisibles.

Lema 1.1.2 (**Lema de Kac**). *Para cualquier sistema dinámico ergódico que preserva la medida μ, la integral del tiempo de retorno sobre el conjunto de estados admisibles está estrictamente normalizada a la unidad:*

$$\int_{\underline{x}\in\sigma} \tau_\sigma(x)\, d\mu(x) = 1 \tag{1.4}$$

Sustituyendo el Lema de Kac en la definición de la esperanza matemática, se llega a una de las conclusiones más elegantes de la física estadística:

$$\mathbb{E}[\tau_\sigma(x)] = \frac{1}{\mu(\sigma)} \tag{1.5}$$

O lo que es lo mismo, el tiempo promedio de recurrencia a una región es exactamente el **inverso del hipervolumen** (medida) de dicho conjunto.

1.2. La Paradoja de los Sistemas de Alta Dimensión

Evaluemos cuantitativamente la magnitud física de este resultado. Supongamos que la región objetivo σ tiene una dimensión lineal de tolerancia ε (un error permitido) dentro de un dominio espacial acotado por una longitud máxima de excursión L. Si el sistema posee N grados de libertad independientes, el volumen relativo o medida de la región se escala exponencialmente con la dimensión:

$$\mu(\sigma) \propto \left(\frac{\varepsilon}{L}\right)^N \quad \text{donde } \frac{L}{\varepsilon} \gg 1 \tag{1.6}$$

Por consiguiente, aplicando el resultado de Kac, el tiempo esperado de retorno es:

$$\mathbb{E}[\tau_\sigma(x)] \propto \left(\frac{L}{\varepsilon}\right)^N \tag{1.7}$$

En un sistema físico macroscópico real (por ejemplo, las partículas de aire en una habitación o variables en problemas complejos de datos), los grados de libertad escalan con el **Número de Avogadro** ($N \sim 10^{23}$). Al elevar una base mayor que uno a una potencia de 10^{23}, el valor numérico resultante es inconmensurable: **el tiempo que tardaría el sistema en regresar a su configuración inicial supera por órdenes de magnitud la edad estimada de nuestro universo**.

1.3. Conclusión: El Nacimiento de los Procesos Estocásticos

Este colapso analítico nos obliga a replantear las estrategias de modelado en Ciencia de Datos cuando salimos de las condiciones controladas del Caso 1:

- **Limitación del Método de Análogos (Caso 2):** Cuando conocemos las variables pero no las leyes dinámicas, una técnica intuitiva es buscar estados pasados idénticos o "análogos" para proyectar el futuro. Sin embargo, debido a la inmensidad del espacio de fases impulsada por los altos grados de libertad ($N \gg 1$), el método de análogos **fracasa rotundamente**, puesto que la probabilidad de encontrar dos configuraciones verdaderamente cercanas en un histórico de datos tiende a cero.
- **Incertidumbre Absoluta (Caso 3):** Cuando no hay leyes ni variables claras, el determinismo queda completamente invalidado.

El Teorema de Poincaré y el Caos demuestran de forma conjunta que la descripción microscópica detallada y exacta es una utopía matemática inaccesible en sistemas complejos.

Para poder realizar pronósticos robustos en finanzas, meteorología o comportamiento computacional, debemos abandonar las trayectorias rígidas y deterministas y adoptar un paradigma probabilístico. Es precisamente en esta frontera donde concluye la física clásica y **comienza el estudio formal de los Procesos Estocásticos**.

2. Stochastic Processes

Antes de empezar a trabajar con procesos estocásticos vamos a recordar los **axiomas de Kolmogoron**, 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
$$P(A) \geqslant 0 \quad \forall A \in \mathcal{F}$$
- **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
$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Tras constatar las limitaciones de los enfoques puramente deterministas para modelar sistemas complejos de alta dimensión, la Ciencia de Datos recurre al paradigma probabilístico formalizado a través de la teoría de la medida.

Definición 1. Un **proceso estocástico** es una colección de variables aleatorias indexadas en el tiempo, definidas sobre un mismo espacio de probabilidad analítico $(\Omega, \mathcal{F}, \mathbb{P})$. En este espacio:

- Ω representa el espacio muestral que contiene todos los eventos elementales posibles ω.
- $\mathcal{F}$ constituye la σ-álgebra de sucesos medibles.
- $\mathbb{P}$ denota la medida de probabilidad asignada a dichos sucesos.

Matemáticamente, un proceso estocástico $X(\omega, t)$ se define como una función con un dominio bivariado que mapea el producto cartesiano del espacio muestral y el conjunto de indexación temporal hacia un espacio de estados observable en la recta real $\mathbb{R}$ (o $\mathbb{R}^d$ si es multidimensional):

$$X(\omega, t) : \ \Omega \times \mathbb{R}_0^+ \longrightarrow \mathbb{R} \tag{2.1}$$

La comprensión profunda de este objeto matemático radica en su **naturaleza dual**, la cual se manifiesta dependiendo de qué variable se mantenga constante: el elemento aleatorio ω o el tiempo t.

2.0.1. Realización o Trayectoria ($\bar{\omega}$ fijo, t variable)

Si fijamos un evento elemental específico del estado de la naturaleza $\bar{\omega} \in \Omega$, el componente aleatorio se disipa por completo. La función resultante

$$X(\bar{\omega}, t)$$

es una función puramente determinista del tiempo denominada **realización**, **trayectoria** o *sample path* del proceso estocástico. Esta curva representa la evolución histórica concreta que ha seguido el fenómeno bajo estudio.

2.0.2. Variable Aleatoria ($t = \bar{t}$ fijo, ω variable)

Si fijamos un instante de tiempo preciso $\bar{t} \in \mathbb{R}_0^+$ y permitimos que ω varíe a lo largo de todo el espacio muestral, el objeto matemático colapsa en una **variable aleatoria** tradicional[cite: 54, 56]. Para que esta variable sea matemáticamente válida, la función debe satisfacer la condición de **medibilidad** respecto a la σ-álgebra $\mathcal{F}$. Esto significa que la anti-imagen de cualquier conjunto de Borel $B \in \mathbb{R}$ debe ser un suceso integrable:

$$X^{-1}(B, \bar{t}) = \{\omega \in \Omega \mid X(\omega, \bar{t}) \in B\} \in \mathcal{F} \tag{2.2}$$

Esta condición analítica asegura que siempre seremos capaces de asignar una medida de probabilidad válida a la ocurrencia de dicho estado en el tiempo.

Observación. Si el conjunto de indexación del proceso estocástico no representa el tiempo sino coordenadas espaciales continuas, el objeto se denomina **Campo Aleatorio** (*Random Field*). Asimismo, si el índice abarca tanto coordenadas del espacio como del tiempo, se define formalmente como un campo aleatorio variable en el tiempo.

2.1. Procesos de Tiempo Discreto y Espacio de Estados Finito

Una vez estructurado el marco teórico continuo, resulta indispensable estudiar la variante discreta y finita, la cual es la base de las aplicaciones computacionales y los modelos algorítmicos. Consideremos un proceso donde el espacio de estados es un **conjunto finito** de tamaño g[cite: 57, 58]:

$$X_m = x_m \quad \text{donde } x_m \in S = \{1, 2, \ldots, g\} \tag{2.3}$$

Al muestrear el proceso en una secuencia ordenada de instantes temporales discretos, obtenemos la sucesión de variables aleatorias $X_1, X_2, \ldots, X_m$.

Para caracterizar estadísticamente y de forma completa el comportamiento conjunto de estas variables, recurrimos a las llamadas **Distribuciones Finito-Dimensionales** (FDD).

2.1.1. Definición de la Distribución Conjunta

La distribución finito-dimensional para m puntos en el tiempo se define como la probabilidad de la intersección lógica ($\cap$) de que cada variable aleatoria tome un valor específico del espacio de estados simultáneamente:

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_m = x_m) = \mathbb{P}\big(\{X_1 = x_1\} \cap \{X_2 = x_2\} \cap \cdots \cap \{X_m = x_m\}\big) \tag{2.4}$$

Para que una familia de distribuciones finito-dimensionales defina un proceso estocástico consistente y bien estructurado, debe satisfacer obligatoriamente dos propiedades axiomáticas esenciales:

Proposición 2.1.1 (**Simetría ante Permutaciones**). *La distribución conjunta del vector de variables aleatorias debe ser estrictamente invariante respecto al orden en el que se listen o registren sus componentes dentro del operador de probabilidad[cite: 62, 66]. Sea $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_m)$ una permutación arbitraria de los índices de tiempo, la igualdad funcional debe mantenerse inalterada:*

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

Proposición 2.1.2 (**Consistencia de Marginalización**). *Si disponemos de la caracterización probabilística conjunta para un sistema de m variables, debemos ser capaces de colapsar la dimensión del problema y recuperar con total fidelidad la distribución de un subconjunto menor de variables (por ejemplo, $m-1$ puntos)[cite: 63, 66]. Analíticamente, esto se logra aplicando el teorema de la probabilidad total, eliminando la variable no deseada (digamos, X_2) mediante la suma sobre todo su soporte finito admisible:*

$$\sum_{x_2 \in \{1,2,\ldots,g\}} \mathbb{P}\big(\{X_1 = x_1\} \cap \{X_2 = x_2\} \cap \cdots \cap \{X_m = x_m\}\big) = \mathbb{P}\big(\{X_1 = x_1\} \cap \cdots \cap \{X_m = x_m\}\big)$$

Estas dos propiedades constituyen las hipótesis del **Teorema de Consistencia de Kolmogorov**, el cual asegura que a partir de una familia consistente de distribuciones finito-dimensionales que cumplan con la simetría y la marginalización, queda garantizada la existencia única de un proceso estocástico real en el espacio $(\Omega, \mathcal{F}, \mathbb{P})$.

2.2. Construcción Formal del Espacio Muestral Colectivo

Consideremos un proceso estocástico compuesto por tres variables aleatorias binarias o dicotómicas X_1, X_2, X_3 que modelan lanzamientos sucesivos de monedas (caras H y cruces T) .

Para una única variable aleatoria elemental X_i, el espacio muestral es $\Omega = \{H, T\}$ y su σ-álgebra correspondiente es el conjunto de las partes $\mathcal{F} = \{\emptyset, \{H\}, \{T\}, \{H, T\}\}$. Definimos la medida de probabilidad elemental mediante:

$$P(\{H\}) = p, \qquad P(\{T\}) = 1 - p \quad \text{con } p \in (0, 1) \tag{2.5}$$

La variable aleatoria se mapea numéricamente asignando $X_i(T) = 0$ y $X_i(H) = 1$.

Al extender el análisis a la combinación conjunta de las tres variables X_1, X_2, X_3, construimos el **espacio de probabilidad producto** $(\Omega_3, \mathcal{F}_3, \mathbb{P}_3)$, donde:

- El espacio muestral conjunto es el producto cartesiano de los espacios individuales, generando un total de $2^3 = 8$ resultados elementales (eventos atómicos) ω :

$$\Omega_3 = \Omega \times \Omega \times \Omega = \{H, T\}^3 = \{(H, H, H), (H, H, T), \ldots, (T, T, T)\} \tag{2.6}$$

- La σ-álgebra global $\mathcal{F}_3$ se genera a partir de la unión y complemento de todas las combinaciones de estos sucesos elementales . Esto se denota formalmente como la σ-álgebra generada por los conjuntos unitarios :

$$\mathcal{F}_3 = \sigma\big(\{(H, H, H)\}, \{(H, H, T)\}, \ldots, \{(T, T, T)\}\big) \tag{2.7}$$

- La medida de probabilidad global $\mathbb{P}_3$ es el producto tensorial de las medidas individuales :

$$\mathbb{P}_3 = \bigotimes_{i=1}^{3} \mathbb{P} \tag{2.8}$$

2.3. Distribución Finito-Dimensional y la Propiedad de Intercambiabilidad

Si asumimos provisionalmente que los ensayos son estadísticamente independientes, la probabilidad conjunta de que ocurra una secuencia de eventos específicos $\{X_1 = x_1\} \cap \{X_2 = x_2\} \cap \{X_3 = x_3\}$ se reduce al producto de sus probabilidades marginales :

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = P(X_1 = x_1) \cdot P(X_2 = x_2) \cdot P(X_3 = x_3) \tag{2.9}$$

Dado que el espacio muestral conjunto contiene 8 casos posibles , para definir por completo esta distribución necesitaríamos en principio conocer $2^3 - 1 = 7$ números independientes (ya que la última probabilidad se deduce por complemento restando a la unidad: $1 - \sum P_i$) .

Sin embargo, al observar las asignaciones numéricas analíticas para cada una de las 8 secuencias (definiendo $q = 1 - p$) , emerge una propiedad fundamental:

$$\begin{aligned} P(0,0,0) &= (1-p)^3 = q^3 \\ P(0,0,1) = P(0,1,0) = P(1,0,0) &= (1-p)^2 p = q^2 p \\ P(0,1,1) = P(1,1,0) = P(1,0,1) &= (1-p) p^2 = q p^2 \\ P(1,1,1) &= p^3 \end{aligned} \tag{2.10}$$

El Concepto de Distribución Intercambiable (Exchangeable)

Nota cómo las secuencias que contienen **el mismo número de éxitos (1) y de fracasos (0) tienen exactamente la misma probabilidad**, sin importar el orden temporal en el que ocurrieron. Por ejemplo, $P(0, 0, 1) = P(0, 1, 0) = P(1, 0, 0)$. Esta es la definición de una **Distribución Intercambiable** .

Definición 2.3.2. A **exchangeable distribution** es aquella en la que cualquier permutación de una secuencia de variables aleatorias tiene exactamente la misma probabilidad. En otras palabras, las secuencias que contienen el mismo número de éxitos y fracasos comparten idéntica probabilidad, sin importar el orden temporal en el que ocurrieron.

Bajo esta condición, la suma de todas las intensidades probabilísticas se reduce notablemente mediante el desarrollo del binomio de Newton :

$$\sum P(\omega) = (p+q)^3 = p^3 + 3p^2q + 3pq^2 + q^3 = 1 \tag{2.11}$$

2.4. Demostración Analítica de la Consistencia de Marginalización

Para corroborar que nuestra estructura conjunta cumple con la propiedad axiomática de marginalización (consistencia de Kolmogorov), procedemos a colapsar la dimensión de la distribución eliminando la variable correspondiente al tercer lanzamiento, X_3, sumando sobre todo su soporte elemental $x_3 \in \{0, 1\}$:

$$P(X_1 = 0, X_2 = 0) = \sum_{x_3 \in \{0,1\}} P(X_1 = 0, X_2 = 0, X_3 = x_3) \tag{2.12}$$

Desglosando los términos correspondientes a la suma del soporte del tercer lanzamiento :

$$P(X_1 = 0, X_2 = 0) = P(0, 0, 1) + P(0, 0, 0) \tag{2.13}$$

Sustituyendo los valores funcionales de la distribución previamente calculados :

$$P(X_1 = 0, X_2 = 0) = (1-p)^2 p + (1-p)^3 \tag{2.14}$$

Factorizando por término común la expresión cuadrática del fracaso, $(1-p)^2$:

$$P(X_1 = 0, X_2 = 0) = (1-p)^2 \big[p + (1-p)\big] \tag{2.15}$$

Dado que el término entre corchetes se reduce a la unidad ($p + 1 - p = 1$), obtenemos el resultado final marginalizado :

$$P(X_1 = 0, X_2 = 0) = (1-p)^2 \cdot [1] = (1-p)^2 \tag{2.16}$$

Conclusión de la Demostración

El resultado corrobora perfectamente que la probabilidad conjunta de dos dimensiones coincide de manera idéntica con el producto de sus probabilidades marginales independientes: $P(X_1 = 0) \cdot P(X_2 = 0) = (1-p) \cdot (1-p) = (1-p)^2$. La distribución finito-dimensional es consistente .

2.5. Modelos de Urnas e Intercambiabilidad Sin Independencia

Sea una urna que contiene inicialmente un total de N bolas, divididas en B bolas negras (black) y W bolas blancas (white), de modo que $N = B + W$. Extraemos una muestra de tamaño n, donde se observan b bolas negras y w bolas blancas ($n = b + w$). Asignamos una variable aleatoria binaria X_i para cada extracción i, de tal manera que:

$$X_i = \begin{cases} 1 & \text{si se extrae una bola Negra } (B) \\ 0 & \text{si se extrae una bola Blanca } (W) \end{cases} \tag{2.17}$$

2.5.1. Muestreo Sin Reemplazo y la Distribución Hipergeométrica

En un esquema de extracción **sin reemplazo**, las probabilidades de cada extracción dependen de la historia pasada, eliminando la independencia estadística.

Análisis de Trayectorias

Evaluemos la probabilidad conjunta de obtener la secuencia específica blanca, blanca, negra (W_1, W_2, B_3):

$$P(X_1 = 0, X_2 = 0, X_3 = 1) = \frac{W}{N} \cdot \frac{W-1}{N-1} \cdot \frac{B}{N-2} = \frac{4}{8} \cdot \frac{3}{7} \cdot \frac{4}{6} \tag{2.18}$$

Si alteramos el orden cronológico de la secuencia a blanca, negra, blanca (W_1, B_2, W_3), la probabilidad conjunta es:

$$P(X_1 = 0, X_2 = 1, X_3 = 0) = \frac{W}{N} \cdot \frac{B}{N-1} \cdot \frac{W-1}{N-2} = \frac{4}{8} \cdot \frac{4}{7} \cdot \frac{3}{6} \tag{2.19}$$

Propiedad de Intercambiabilidad

A pesar de la falta de independencia, ambas secuencias arrojan exactamente el mismo valor numérico porque sus numeradores y denominadores contienen los mismos factores en diferente orden. Para una secuencia genérica ordenada con w blancas primero y b negras después, la probabilidad se formaliza mediante el **factorial descendente** (denotado como $x_{(y)} = x(x-1)\cdots(x-y+1)$):

$$P(\underbrace{W_1, \ldots, W_w}_{w}, \underbrace{B_{w+1}, \ldots, B_n}_{b}) = \frac{W_{(w)} B_{(b)}}{N_{(n)}} = \frac{W(W-1)\cdots(W-w+1) \cdot B(B-1)\cdots(B-b+1)}{N(N-1)\cdots(N-n+1)} \tag{2.20}$$

Dado que cualquier permutación de esta secuencia tiene la misma probabilidad, para hallar la probabilidad de obtener w blancas y b negras en *cualquier* orden, multiplicamos por el coeficiente combinatorio $\binom{n}{w} = \frac{n!}{w!b!}$, resultando en la **Distribución Hipergeométrica**:

$$P(w, b) = \frac{n!}{w!b!} \cdot \frac{W_{(w)} B_{(b)}}{N_{(n)}} \tag{2.21}$$

2.5.2. Muestreo con Reforzamiento: El Modelo de Urnas de Pólya

En el muestreo **con reforzamiento**, cada vez que se extrae una bola, se observa su color, se devuelve a la urna y **se añade una bola extra de ese mismo color**. Este mecanismo introduce una dependencia positiva (correlación).

Análisis de Trayectorias

Evaluemos la probabilidad de la secuencia blanca, blanca, negra (W_1, W_2, B_3):

$$P(W_1, W_2, B_3) = \frac{W}{N} \cdot \frac{W+1}{N+1} \cdot \frac{B}{N+2} = \frac{4}{8} \cdot \frac{5}{9} \cdot \frac{4}{10} \tag{2.22}$$

Generalización y Factorial Ascendente

Para una secuencia genérica con w blancas y b negras, el comportamiento se modela mediante el **factorial ascendente** (denotado como $x^{(y)} = x(x+1)\cdots(x+y-1)$):

$$P(W_1, \ldots, W_w, B_{w+1}, \ldots, B_n) = \frac{W^{(w)} B^{(b)}}{N^{(n)}} = \frac{W(W+1)\cdots(W+w-1) \cdot B(B+1)\cdots(B+b-1)}{N(N+1)\cdots(N+n-1)} \tag{2.23}$$

Como el orden de los factores en el numerador no altera el producto, el proceso sigue siendo **intercambiable**. Multiplicando por todas las combinaciones posibles de ordenar la secuencia, obtenemos la **Distribución de Pólya**:

$$P(w, b) = \frac{n!}{w!b!} \cdot \frac{W^{(w)} B^{(b)}}{N^{(n)}} \tag{2.24}$$

2.6. Modelos de Asignación Combinatoria

El análisis macroscópico de los procesos estocásticos en ciencia de datos requiere definir formalmente cómo se distribuyen e indexan un conjunto de datos u objetos dentro de un espacio de estados acotado. Formalmente, consideramos el problema de asignar m objetos en g categorías (o clases) diferenciables .

Para estructurar este problema, definimos el marco de trabajo mediante las siguientes condiciones axiomáticas:

1. Los m objetos son discernibles y pueden ser etiquetados unívocamente con valores enteros del conjunto $\{1, 2, \ldots, m\}$.
2. Las g categorías o clases son igualmente discernibles y se indexan mediante el conjunto $\{1, 2, \ldots, g\}$.

Para realizar un seguimiento probabilístico del sistema, se pueden adoptar diferentes niveles de descripción matemática según el grado de detalle requerido.

2.6.1. Individual Description

La **descripción individual** constituye el nivel microscópico más detallado de asignación. En este enfoque, registramos de forma exacta y ordenada a qué categoría específica pertenece cada uno de los objetos examinados .

Matemáticamente, definimos la variable de estado individual como:

$$X_i = j \quad \longrightarrow \quad \text{El objeto } i \text{ se encuentra asignado a la categoría } j \tag{2.25}$$

El número total de configuraciones microscópicas u ordenaciones posibles para situar m objetos en g cajas está determinado por las variaciones con repetición, lo cual equivale a multiplicar el número de opciones de cajas disponibles para cada objeto :

$$\text{Número total de combinaciones individuales} = g^m \tag{2.26}$$

Ejemplo Ilustrativo ($m = 3$ objetos, $g = 3$ categorías)

Consideremos el caso expuesto en la imagen donde disponemos de 3 objetos y 3 cajas . Una descripción individual específica sería determinar que el primer objeto está en la caja 2, el segundo en la caja 1, y el tercero en la caja 2 ($x_1 = 2, x_2 = 1, x_3 = 2$) . El total de mundos microscópicos posibles para este ejemplo es $3^3 = 27$ realizaciones .

2.6.2. Statistical Description

En muchas aplicaciones prácticas de la ciencia de datos, no nos interesa conocer la identidad exacta de qué objeto específico cayó dentro de una caja, sino únicamente el **conteo agregado** o la frecuencia final de ocupación de cada categoría . A este nivel intermedio se le denomina **descripción estadística** .

Formalmente, definimos la variable de conteo macroscópico mediante:

$$Y_i = m_i \quad \longrightarrow \quad \text{La categoría (caja) } i \text{ contiene exactamente } m_i \text{ objetos} \tag{2.27}$$

Sujeto a la restricción física de que la suma de todos los objetos en las categorías debe igualar al total de la muestra m :

$$\sum_{i=1}^{g} Y_i = m \tag{2.28}$$

El número total de descripciones estadísticas posibles corresponde al cálculo combinatorio de **combinaciones con repetición** (la clásica técnica de "barras y estrellas"), modelada mediante el coeficiente combinatorio :

$$\text{Total de descripciones estadísticas} = \binom{m+g-1}{m} = \binom{m+g-1}{g-1} = \frac{(m+g-1)!}{m!(g-1)!} \tag{2.29}$$

Para nuestro ejemplo base ($m = 3, g = 3$), el número de estados estadísticos es :

$$\binom{3+3-1}{3} = \binom{5}{3} = 10 \text{ configuraciones macroscópicas}$$

Estas 10 configuraciones equivalen a los vectores de ocupación lineal mostrados en los apuntes : $(3,0,0)$, $(0,3,0)$, $(0,0,3)$, $(1,2,0)$, $(1,0,2)$, $(0,1,2)$, $(2,1,0)$, $(2,0,1)$, $(0,2,1)$ y $(1,1,1)$.

2.6.3. Frequencies of frequencies descriptors (f.o.f.)

En el estudio macroscópico de asignación de m objetos en g categorías, el nivel máximo de compresión abstracta se alcanza mediante las **descripciones de frecuencias**. En este enfoque, se pierde por completo tanto la identidad de los objetos individuales como la identidad de las categorías particulares.

Definición 2.6.3. Sea Z_i la variable aleatoria que contabiliza la estructura global del sistema. Se define como:

$$Z_i = j \quad \longrightarrow \quad \text{Existen exactamente } j \text{ categorías que contienen } i \text{ objetos} \tag{2.30}$$

Ecuaciones Fundamentales de Restricción

Cualquier vector válido de frecuencias de frecuencias $\underline{Z} = (Z_0, Z_1, \ldots, Z_m)$ debe satisfacer de manera obligatoria las siguientes dos leyes de conservación algebraica:

Proposición 2.6.3 (**Conservación de Categorías**). *La suma de todas las componentes del vector debe igualar al número total de categorías g del sistema:*

$$\sum_{i=0}^{m} Z_i = g \tag{2.31}$$

Proposición 2.6.4 (**Conservación de Objetos**). *La suma ponderada de las componentes por su respectivo número de objetos debe igualar al tamaño total de la muestra m:*

$$\sum_{i=0}^{m} i \cdot Z_i = m \tag{2.32}$$

—

Ejercicio 2.6.1. ¿Cuál es el número de descripciones estadísticas (vectores de conteo de ocupación $\underline{Y}$) correspondientes a un vector de frecuencias de frecuencias dado $\underline{Z} = \underline{z}$?

Dado que una descripción f.o.f. representa la estructura abstracta de ocupación sin importar qué etiqueta tenga cada caja, el número de formas distintas en que las categorías pueden ordenarse para cumplir dicha distribución se calcula mediante el coeficiente multinomial de permutación de categorías:

$$\text{Número de descripciones estadísticas} = \frac{g!}{Z_0! \cdot Z_1! \cdot Z_2! \cdot \cdots \cdot Z_m!} \tag{2.33}$$

Ejemplo. Para un estado donde una caja está vacía, otra tiene un objeto y otra tiene dos objetos, el vector f.o.f. es $\underline{z} = (1,1,1,0)$. El número de descripciones estadísticas asociadas es:

$$\text{Descripciones Estadísticas} = \frac{3!}{1! \cdot 1! \cdot 1! \cdot 0!} = 6$$

Correspondientes a las permutaciones simétricas de ocupación del espacio de estados: $(1,2,0)$, $(2,1,0)$, $(1,0,2)$, $(2,0,1)$, $(0,1,2)$ y $(0,2,1)$.

2.7. Probabilidades Predictivas

Definición 2.7.4. La **probabilidad predictiva** es la probabilidad condicional de que un proceso estocástico tome un estado específico en el futuro ($t = m$), dado que se conoce el registro histórico completo de sus realizaciones pasadas.

Descomposición por la Regla de la Cadena

Utilizando la definición formal de probabilidad condicional ($P(A \cap B) = P(B \mid A)P(A)$), cualquier distribución conjunto o trayectoria finito-dimensional se puede descomponer recursivamente como el producto de una probabilidad inicial y una secuencia de probabilidades predictivas:

$$P(X_1 = x_1, \ldots, X_m = x_m) = P(X_1 = x_1) \prod_{k=2}^{m} P(X_k = x_k \mid X_1 = x_1, \ldots, X_{k-1} = x_{k-1})$$

Para un sistema de tres variables, la expansión exacta se escribe como:

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = P(X_3 = x_3 \mid X_2 = x_2, X_1 = x_1) \cdot P(X_2 = x_2 \mid X_1 = x_1) \cdot P(X_1 = x_1)$$

Ejemplo. Evaluación cronológica paso a paso de la trayectoria blanca, blanca, negra (W, W, B) en una urna con composición inicial $N = B + W$:

1. **Estado Inicial ($t = 1$):** Probabilidad marginal de extraer la primera blanca:
$$P(X_1 = W) = \frac{W}{N} \tag{2.34}$$
2. **Primera Predicción ($t = 2$):** Condicionado a la primera blanca, se añade una bola extra de ese color ($W + 1$ blancas, $N + 1$ totales):
$$P(X_2 = W \mid X_1 = W) = \frac{W + 1}{N + 1} \tag{2.35}$$
3. **Segunda Predicción ($t = 3$):** Condicionado a la historia previa (W, W), el total de la urna sube a $N + 2$. La cantidad de negras se mantiene intacta:
$$P(X_3 = B \mid X_1 = W, X_2 = W) = \frac{B}{N + 2} \tag{2.36}$$

Multiplicando los tres factores predictivos secuenciales obtenemos la probabilidad conjunta de la trayectoria:

$$P(W, W, B) = \frac{W}{N} \cdot \frac{W + 1}{N + 1} \cdot \frac{B}{N + 2}$$

Definición 2.7.5. Para cualquier paso temporal m, la probabilidad predictiva de observar un éxito (bola Negra) se calcula directamente sumando la memoria del proceso (éxitos acumulados $b = \sum x_i$) al estado inicial, usando la **fórmula de actualización generalizada**:

$$P(X_m = B \mid X_1 = x_1, \ldots, X_{m-1} = x_{m-1}) = \frac{B + b}{N + m - 1}$$

2.8. El Proceso de Pólya Generalizado

Consideremos una urna generalizada que contiene bolas distribuidas en g colores o categorías diferenciables . El estado inicial del sistema se parametriza mediante el número teórico de bolas asignado a cada componente :

$$\alpha_1 > 0,\ \alpha_2 > 0,\ \ldots,\ \alpha_g > 0$$

Aunque estos parámetros representan físicamente el conteo inicial de bolas (enteros), el formalismo matemático permite que $\alpha_i \in \mathbb{R}$ de forma continua sin alterar la validez del proceso . El volumen o masa total inicial de la urna se define como la suma indexada de sus componentes :

$$\alpha = \sum_{i=1}^{g} \alpha_i$$

El proceso evoluciona de forma discreta paso a paso. En cada extracción, se observa el color de la bola, se devuelve a la urna y se añade una bola extra de esa misma categoría (mecanismo de reforzamiento positivo) . Tras haber efectuado un número n de extracciones totales, denotamos como n_i al número de bolas acumuladas del tipo i, cumpliendo la condición de conservación de la muestra :

$$n = \sum_{i=1}^{g} n_i \quad \text{donde } n_i = \sum_{j=1}^{n} I_{\{X_j=i\}}$$

Aquí, $I_{\{X_j=i\}}$ representa la función indicadora que computa si la extracción en el tiempo j perteneció a la clase i .

Probabilidad Predictiva Generalizada

Utilizando la regla de la cadena, la probabilidad condicional de que la siguiente bola extraída en el instante $n+1$ sea de la categoría específica 1 depende de la historia acumulada a través de la siguiente probabilidad predictiva :

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

Esta asignación es matemáticamente consistente con los axiomas de Kolmogorov, ya que la suma de las probabilidades predictivas sobre todo el espacio de estados finito de las g categorías normaliza estrictamente a la unidad :

$$\sum_{i=1}^{g} P(X_{n+1} = i \mid X_1 = x_1, \ldots, X_n = x_n) = \frac{\sum_{i=1}^{g}(\alpha_i + n_i)}{\alpha + n} = \frac{\alpha + n}{\alpha + n} = 1$$

2.8.1. Distribución Finito-Dimensional y la Medida de Pólya Multivariante

Deseamos calcular la probabilidad conjunta de observar una trayectoria microscópica particular. Por conveniencia analítica y explotando la propiedad de **intercambiabilidad** del proceso, evaluamos la secuencia ordenada donde se agrupan primero los éxitos de la categoría 1, luego la categoría 2, hasta llegar a la componente g :

$$P(\underbrace{1, \ldots, 1}_{n_1}, \underbrace{2, \ldots, 2}_{n_2}, \ldots, \underbrace{g, \ldots, g}_{n_g})$$

Aplicando el desarrollo multiplicativo de las probabilidades predictivas secuenciales, la estructura de la fracción conjunta se desglosa como :

$$\left[\frac{\alpha_1}{\alpha} \cdots \frac{\alpha_1 + n_1 - 1}{\alpha + n_1 - 1}\right] \cdot \left[\frac{\alpha_2}{\alpha + n_1} \cdots \frac{\alpha_2 + n_2 - 1}{\alpha + n_1 + n_2 - 1}\right] \cdots \left[\frac{\alpha_g}{\alpha + n - n_g} \cdots \frac{\alpha_g + n_g - 1}{\alpha + n - 1}\right]$$

Asociando los factores de cada categoría mediante la definición del **factorial ascendente** ($x^{(y)} = x(x+1)\cdots(x+y-1)$), la probabilidad de esta trayectoria única se simplifica de forma compacta como :

$$P(\text{Trayectoria Ordenada}) = \frac{\alpha_1^{(n_1)} \alpha_2^{(n_2)} \ldots \alpha_g^{(n_g)}}{\alpha^{(n)}} = \frac{1}{\alpha^{(n)}} \prod_{i=1}^{g} \alpha_i^{(n_i)}$$

Función de Masa de Probabilidad de Ocupación Colectiva

Dado que el proceso es intercambiable, cualquier permutación de esta secuencia comparte exactamente la misma intensidad probabilística . El número de formas distintas en que podemos organizar n objetos donde el color i se repite n_i veces se rige por el coeficiente multinomial $\frac{n!}{n_1! n_2! \ldots n_g!}$.

Multiplicando la multiplicidad por la probabilidad de un solo camino, se define la probabilidad macroscópica del vector de conteo $\underline{Y}(n) = (n_1, \ldots, n_g)$, dando origen a la **Distribución de Pólya Multivariante** :

$$P(\underline{Y}(n) = \underline{n}) = \frac{n!}{\alpha^{(n)}} \prod_{i=1}^{g} \frac{\alpha_i^{(n_i)}}{n_i!}$$

2.8.2. Consistencia de la Distribución: El Caso Marginal Bivariado

Para corroborar la consistencia interna del modelo bajo el colapso de dimensiones (marginalización), consideremos el caso donde agrupamos las categorías en dos macro-clases ($g = 2$) . Sea la primera clase de interés $Y_1 = n_1$ y el residuo agrupado $Y_2 = n - n_1$. Redefiniendo los parámetros de la urna bajo la condición $\alpha = \alpha_1 + \alpha_2 \implies \alpha_2 = \alpha - \alpha_1$, la probabilidad conjunta colapsa en :

$$P(Y_1 = n_1) = \frac{n!}{\alpha^{(n)}} \frac{\alpha_1^{(n_1)}}{n_1!} \frac{(\alpha - \alpha_1)^{(n-n_1)}}{(n-n_1)!}$$

Reordenando los términos factoriales en la combinación binaria tradicional $\binom{n}{n_1} = \frac{n!}{n_1!(n-n_1)!}$, recuperamos la **Distribución Beta-Binomial discreta** (Pólya binaria) :

$$P(Y_1 = n_1) = \binom{n}{n_1} \frac{\alpha_1^{(n_1)} (\alpha - \alpha_1)^{(n-n_1)}}{\alpha^{(n)}}$$

Si tomamos el límite asintótico donde la masa inicial de la urna tiende a infinito ($\alpha \to \infty$) lo que ocurre es que el sacar una bola de una categoría concreta y añadir una más de esta, no altera las probabilidades (hay millones de millones de bola y no cambia apenas la probabilidad), por lo que se mantienen las proporciones constantes

$$\frac{\alpha_1}{\alpha} \to p_1 \quad \text{y} \quad \frac{\alpha_2}{\alpha} \to p_2$$

el proceso pierde el efecto de reforzamiento y converge de forma exacta a la **Distribución Multinomial** de ensayos independientes .

2.8.3. Conexión con Procesos No Markovianos: El Elephant Random Walk

El tramo final de la hoja conecta el Proceso de Pólya con una aplicación física y biológica de vanguardia en ciencia de datos: el ***Elephant Random Walk*** (Caminata aleatoria del elefante) . Este modelo simula una caminata con **memoria infinita**, donde un agente decide su siguiente paso basándose en todo su recorrido histórico pasado .

Si restringimos las variables de actualización predictiva a dos estados de movimiento dicotómicos $\{1, -1\}$ (donde 1 representa dar un paso a la derecha y -1 representa un paso a la izquierda), las ecuaciones predictivas se mapean directamente con la urna de Pólya binaria :

$$P(X_{n+1} = 1 \mid X_1 = x_1, \ldots, X_n = x_n) = \frac{\alpha_1 + n_1}{\alpha + n}$$
$$P(X_{n+1} = -1 \mid X_1 = x_1, \ldots, X_n = x_n) = \frac{\alpha_2 + n_2}{\alpha + n} = 1 - \frac{\alpha_1 + n_1}{\alpha + n}$$

Este mapeo analítico demuestra que el *Elephant Random Walk* es, en esencia, un proceso de Pólya encubierto, heredando su propiedad de **intercambiabilidad** y permitiendo resolver analíticamente la posición asintótica del agente a pesar de romper la propiedad de Markov por su dependencia histórica absoluta .

2.9. Caminatas Aleatorias y la Versión Finita de de Finetti

2.9.1. Comportamiento Asintótico del Elephant Random Walk

Consideremos una caminata aleatoria (*random walk*) definida por la suma acumulada de un proceso de variables aleatorias binarias no independientes (no i.i.d.) $X_i \in \{-1, +1\}$:

$$Z_n = \sum_{i=1}^{n} X_i$$

Heredando las probabilidades predictivas del proceso de Pólya parametrizado por los valores iniciales de la urna (α_1, α_2), el comportamiento asintótico (límite casi seguro, c.s.) del agente exhibe tres regímenes dinámicos drásticamente diferentes según la simetría original del sistema :

$$\begin{array}{lll} \alpha_1 > \alpha_2 & \implies & \lim_{n\to\infty} Z_n = +\infty \quad \text{c.s.} \\ \alpha_2 > \alpha_1 & \implies & \lim_{n\to\infty} Z_n = -\infty \quad \text{c.s.} \\ \alpha_1 = \alpha_2 & \implies & \lim_{n\to\infty} |Z_n| = +\infty \quad \text{c.s.} \end{array}$$

Nota sobre la Convergencia Clásica

Si el sistema es estrictamente simétrico y los parámetros iniciales son masivos ($\alpha_1 = \alpha_2 \gg 1$), el efecto de memoria del reforzamiento se diluye. El proceso colapsa asintóticamente en una caminata aleatoria simétrica clásica estándar (*unbiased random walk*) gobernada por variables X_i independientes e idénticamente distribuidas (i.i.d.) .

—

2.9.2. Propiedad Fundamental de la Distribución Hipergeométrica

Antes de abordar la representación general de de Finetti, es mandatorio aislar el comportamiento de un proceso intercambiable bajo el esquema de muestreo sin reemplazo (Distribución Hipergeométrica) .

Lema 2.9.5. *Dada una secuencia de variables binarias de longitud n bajo una distribución hipergeométrica puramente determinista condicionada a registrar exactamente un número fijo de h éxitos, la probabilidad conjunta de observar cualquier trayectoria microscópica específica es **uniforme** :*

$$P\left(X_1 = x_1, \ldots, X_n = x_n \;\middle|\; \sum_{i=1}^{n} X_i = h\right) = \frac{1}{\binom{n}{h}}$$

Observación. Lo que realmente se esta representando en este lema, es que dado por ejemplo un número de lanzamientos de un moneda n y un número de éxitos h (que salga cara por ejemplo), la probabilidad de que salga una secuencia específica de caras y cruces es uniforme, es decir, todas las secuencias posibles que tengan exactamente h caras y $n - h$ cruces tienen la misma probabilidad de ocurrir. Esto refleja la simetría del problema y la naturaleza combinatoria de las posibles secuencias.

Observación. $\binom{n}{h}$ representa el número de formas distintas de elegir h posiciones para los éxitos (caras) entre n lanzamientos, y cada una de estas configuraciones es igualmente probable bajo la condición de tener exactamente h éxitos.

Verificación Empírica mediante la Función de Masa

La función de masa de probabilidad hipergeométrica tradicional que computa la obtención de h éxitos en n extracciones a partir de una urna con N elementos totales y H éxitos iniciales cumple la simetría algebraica de conmutación de índices :

$$P(Y_n = h) = \frac{\binom{H}{h}\binom{N-H}{n-h}}{\binom{N}{n}} = \frac{\binom{n}{h}\binom{N-n}{H-h}}{\binom{N}{H}}$$

Si evaluamos casos frontera cerrados para una muestra fija de $n = 3$, se demuestra la naturaleza uniforme del soporte condicional :

- **Para** $h = 0$**:** $P(Y_3 = 0 \mid H = 0) = 1 \implies P(0,0,0) = 1 = \frac{1}{\binom{3}{0}}$
- **Para** $h = 1$**:** $P(Y_3 = 1 \mid H = 1) = 1 \implies P(0,0,1) = P(0,1,0) = P(1,0,0) = \frac{1}{3} = \frac{1}{\binom{3}{1}}$
- **Para** $h = 2$**:** $P(Y_3 = 2 \mid H = 2) = 1 \implies P(1,1,0) = P(1,0,1) = P(0,1,1) = \frac{1}{3} = \frac{1}{\binom{3}{2}}$

- **Para $h = 3$:** $P(Y_3 = 3 \mid H = 3) = 1 \implies P(1,1,1) = 1 = \frac{1}{\binom{3}{3}}$

De este modo, queda demostrado que la probabilidad condicional de una trayectoria bajo el marco hipergeométrico depende de forma exclusiva de la masa del hipervolumen combinatorio :

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \mid H = h) = \begin{cases} \frac{1}{\binom{n}{h}} & \text{si } \sum_{i=1}^{n} X_i = h \\ 0 & \text{en otro caso} \end{cases}$$

Intuición Física Asintótica: En una urna que posea un número masivo de elementos ($N \to \infty$), las fronteras estadísticas entre el muestreo con o sin reemplazo se vuelven idénticas durante un horizonte temporal finito, ya que la remoción de unos pocos elementos apenas perturba las intensidades de probabilidad marginales.

2.9.3. La Versión Finita del Teorema de de Finetti

El resultado hipergeométrico anterior es la pieza clave para demostrar el teorema fundamental de representación en espacios finitos .

Teorema 2.9.6 (**Teorema de de Finetti - Versión Finita**). *Sea P una distribución de probabilidad estrictamente intercambiable sobre una secuencia de variables aleatorias binarias de longitud finita n. Entonces, existe una única distribución de probabilidad discreta a priori $\mu(h)$ definida sobre el soporte $\{0, 1, \ldots, n\}$ tal que la probabilidad de cualquier trayectoria se puede expresar como una mixtura ponderada de densidades hipergeométricas uniformes :*

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

donde $\mu(h)$ es la distribución a priori, que indica el peso que le asignas a cada escenario. Contiene la "verdadera naturaleza física" del experimento particular (si la moneda está trucada, la composición inicial de la urna, etc.).

Demostración. Definamos el suceso intersección que caracteriza la trayectoria microscópica completa como $A = \bigcap_{i=1}^{n} \{X_i = x_i\}$. Aplicando el teorema de la probabilidad total, podemos expandir la probabilidad de este evento condicionándolo sobre el número total de éxitos reales h registrados por el proceso :

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

Por la hipótesis fundamental de **intercambiabilidad**, la distribución P está obligada a asignar una probabilidad idéntica a cada una de las $\binom{n}{h}$ secuencias distintas que comparten exactamente el mismo número de éxitos h, sin importar su orden cronológico . Por consiguiente, la probabilidad condicional de observar la trayectoria exacta A coincide de forma unívoca con la densidad de una distribución hipergeométrica pura :

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

Sustituyendo esta equivalencia geométrica en la sumatoria de la probabilidad total, y definiendo la medida de ponderación discreta como la probabilidad marginal del conteo de éxitos, $\mu(h) = P\left(\sum X_i = h\right)$, el teorema queda plenamente demostrado :

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

□

Para comprender la mecánica operacional de la versión finita del Teorema de de Finetti, analizamos el comportamiento de un proceso estocástico en tiempo discreto y espacio de estados finito, para ellos veámos el siguiente ejemplo.

Ejemplo. Consideremos un experimento compuesto por $m = 3$ variables aleatorias binarias, donde deseamos evaluar analíticamente la probabilidad de observar la trayectoria microscópica específica:

$$(x_1, x_2, x_3) = (0, 1, 0)$$

Si asumimos provisionalmente que los datos provienen de ensayos independientes e idénticamente distribuidos (i.i.d.) de Bernoulli con probabilidad de éxito p, la probabilidad conjunta real de dicha secuencia se calcula mediante el producto clásico de sus intensidades marginales:

$$P(X_1 = 0, X_2 = 1, X_3 = 0) = (1 - p) \cdot p \cdot (1 - p) = (1 - p)^2 p$$

El objetivo de este ejemplo es proyectar esta trayectoria sobre la sumatoria de mixtura de de Finetti para deducir la estructura analítica de la distribución a priori $\mu(h)$:

$$P(0, 1, 0) = \sum_{h=0}^{3} P_H \left(X_1 = 0, X_2 = 1, X_3 = 0 \;\middle|\; \sum_{i=1}^{3} X_i = h \right) \cdot \mu(h)$$

Evaluamos de forma indexada el comportamiento del término de probabilidad condicional hipergeométrica P_H para cada valor posible del número total de éxitos $h \in \{0, 1, 2, 3\}$:

1. **Para $h = 0$:** El soporte condicional contiene de forma exclusiva a la secuencia $(0, 0, 0)$. Dado que nuestra trayectoria objetivo es $(0, 1, 0)$, la probabilidad de coincidencia es nula:

$$P_H \left((0, 1, 0) \;\middle|\; \sum X_i = 0 \right) = 0$$

2. **Para $h = 1$:** El espacio condicional contiene las $\binom{3}{1} = 3$ permutaciones uniformes con un único éxito: $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$. Al ser un proceso intercambiable, la masa se distribuye equitativamente, otorgando a cada camino un peso de $1/3$. Como nuestra secuencia pertenece a este conjunto, el término sobrevive:

$$P_H \left((0, 1, 0) \;\middle|\; \sum X_i = 1 \right) = \frac{1}{3}$$

3. **Para $h = 2$:** El soporte está restringido a las secuencias con dos éxitos: $\{(1, 1, 0), (0, 1, 1), (1, 0, 1)\}$. Al no corresponder con el conteo de nuestra secuencia, el término colapsa a cero:

$$P_H \left((0, 1, 0) \;\middle|\; \sum X_i = 2 \right) = 0$$

4. **Para $h = 3$:** El espacio condicional se reduce unívocamente al evento atómico $(1, 1, 1)$, resultando en:

$$P_H \left((0, 1, 0) \;\middle|\; \sum X_i = 3 \right) = 0$$

Sustituyendo los coeficientes condicionales calculados dentro de la sumatoria general del teorema de representación, todos los sumandos se anulan con excepción del estado asociado a un único éxito ($h = 1$):

$$P(0, 1, 0) = 0 \cdot \mu(0) + \frac{1}{3} \cdot \mu(1) + 0 \cdot \mu(2) + 0 \cdot \mu(3) \implies P(0, 1, 0) = \frac{1}{3}\mu(1)$$

Igualando esta expresión con el valor real del proceso i.i.d. calculado al inicio, podemos despejar el valor de la componente $\mu(1)$:

$$(1 - p)^2 p = \frac{1}{3}\mu(1) \implies \mu(1) = 3(1 - p)^2 p$$

Si generalizamos de forma sistemática este mismo procedimiento algebraico para el resto de las trayectorias del espacio muestral, el sistema mapea unívocamente los cuatro pesos del proceso:

$$\mu(0) = (1-p)^3 = \binom{3}{0} p^0 (1-p)^3$$
$$\mu(1) = 3(1-p)^2 p = \binom{3}{1} p^1 (1-p)^2$$
$$\mu(2) = 3(1-p) p^2 = \binom{3}{2} p^2 (1-p)^1$$
$$\mu(3) = p^3 = \binom{3}{3} p^3 (1-p)^0$$

La función de ponderación discreta $\mu(h)$ recuperada de forma automática por el teorema coincide de manera exacta con la función de masa de la **Distribución Binomial**. El experimento verifica que la suma de todas las intensidades probabilísticas cumple la condición de normalización del binomio de Newton:

$$\sum_{h=0}^{3} \mu(h) = \mu(0) + \mu(1) + \mu(2) + \mu(3) = (1 - p + p)^3 = 1$$

Antes de ver el siguiente teorema hay que tener claro la definición de integral de Stieltjes, que es una generalización de la integral de Riemann. La integral de Stieltjes permite integrar funciones con respecto a funciones de distribución, lo que es útil en probabilidad y estadística.

Definición 2.9.6 (Integral de Riemann-Stieltjes). La **integral de Riemann-Stieltjes** es una generalización de la integral clásica donde el integrando $f(\theta)$ se evalúa respecto a una función ponderadora $Q(\theta)$ (llamada **integrador**), en lugar de avanzar uniformemente sobre el eje X. Su expresión formal es

$$\int_a^b f(\theta)\, dQ(\theta)$$

El comportamiento y el valor del diferencial $dQ(\theta)$ dependen estrictamente de la naturaleza de la función $Q(\theta)$ a lo largo del eje X

- Si $Q(\theta)$ es una función continua y diferenciable: el proceso se comporta como una integral clásica orientada al cálculo de áreas continuas, donde el diferencial se expande mediante la derivada ordinaria:
$$dQ(\theta) = Q'(\theta)\, d\theta$$
- Si $Q(\theta)$ es una función escalonada pura (**Caso de de Finetti**): la función solo cambia mediante saltos verticales discretos en ciertos nodos del eje X y permanece completamente constante (plana) en los intervalos intermedios. Esto altera el diferencial de la siguiente manera:
  - En los tramos planos ($Q(\theta)$ constante): no hay variación vertical, por lo que el diferencial se anula por completo:
  $$dQ(\theta) = 0 \implies \int_{\text{tramo}} f(\theta) \cdot 0 = 0$$
  - En el punto exacto de un escalón ($\theta = \theta_k$): la función experimenta un salto vertical abrupto. El diferencial "despierta" y absorbe exactamente la magnitud o altura de dicho salto (que en probabilidad equivale a la masa discreta P_k):
  $$dQ(\theta_k) = \Delta Q(\theta_k) = P(Y = y_k)$$

Teorema 2.9.7 (**Teorema de Finetti**). *Sea $\{X_i\}_{i \in \mathbb{N}^+}$ una sucesión infinita e intercambiable de variables aleatorias binarias ($X_i \in \{0,1\}$). Entonces, existe una única función de distribución acumulada $Q(\theta)$ definida sobre el intervalo $[0,1]$ tal que, para cualquier tamaño de muestra n, la distribución de probabilidad conjunta de la trayectoria viene dada por la mixtura integral:*

$$P(X_1 = x_1, \ldots, X_n = x_n) = \int_0^1 \prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i}\, dQ(\theta)$$

Demostración. Sea

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

la variable macroscópica que contabiliza el número de éxitos en una submuestra de longitud m. Por la propiedad de intercambiabilidad, la probabilidad de observar un conteo de éxitos exacto $Y_m = y_m$ es el producto del coeficiente combinatorio por la intensidad de una única trayectoria microscópica:

$$P(Y_m = y_m) = \binom{m}{y_m} P(X_1 = x_1, \dots, X_m = x_m)$$

Para proyectar este sistema hacia el infinito, introducimos una ventana de observación mayor de longitud N, tal que $N \geqslant m \geqslant y_m \geqslant 0$. Aplicando el teorema de la probabilidad total sobre el número total de éxitos intermedios $Y_N = y_N$, la distribución se expande como una mixtura condicionada a la distribución hipergeométrica:

$$P(Y_m = y_m) = \sum_{y_N} P_H(Y_m = y_m \mid Y_N = y_N) \cdot P(Y_N = y_N)$$

donde tenemos que

$$P_H(Y_m = y_m \mid Y_N = y_N) = \frac{\binom{Y_N}{y_m}\binom{N-Y_N}{m-y_m}}{\binom{N}{m}} = \frac{\dfrac{Y_{N_{(y_m)}}}{y_m!} \cdot \dfrac{(N-Y_N)_{(m-y_m)}}{(m-y_m)!}}{\dfrac{N_{(m)}}{m!}} =$$

$$= \binom{m}{y_m} \frac{Y_{N_{(y_m)}} \cdot (N-Y_N)_{(m-y_m)}}{N_{(m)}}$$

Sustituyendo la masa de la distribución hipergeométrica expresada mediante la notación de **factoriales descendentes** ($x_{(y)} = x(x-1)\cdots(x-y+1)$), extraemos el coeficiente binomial y estructuramos la sumatoria:

$$P(Y_m = y_m) = \binom{m}{y_m} \sum_{y_N} \frac{Y_{N_{(y_m)}} \cdot (N-Y_N)_{(m-y_m)}}{N_{(m)}} \cdot P(Y_N = y_N)$$

Para transformar esta suma discreta en un operador continuo, introducimos la función de distribución empírica escalonada $Q_N(\theta)$, cuyos saltos discretos ocurren en los nodos discretos

$$\theta = \frac{Y_N}{N}$$

con magnitudes de probabilidad iguales a $P(Y_N = y_N)$. Al realizar este cambio de variable, la sumatoria se convierte formalmente en una integral de Stieltjes sobre el soporte continuo $[0, 1]$:

$$P(Y_m = y_m) = \binom{m}{y_m} \int_0^1 \frac{(\theta N)_{(y_m)} \cdot ((1-\theta)N)_{(m-y_m)}}{N_{(m)}} \, dQ_N(\theta)$$

Estudiamos el comportamiento asintótico del integrando cuando el tamaño del bloque superior tiende a infinito ($N \to \infty$) mientras mantenemos fija la submuestra m. Sabiendo que para valores masivos $x \gg 1$ el factorial descendente se aproxima a la potencia tradicional ($x_{(m)} \approx x^m$), el cociente de combinaciones converge de forma determinista hacia el núcleo de una distribución binomial i.i.d.:

$$\lim_{N\to\infty} \frac{(\theta N)_{(y_m)} \cdot ((1-\theta)N)_{(m-y_m)}}{N_{(m)}} = \theta^{y_m}(1-\theta)^{m-y_m} = \prod_{i=1}^{m} \theta^{x_i}(1-\theta)^{1-x_i}$$

donde la última igualdad se obtiene al reescribir el exponente de cada factor como una función indicadora de la trayectoria microscópica $x_i \in \{0, 1\}$, es decir, si el experimento i-ésimo es un éxito ($x_i = 1$) o un fracaso ($x_i = 0$).

Para garantizar la validez matemática de intercambiar el operador de límite con el signo de la integral, recurrimos al **Teorema de Convergencia Dominada**, es decir, necesitamos demostrar que el integrando está acotado por una función integrable independiente de N. Dado que las fracciones hipergeométricas están acotadas superiormente por la unidad (cualquier probabilidad lo está), la integral de la función dominante converge de forma trivial:

$$\int_0^1 \frac{(\theta N)_{(y_m)} \cdot ((1-\theta)N)_{(m-y_m)}}{N_{(m)}}\, dQ_N(\theta) \leqslant \int_0^1 1\, dQ_N(\theta) = 1 < \infty$$

El cumplimiento de esta cota nos faculta para introducir el límite dentro de la integral:

$$\lim_{N\to\infty} P(Y_m = y_m) = \binom{m}{y_m} \int_0^1 \left[\lim_{N\to\infty} \frac{(\theta N)_{(y_m)} \cdot ((1-\theta)N)_{(m-y_m)}}{N_{(m)}} \right] dQ_N(\theta)$$

Finalmente, invocando el **Teorema de Selección de Helly** (Helly's Selection Theorem), se garantiza que la secuencia de medidas de probabilidad empíricas $Q_N(\theta)$ converge debilmente hacia una medida de probabilidad límite única $Q(\theta)$ en el intervalo compacto $[0, 1]$. Sustituyendo ambos límites, el teorema de representación queda plenamente demostrado:

$$P(Y_m = y_m) = \binom{m}{y_m} \int_0^1 \theta^{y_m}(1-\theta)^{m-y_m}\, dQ(\theta) \implies$$

$$\implies P(X_1 = x_1, \ldots, X_m = x_m) = \int_0^1 \prod_{i=1}^{m} \theta^{x_i}(1-\theta)^{1-x_i}\, dQ(\theta)$$

donde hemos usado que

$$P(X_1 = x_1, \ldots, X_m = x_m) = \frac{P(Y_m = y_m)}{\binom{m}{y_m}}$$

□

Observación. Si la medida de probabilidad Q admite una función de densidad $q(\theta) = \frac{dQ}{d\theta}$, la ecuación puede expresarse en términos del formalismo bayesiano clásico como:

$$P(X_1 = x_1, \ldots, X_n = x_n) = \int_0^1 \left[\prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} \right] \cdot q(\theta)\, d\theta$$

donde conceptualmente identificamos:

- **Distribución Conjunta Marginal / Posterior:** $P(X_1 = x_1, \ldots, X_n = x_n)$.
- **Verosimilitud de Ensayos Independientes (Likelihood):** $\prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$, que asume un comportamiento i.i.d. condicionado a un parámetro θ fijo.
- **Distribución a Priori (Prior):** $q(\theta)d\theta$, que dota al parámetro de naturaleza aleatoria y actúa como la medida de ponderación de la mixtura.

El parámetro aleatorio θ se define formalmente como la proporción límite del conteo de éxitos ($Y_n = \sum_{i=1}^{n} X_i$) cuando el tamaño muestral tiende a infinito, cuya existencia queda garantizada bajo convergencia **casi segura (almost surely)** por la Ley Fuertísima de los Grandes Números para procesos intercambiables:

$$\theta \overset{\text{def}}{=} \lim_{n\to\infty} \frac{Y_n}{n} \implies Q(\theta) = \lim_{n\to\infty} P\left(\frac{Y_n}{n} \leqslant \theta \right)$$

2.10. Teorema de Helly-Bray y Johnson

El Teorema de Helly-Bray (y en un espectro más amplio, el Lema de Portmanteau) constituye el soporte analítico definitivo para validar el paso al límite en la demostración de Finetti infinita. Su propósito fundamental es establecer la equivalencia entre la convergencia en distribución de

variables aleatorias y la convergencia de sus operadores de esperanza matemática sobre distintas clases de funciones de prueba.

Sea $\{X_n\}_{n\in\mathbb{N}^+}$ una secuencia de variables aleatorias y X la variable aleatoria límite. Las siguientes cuatro afirmaciones macroscópicas son matemáticamente equivalentes entre sí, sirviendo todas como definición operacional de la **convergencia débil**, denotada como

$$X_n \xrightarrow[n\to\infty]{d} X$$

a) Convergencia en Distribución clásica: Las funciones de distribución acumulada convergen en todos los puntos donde la función límite es continua:

$$F_n(x) \xrightarrow[n\to\infty]{} F(x) \quad \forall x \in \mathcal{C}(F)$$

b) Convergencia sobre Funciones Continuas con Soporte Compacto: Las esperanzas matemáticas convergen para toda función de prueba $h : \mathbb{R} \to \mathbb{R}$ que sea continua y difiera de cero estrictamente dentro de un conjunto cerrado y acotado (espacio compacto):

$$\mathbb{E}[h(X_n)] \xrightarrow[n\to\infty]{} \mathbb{E}[h(X)]$$

c) Convergencia sobre Funciones Continuas y Acotadas (Extensión de b): Las esperanzas convergen para toda función continua $h : \mathbb{R} \to \mathbb{R}$ cuya magnitud esté acotada globalmente ($|h(x)| \leqslant M < \infty$), garantizando que la masa de probabilidad no se escape en los extremos del soporte:

$$\int_{\mathbb{R}} h(x)\, dF_n(x) \xrightarrow[n\to\infty]{} \int_{\mathbb{R}} h(x)\, dF(x)$$

es la usada en la demostración de Finetti.

d) Convergencia sobre Funciones Medibles Acotadas casi seguro: Las esperanzas convergen para cualquier función medible y acotada h cuyo conjunto de discontinuidades tenga medida cero respecto a la ley de probabilidad del límite (puntos de continuidad con masa $\mathbb{P}(\mathcal{C}_h) = 1$):

$$\mathbb{E}[h(X_n)] \xrightarrow[n\to\infty]{} \mathbb{E}[h(X)]$$

2.10.1. Postulados y Teoremade Johnson

Dejando atrás el espacio de estados estrictamente binario (0-1), el formalismo de W.E. Johnson extiende el análisis estocástico predictivo hacia un entorno generalizado con g categorías o colores simultáneos. Para monitorizar la influencia de la historia sobre las transiciones consecutivas, se introduce el **Coeficiente de Heterorrelevancia** Q_j^i.

Definición 2.10.7 (Coeficiente de Heterorrelevancia). El **coeficiente de heterorrelevancia** Q_j^i es un operador estocástico que cuantifica la influencia relativa de la última categoría observada j sobre la probabilidad predictiva de extraer la categoría i en el siguiente paso, en comparación con la probabilidad marginal de extraer i sin considerar el efecto inmediato de j, y viene dada de la siguiente manera

$$Q_j^i(X_1 = x_1, \ldots, X_m = x_m) = \frac{P(X_{m+2} = i \mid X_1 = x_1, \ldots, X_m = x_m, X_{m+1} = j)}{P(X_{m+1} = i \mid X_1 = x_1, \ldots, X_m = x_m)}$$

Para parametrizar el proceso de forma consistente, se establecen tres hipótesis restrictivas sobre la naturaleza del coeficiente:

Hipótesis 1: Independencia de la Última Observación

El coeficiente de heterorrelevancia evaluado en el instante $m+1$ es invariante ante la última categoría registrada x_{m+1}, dependiendo exclusivamente de la trayectoria acumulada previa de longitud m:

$$Q_j^i(X_1 = x_1, \ldots, X_m = x_m, X_{m+1} = x_{m+1}) = Q_j^i(X_1 = x_1, \ldots, X_m = x_m)$$

Implicación Crítica: Esta condición fuerza de forma directa la **intercambiabilidad bivariada local** de la secuencia, obligando a que el orden cronológico de aparición de dos categorías consecutivas (i y j) no altere su intensidad de probabilidad conjunta condicional:

$$P(X_{m+2} = i, X_{m+1} = j \mid X_1 = x_1, \ldots, X_m = x_m) = P(X_{m+2} = j, X_{m+1} = i \mid X_1 = x_1, \ldots, X_m = x_m)$$

Hipótesis 2: Homogeneidad Categórica del Coeficiente

El coeficiente se desliga de las etiquetas de las categorías específicas i y j, volviéndose un factor escalar Q_m que evoluciona únicamente en función del tiempo transcurrido (paso discreto m):

$$Q_m = Q_i^j(X_1 = x_1, \ldots, X_m = x_m) \quad \forall i \neq j$$

Hipótesis 3: Suficiencia del Conteo Marginal

La probabilidad predictiva de que la siguiente extracción pertenezca a la categoría j se denota de forma compacta como $p(j \mid m_j, m)$. Esta intensidad depende única y exclusivamente del número absoluto de observaciones acumuladas de esa misma categoría (m_j) y del tiempo total transcurrido (m), ignorando la ordenación interna o la distribución de las categorías restantes:

$$P(X_{m+1} = j \mid X_1 = x_1, \ldots, X_m = x_m) = p(j \mid m_j, m)$$

Bajo la convergencia de estas tres hipótesis, el Coeficiente de Heterorrelevancia temporal Q_m se reduce elegantemente al cociente de intensidades predictivas acopladas según el avance temporal del conteo de éxitos:

$$Q_m = \frac{p(k \mid m_k, m+1)}{p(k \mid m_k, m)}$$

Teorema 2.10.8 (Teorema de Johnson). *Bajo esta hipótesis, dada por sus tres postulados, el coeficiente de heterorrelevancia temporal Q_m se colapsa a la forma explícita*

$$Q_m = \frac{\alpha + m}{\alpha + m + 1}$$

lo que implica que la probabilidad predictiva de la categoría j se puede expresar como

$$p(j \mid m_j, m) = \frac{\alpha_j + m_j}{\alpha + m}$$

donde α y α_j se calculan durante la demostración.

Demostración. No se estudiará de cara al examen. □

Observación. Lo que demuestra el teorema es que tu predicción del futuro siempre será un promedio ponderado (combinación convexa) entre tu intuición inicial (α_j/α) y la evidencia de los datos reales (m_j/m).

Observación. Esta forma de $p(j \mid m_j, m)$ conduce al proceso de Pólya generalizado que ya hemos visto.

El resultado fundamental del teorema, expresado como:

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

indica que la probabilidad predictiva es una función lineal racional que amalgama dos fuerzas estadísticas distintas: la información histórica observada (los datos reales empíricos) y la información

teórica inicial (el modelo bayesiano a priori).

Podemos reescribir esta probabilidad predictiva exacta como una **combinación convexa** (un promedio ponderado) que ilustra de forma transparente la transición del aprendizaje inductivo:

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

Al desglosar esta equivalencia, identificamos la convergencia de dos componentes:

1. $\frac{m_j}{m}$: La **frecuencia relativa empírica** (la proporción real de éxitos observada en la muestra).
2. $\frac{\alpha_j}{\alpha}$: La **probabilidad a priori pura** (la creencia inicial del investigador antes de iniciar el experimento).

Esta formulación en combinación convexa justifica analíticamente cómo interactúa el flujo temporal del proceso estocástico con el conocimiento almacenado:

- **Régimen de Datos Escasos ($m \to 0$):** Cuando el tamaño de la muestra real es nulo o despreciable, el peso del primer término se desvanece, provocando que la predicción dependa exclusivamente de la estructura a priori del investigador:

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

- **Régimen de Datos Masivos ($m \to \infty$):** Cuando el horizonte temporal avanza hacia el infinito, el factor de ponderación empírica tiende a la unidad ($\frac{m}{m+\alpha} \to 1$), mientras que el peso de la opinión a priori se diluye por completo hacia cero. El sistema olvida su sesgo inicial y la predicción converge exactamente a la proporción real observada en los datos:

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

El Teorema de Johnson demuestra que **la regla predictiva de la Urna de Pólya Generalizada es la única solución matemática consistente** para un proceso predictivo multi-categoría que sea simétrico (intercambiable) y que reduzca la historia a conteos marginales suficientes.

3. Markov Chain

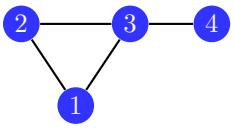
Hasta este punto, el núcleo del análisis estocástico se ha sustentado sobre procesos **intercambiables**, donde el orden cronológico de las observaciones carece de impacto en la asignación de la probabilidad conjunta:

$$P(X_1 = x_1, \ldots, X_m = x_m) = P(X_{\sigma_1} = x_{\sigma_1}, \ldots, X_{\sigma_m} = x_{\sigma_m})$$

Las **redes bayesianas** (*Bayesian Networks*), formalizadas intensamente por **Judea Pearl**, rompen este supuesto simétrico introduciendo una ordenación causal y asimétrica a través de estructuras de grafos.

Definición 1. Una **red bayesiana** parametriza las dependencias de un vector de variables aleatorias $\{X_1, X_2, \ldots, X_n\}$ mediante un Grafo $G = (V, E)$, donde los vértices V representan las variables (1 v.a. $\in V$) y las aristas E codifican las relaciones directas. Para garantizar la consistencia probabilística, el grafo debe ser un **DAG** (*Directed Acyclic Graph*): dirigido (relaciones causa-efecto unívocas) y acíclico (ausencia de bucles cerrados).

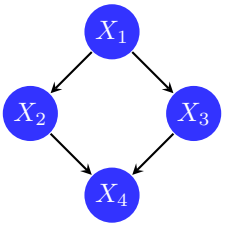
Grafo No Válido (Con Ciclos)



Matriz de Adyacencia (Grafo No Dirigido)

	1	2	3	4
1	0	1	1	0
2	1	0	1	0
3	1	1	0	1
4	0	0	1	0

DAG Válido (Red Bayesiana)



Asumiendo por simplicidad que las variables aleatorias discretas toman valores sobre un soporte finito, nos interesa calcular su distribución finito-dimensional conjunta. De primeras para un sistema de cuatro variables aleatorias $\{X_1, X_2, X_3, X_4\}$, la probabilidad conjunta se puede factorizar mediante la regla de la cadena

$$\begin{aligned} P(X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4) = {} & P(X_4 = x_4 \mid X_1 = x_1, X_2 = x_2, X_3 = x_3) \\ & \cdot P(X_3 = x_3 \mid X_1 = x_1, X_2 = x_2) \\ & \cdot P(X_2 = x_2 \mid X_1 = x_1) \\ & \cdot P(X_1 = x_1) \end{aligned}$$

Definición 2. La **propieda 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.

Sabiendo que las flechas de un DAG codifican una **relación de causalidad directa**. Bajo la *Propiedad de Márkov Local*, y sabiendo que los padres directos del nodo X_4 son exclusivamente X_2 y X_3, tenemos que

$$P(X_4 = x_4 \mid X_1 = x_1, X_2 = x_2, X_3 = x_3) = P(X_4 = x_4 \mid X_2 = x_2, X_3 = x_3)$$

Aplicando este filtro de independencia condicional a cada nodo, la probabilidad conjunta de la Red Bayesiana colapsa de forma exacta en el siguiente producto simplificado:

$$\begin{aligned} P(X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4) = P(X_4 = x_4 \mid X_2 = x_2, X_3 = x_3) \\ \cdot P(X_3 = x_3 \mid X_1 = x_1) \\ \cdot P(X_2 = x_2 \mid X_1 = x_1) \\ \cdot P(X_1 = x_1) \end{aligned}$$

Las cadenas de márkov constituyen uno de los pilares más trascendentales de la teoría de procesos estocásticos y la estadística bayesiana moderna. Sus aplicaciones principales incluyen:

- **MCMC (*Markov Chain Monte Carlo*):** algoritmos de simulación computacional (como el muestreo de Gibbs o Metropolis-Hastings) orientados a extraer muestras de distribuciones posteriores complejas.
- **HMM (*Hidden Markov Models*):** modelos donde los estados reales del sistema son latentes (ocultos) y solo se infieren a través de variables observables correlacionadas.
- **Modelización de Sistemas Físicos:** dinámica de fluidos, colas de espera masivas y transiciones en sistemas computacionales de estados finitos.

En esta sección trabajaremos con $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad indexado por un parámetro de tiempo discreto $T = \{0, 1, 2, \dots\}$.

Definición 3 (Cadena de Márkov). Un proceso estocástico en tiempo discreto $\{X_m\}_{m \geqslant 0}$ 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 $m \geqslant 0$ y para cualquier secuencia de estados $\{x_0, x_1, \dots, x_{m+1}\} \in S$ tales que $P(X_0 = x_0, \dots, X_m = x_m) > 0$.

Observación. La esencia de esta definición formal radica en el aislamiento de la memoria: el comportamiento futuro del proceso (X_{m+1}) condicionado a toda su historia evolutiva real depende única y exclusivamente del estado presente del sistema (X_m), volviéndose el pasado remoto completamente redundante.

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:

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

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:

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

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:

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

Ejemplo. Vamos a ver un ejemplo concreto de una cadena de Márkov con un espacio de estados finito $S = \{1, 2, 3\}$. Consideremos la probabilidad conjunta de que el proceso estocástico tome valores específicos en los tres primeros pasos de tiempo. Aplicando la regla de la cadena y la Propiedad de Markov (falta de memoria), simplificamos la dependencia del pasado lejano:

$$\begin{aligned} P(X_1 = x, X_2 = y, X_3 = z) &= P(X_3 = z \mid X_2 = y, X_1 = x) \cdot P(X_2 = y \mid X_1 = x) \cdot P(X_1 = x) \\ &= P(y, z) \cdot P(x, y) \cdot p(x) \end{aligned}$$

Para hallar la función de masa de probabilidad marginal en el tiempo $t = 3$, aplicamos el teorema de la probabilidad total (marginalizando sobre todos los estados intermedios posibles $x, y \in S$):

$$P(X_3 = z) = p_3(z) = \sum_{x,y \in S} P(y, z) P(x, y) \cdot p(x)$$

Nos centramos ahora en un caso más simple con un espacio de estados binario $S = \{1, 2\}$, donde tenemos

- **Vector de Distribución Inicial (λ_0):** para un espacio de dos estados $S = \{1, 2\}$, definimos:

$$\lambda_0 = (p(1), p(2)) = \left(\frac{3}{4}, \frac{1}{4}\right)$$

- **Matriz de Transición (P):** es una matriz estocástica de dimensiones $g \times g$ (donde $g = |S|$) con entradas no negativas cuyas filas suman estrictamente 1.

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

  y viene dada de la siguiente forma

$$P = \begin{pmatrix} \frac{1}{4} & \frac{3}{4} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

  donde cada fila representa un estado origen y cada columna un estado destino, cumpliendo la condición de que la suma de cada fila sea igual a 1.

Calculemos de forma explícita la distribución en el paso $t = 2$ analizando cada estado individualmente mediante la Ley de Probabilidad Total:

$$\begin{aligned} P(X_2 = 1) &= P(X_2 = 1 \mid X_1 = 1)P(X_1 = 1) + P(X_2 = 1 \mid X_1 = 2)P(X_1 = 2) \\ &= \frac{1}{4} \cdot \frac{3}{4} + \frac{1}{2} \cdot \frac{1}{4} = \frac{3}{16} + \frac{1}{8} = \frac{5}{16} \end{aligned}$$

$$\begin{aligned} P(X_2 = 2) &= P(X_2 = 2 \mid X_1 = 1)P(X_1 = 1) + P(X_2 = 2 \mid X_1 = 2)P(X_1 = 2) \\ &= \frac{3}{4} \cdot \frac{3}{4} + \frac{1}{2} \cdot \frac{1}{4} = \frac{9}{16} + \frac{1}{8} = \frac{11}{16} \end{aligned}$$

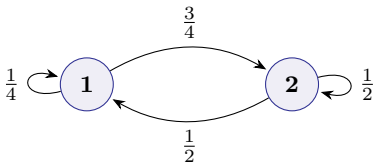
Validación de la medida de probabilidad: $\frac{5}{16} + \frac{11}{16} = 1$

Este mecanismo algebraico tedioso se resuelve elegantemente en Data Science mediante el **álgebra lineal numérico** operando el vector fila por la matriz estocástica:

Paso 1 (Evolución a λ_1):

$$\underbrace{\left(\frac{3}{4}, \frac{1}{4}\right)}_{\text{Estado Inicial } \lambda_0} \overbrace{\begin{pmatrix} \frac{1}{4} & \frac{3}{4} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}}^{\text{Matriz } P} = \left(\frac{5}{16}, \frac{11}{16}\right)$$

Grafo de Transición



Paso 2 (Evolución a λ_2):

$$\left(\frac{5}{16}, \frac{11}{16}\right) \begin{pmatrix} \frac{1}{4} & \frac{3}{4} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} = \left(\frac{3}{4}, \frac{1}{4}\right) \begin{pmatrix} \frac{1}{4} & \frac{3}{4} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}^2$$

3.2. Time Homogeneous Markov Chains

Definición 3.2.4. 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 3.2.5. 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 Finito-Dimensional del Proceso

En esta sección, veremos cómo calcular la probabilidad de que una Cadena de Markov describa una trayectoria exacta y específica a lo largo del tiempo. Para ello, definimos las funciones de masa de probabilidad en etapas tempranas o iniciales del sistema mediante vectores de estado:

- **Distribución Inicial (p_0):** $p_0(x) = P(X_0 = x)$, la probabilidad del estado inicial en el tiempo $t = 0$.
- **Distribución en el Paso Uno (p_1):** $p_1(x) = P(X_1 = x)$.

La distribución de probabilidad conjunta (distribución finito-dimensional) para una secuencia histórica de tamaño n se descompone recursivamente mediante la propiedad de Markov como el producto encadenado de $n-1$ transiciones individuales y la masa inicial:

$$\begin{aligned} P(X_1 = x_1, \ldots, X_n = x_n) &= P(X_n = x_n \mid X_{n-1} = x_{n-1}) \cdots P(X_2 = x_2 \mid X_1 = x_1) \cdot P(X_1 = x_1) \\ &= \underbrace{P(x_{n-1}, x_n) \cdot P(x_{n-2}, x_{n-1}) \cdots P(x_1, x_2)}_{n-1 \text{ factores de transición}} \cdot p_1(x_1) \end{aligned}$$

Desplazamiento del Índice Temporal ($t = 0$): Si reordenamos la indexación temporal para hacer explícito el origen cronológico desde el estado base $X_0 = x_0$, obtenemos la trayectoria exacta para un horizonte de n pasos intermedios:

$$P(X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) = \underbrace{P(x_{n-1}, x_n) \cdot P(x_{n-2}, x_{n-1}) \cdots P(x_0, x_1)}_{n \text{ factores de transición}} \cdot p_0(x_0)$$

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 $x_0, x_1, \ldots, x_{n-1} \in S$.

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 $x_0, x_1, \ldots, x_{n-1} \in S$:

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

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

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

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 anteriormenten 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 3.3.6. 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

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

Teorema 3.3.1 (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

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

□

Nos interesa analizar la tasa de cambio de la probabilidad de estar en un estado específico j entre el paso m y el paso $m+1$. Evaluamos la diferencia:

$$P(X_{m+1}=j) - P(X_m=j)$$

Utilizando el Teorema de la Probabilidad Total y la propiedad de Markov, expandimos el primer término. Asimismo, multiplicamos el término $P(X_m=j)$ por 1 en su forma estocástica unitaria, bajo la identidad indiscutible $\sum_{x\in S} P(X_{m+1}=x \mid X_m=j)=1$:

$$\begin{aligned}
P(X_{m+1}=j) - P(X_m=j) &= \\
&= \sum_{x\in S} P(X_{m+1}=j \mid X_m=x)P(X_m=x) - \sum_{x\in S} P(X_{m+1}=x \mid X_m=j)P(X_m=j) = \\
&= \sum_{x\in S} \Big[\underbrace{P(X_{m+1}=j \mid X_m=x)P(X_m=x)}_{\text{Flujo de entrada hacia } j} - \underbrace{P(X_{m+1}=x \mid X_m=j)P(X_m=j)}_{\text{Flujo de salida desde } j} \Big]
\end{aligned}$$

Reescribiendo en la notación simplificada de vectores de estado $p_m(j) = P(X_m=j)$ y elementos de la matriz de transición $P(x,j)$, obtenemos la forma discreta de la ecuación de balance:

$$p_{m+1}(j) - p_m(j) = \sum_{x\in S} \big[P(x,j)p_m(x) - P(j,x)p_m(j)\big]$$

Para transicionar al tiempo continuo, parametrizamos los pasos temporales en función de intervalos pequeños Δt, de tal manera que el tiempo físico sea $t_m = m\Delta t$. Dividiendo toda la expresión por Δt, la ecuación se transforma en:

$$\frac{p_{t_m+\Delta t}(j) - p_{t_m}(j)}{\Delta t} = \sum_{x\in S} \big[P(x,j)p_{t_m}(x) - P(j,x)p_{t_m}(j)\big]\cdot\frac{1}{\Delta t}$$

Al tomar el límite infinitesimal cuando $\Delta t \to 0$, las diferencias finitas se convierten en derivadas respecto al tiempo. Definimos ahora

$$\lim_{\Delta t\to 0} \frac{P(x,y)}{\Delta t} = Q(x,y) \quad (\text{para } x\neq y)$$

que son las **tasas de transición continuas** (elementos del generador infinitesimal o matriz Q) con la que el sistema decide abandonar tu estado j para irse al estado x.

Llegamos formalmente a la **Ecuación Maestra en Tiempo Continuo**:

$$\frac{dp_t(j)}{dt} = \sum_{x\in S} \big[Q(x,j)p_t(x) - Q(j,x)p_t(j)\big]$$

que representa la velocidad instantánea a la que cambia la probabilidad de estar en el estado j en el momento exacto t, es decir, si es positiva el flujo neto de probabilidad hacia j es mayor que el flujo de salida, y viceversa.

Definición 3.3.7. 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.

Observación. El vector de probabilidad π representa el estado de equilibrio estadístico a largo plazo de la cadena de Markov. Una vez que el sistema adopta esta distribución de probabilidad, se congela dinámicamente: la probabilidad de estar en cada estado no volverá a cambiar jamás, sin importar cuántos pasos adicionales dé la cadena.

3.3.1. Global Balance vs Detailed Balance

Definición 3.3.8. 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

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

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 3.3.9. 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.

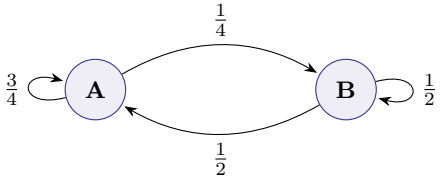
Consecuencia Inmediata (Matrices de Transición Simétricas): Si una cadena de Markov posee una probabilidad de transición estrictamente simétrica entre sus estados intermedios ($P(j,x) = P(x,j)$ para todo $j \neq x$), se deduce directamente de la condición de reversibilidad que:

$$\frac{\pi(x)}{\pi(j)} = 1 \implies \pi(x) = \pi(j) \quad \forall x, j \in S$$

Esto demuestra matemáticamente que la distribución invariante de cualquier cadena con transiciones simétricas es una **distribución uniforme**, es decir, todos los estados tienen exactamente la misma probabilidad a largo plazo.

Ejemplo. Consideremos una cadena de Markov de dos estados, $S = \{A, B\}$, caracterizada por la siguiente matriz estocástica de transición y su correspondiente grafo de estados:

$$P = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$



donde en la matriz de transición, la primera fila corresponde al estado A y la segunda fila al estado B, mientras que las columnas representan los estados destino. Por ejemplo, $P(A,B) = \frac{1}{4}$ indica que la probabilidad de transicionar del estado A al estado B en un solo paso es del 25 %.

Buscamos un vector de probabilidad estacionario $\pi = (\pi_1, \pi_2)$ que actúe como punto fijo del sistema, es decir, que satisfaga la condición de invarianza $\pi P = \pi$:

$$(\pi_1, \pi_2) \begin{pmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} = (\pi_1, \pi_2)$$

Al realizar el producto matriz-vector, el problema algebraico se reduce a resolver el siguiente sistema de ecuaciones lineales simultáneas:

$$\begin{cases} \frac{3}{4}\pi_1 + \frac{1}{2}\pi_2 = \pi_1 \\ \frac{1}{4}\pi_1 + \frac{1}{2}\pi_2 = \pi_2 \end{cases} \implies \frac{1}{2}\pi_2 = \frac{1}{4}\pi_1 \implies \pi_1 = 2\pi_2$$

Imponiendo la condición necesaria de normalización para toda medida de probabilidad ($\pi_1 + \pi_2 = 1$):

$$2\pi_2 + \pi_2 = 1 \implies 3\pi_2 = 1 \implies \pi_2 = \frac{1}{3}, \quad \pi_1 = \frac{2}{3}$$

Ejemplo (Descomposición Espectral (Eigen-extension)). Consideremos una cadena de Markov de dos estados cuya matriz de transición P es **simétrica** y está parametrizada por las variables a y b:

$$P = \begin{pmatrix} a & b \\ b & a \end{pmatrix} \qquad \text{donde } 0 \leq a, b \leq 1 \quad \text{y} \quad a + b = 1$$

Como las filas de toda matriz estocástica deben sumar estricta unidad, se cumple que $b = 1 - a$. Dado que la matriz es simétrica ($P = P^T$), sus autovectores por la izquierda (distribuciones) coinciden espacialmente con sus autovectores por la derecha. El objetivo será ver el procedimiento a seguir para poder obtener

$$P^n$$

de forma sencilla.

Calcular autovalores. Para hallar los autovalores de la matriz de transición, resolvemos el polinomio característico mediante el determinante $\det(P - \lambda I) = 0$:

$$\det \begin{pmatrix} a - \lambda & b \\ b & a - \lambda \end{pmatrix} = 0 \implies (a - \lambda)^2 - b^2 = 0$$

Extrayendo la raíz cuadrada en ambos miembros obtenemos las dos soluciones del sistema:

$$\begin{cases} a - \lambda = -b \implies \lambda_1 = a + b = 1 \\ a - \lambda = +b \implies \lambda_2 = a - b \end{cases}$$

Observación. El primer autovalor siempre será $\lambda_1 = 1$ para cualquier matriz estocástica. El vector asociado a este autovalor determina la **distribución invariante** del sistema. Operando por la izquierda:

$$(\pi_1, \pi_2) \begin{pmatrix} a & b \\ b & a \end{pmatrix} = 1 \cdot (\pi_1, \pi_2) \implies \pi_1 = \pi_2 = \frac{1}{2} \implies \pi = \left(\frac{1}{2}, \frac{1}{2}\right)$$

Cálculo y Normalización Autovectores. Buscamos los autovectores columna por la derecha resolviendo el sistema homogéneo $(P - \lambda_i I)v = 0$:

- **Para $\lambda_1 = 1$:**

$$\begin{pmatrix} a - 1 & b \\ b & a - 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \xrightarrow{a-1=-b} \begin{pmatrix} -b & b \\ b & -b \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \implies x = y \implies v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

  Normalizando vectorialmente bajo la norma euclídea ($\|v\| = \sqrt{x^2 + y^2}$), obtenemos

$$v_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

- **Para $\lambda_2 = a - b$:**

$$\begin{pmatrix} a - (a - b) & b \\ b & a - (a - b) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \implies \begin{pmatrix} b & b \\ b & b \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \implies x = -y \implies v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

  Normalizando euclídeamente obtenemos

$$v_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}$$

Diagonalización y Matriz de Transición Ortogonal. Al ser P una matriz simétrica real, el teorema espectral garantiza que puede ser diagonalizada mediante una matriz ortogonal U (donde la inversa coincide con su transpuesta, $U^{-1} = U^T$), construida acoplando los autovectores normalizados:

$$U = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \implies U^T U = I$$

Por lo tanto, la descomposición espectral de la matriz $P = U\Lambda U^T$ se escribe como:

$$P = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & a-b \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} + \frac{a-b}{2} & \frac{1}{2} - \frac{a-b}{2} \\ \frac{1}{2} - \frac{a-b}{2} & \frac{1}{2} + \frac{a-b}{2} \end{pmatrix}$$

Evolución Temporal. La gran ventaja de la diagonalización es que calcular la evolución temporal a n pasos se reduce a elevar el núcleo central de autavalores a la potencia n, dado que los términos ortogonales intermedios colapsan recursivamente ($U^T U = I$):

$$P^n = (U\Lambda U^T)(U\Lambda U^T)\cdots(U\Lambda U^T) = U\Lambda^n U^T$$

$$P^n = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 1^n & 0 \\ 0 & (a-b)^n \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} + \frac{(a-b)^n}{2} & \frac{1}{2} - \frac{(a-b)^n}{2} \\ \frac{1}{2} - \frac{(a-b)^n}{2} & \frac{1}{2} + \frac{(a-b)^n}{2} \end{pmatrix}$$

Ahora calculemos el límite para $n \to \infty$ y veámos que obtenemos.

Como $0 \leq a, b \leq 1$ y $a + b = 1$, la diferencia en valor absoluto satisface estrictamente $|a - b| < 1$ (salvo en casos triviales de matrices identidad o de permutación pura). Al calcular el límite cuando el tiempo tiende a infinito ($n \to \infty$), el segundo autovalor se desvanece por completo:

$$\lim_{n\to\infty} (a-b)^n = 0$$

De este modo, la matriz de transiciones colapsa en el equilibrio estable y de largo plazo del sistema:

$$\lim_{n\to\infty} P^n = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

donde

- La primera fila te dice qué pasa a largo plazo si empezaste en el estado 1: terminas con probabilidad $\left(\frac{1}{2}, \frac{1}{2}\right)$.
- La segunda fila te dice qué pasa a largo plazo si empezaste en el estado 2: terminas con probabilidad $\left(\frac{1}{2}, \frac{1}{2}\right)$.

Cada fila de la matriz resultante converge de manera idéntica hacia la distribución invariante uniforme $\pi = \left(\frac{1}{2}, \frac{1}{2}\right)$, demostrando que el sistema olvida por completo su estado inicial tras un horizonte temporal prolongado.

3.3.2. Teorema Espectral para Cadenas de Markov Reversibles

Teorema 3.3.2 (Teorema Espectral para Cadenas de Markov Reversibles). *Para cualquier cadena de Markov homogénea en el tiempo que sea* ***reversible****, la probabilidad de transición en n pasos desde un estado x a un estado y puede descomponerse analíticamente en función de sus autovalores y autovectores mediante el teorema espectral*

$$P^n(x, y) = \pi(y) + \sum_{i=2}^{g} \varphi_i(x)\varphi_i(y)\lambda_i^n \tag{3.1}$$

donde:

- $\pi(y)$ *representa la distribución invariante (asociada al primer autovalor* $\lambda_1 = 1$*).*
- λ_i *son los autovalores secundarios, los cuales satisfacen rigurosamente* $|\lambda_i| < 1$ *(garantizando la convergencia).*
- φ_i *son los autovectores ortonormalizados de la matriz de transición.*

Ejemplo. Validamos el teorema general utilizando la matriz simétrica de dos estados $S = \{1, 2\}$ analizada previamente:

$$P = \begin{pmatrix} a & b \\ b & a \end{pmatrix} \qquad \text{con } a + b = 1$$

Los elementos espectrales normalizados calculados para este sistema son:

$$\text{Distribución Invariante:} \quad \pi = \left(\frac{1}{2}, \frac{1}{2}\right)$$

$$\text{Primer Autovalor:} \quad \lambda_1 = 1 \longrightarrow \varphi_1 = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$$

$$\text{Segundo Autovalor:} \quad \lambda_2 = a - b \longrightarrow \varphi_2 = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$$

La matriz de cambio de base ortogonal U (construida acoplando los autovectores en columnas) cumple la condición de ortogonalidad unitaria ($U^T U = I$):

$$U = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}$$

Veámos ahora si se cumple el teorema espectral al realizar la multiplicación de matrices $P^n = U\Lambda^n U^T$, la matriz de transición en n pasos resulta:

$$P^{(n)} = P^n = \begin{pmatrix} \frac{1}{2} + \frac{1}{2}(a-b)^n & \frac{1}{2} - \frac{1}{2}(a-b)^n \\ \frac{1}{2} - \frac{1}{2}(a-b)^n & \frac{1}{2} + \frac{1}{2}(a-b)^n \end{pmatrix}$$

A continuación, descomponemos cada celda de la matriz para demostrar que coincide con la estructura del teorema general

$$P^n(x, y) = \pi(y) + \varphi_2(x)\varphi_2(y)\lambda_2^n$$

véamoslo

$$P^n(1,1) = \frac{1}{2} + \frac{1}{2}(a-b)^n = \pi(1) + \varphi_2^2(1)\lambda_2^n$$

$$P^n(1,2) = \frac{1}{2} - \frac{1}{2}(a-b)^n = \pi(2) + \varphi_2(1)\varphi_2(2)\lambda_2^n$$

$$P^n(2,1) = \frac{1}{2} - \frac{1}{2}(a-b)^n = \pi(1) + \varphi_2(2)\varphi_2(1)\lambda_2^n$$

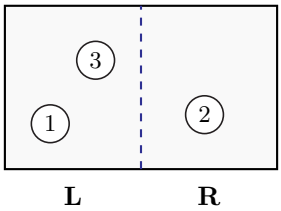
$$P^n(2,2) = \frac{1}{2} + \frac{1}{2}(a-b)^n = \pi(2) + \varphi_2^2(2)\lambda_2^n$$

3.4. Ehrenfest Urn Model

El Modelo de Urna de Ehrenfest fue publicado originalmente por Paul y Tatiana Ehrenfest en 1907 y 1911. Es una herramienta clásica para modelar la difusión de partículas y estudiar la mecánica estadística mediante cadenas de Markov.

Configuración del Modelo de Microestados

Definiremos el modelo de urnas de Ehrenfest con un ejemplo simple usando $n = 3$ bolas y $g = 2$ contenedores. Imaginemos una urna dividida en dos compartimentos: izquierdo (L) y derecho (R), que albergan un total de $n = 3$ bolas numeradas de forma única. En cada unidad discreta de tiempo, se selecciona una bola al azar con probabilidad uniforme ($1/n = 1/3$) y se traslada al compartimento opuesto.



Asignamos un indicador binario X_i para denotar la posición de la bola i:

$$\widetilde{X}_1 = L \longleftrightarrow X_1 = 1 \quad \text{(Bola 1 en } L\text{)}$$
$$\widetilde{X}_2 = R \longleftrightarrow X_2 = 0 \quad \text{(Bola 2 en } R\text{)}$$
$$\widetilde{X}_3 = L \longleftrightarrow X_3 = 1 \quad \text{(Bola 3 en } L\text{)}$$

El vector binario del sistema describe por completo el **microestado** exacto del proceso (e.g., el estado actual es 101).

Cadena de Markov de Microestados (S_I)

Si definimos el espacio de estados S_I basándonos en la configuración individual de cada partícula, existen $2^n = 2^3 = 8$ estados posibles indexados en binario desde 000 hasta 111. Dado que en cada paso elegimos exactamente una bola entre las 3 disponibles con igual probabilidad, la matriz estocástica de transición es:

$$P_I = \begin{array}{c} \\ 000 \\ 001 \\ 010 \\ 100 \\ 011 \\ 101 \\ 110 \\ 111 \end{array} \begin{array}{c} \begin{array}{cccccccc} 000 & 001 & 010 & 100 & 011 & 101 & 110 & 111 \end{array} \\ \begin{bmatrix} 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 1/3 & 0 & 1/3 & 0 \\ 1/3 & 0 & 0 & 0 & 0 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 1/3 \\ 0 & 1/3 & 0 & 1/3 & 0 & 0 & 0 & 1/3 \\ 0 & 0 & 1/3 & 1/3 & 0 & 0 & 0 & 1/3 \\ 0 & 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 \end{bmatrix} \end{array}$$

La distribución invariante uniforme para este espacio de microestados es simétrica y de igual probabilidad:

$$\pi_I = \Big(\underbrace{\frac{1}{8}}_{000}, \underbrace{\frac{1}{8}}_{001}, \underbrace{\frac{1}{8}}_{010}, \underbrace{\frac{1}{8}}_{100}, \underbrace{\frac{1}{8}}_{011}, \underbrace{\frac{1}{8}}_{101}, \underbrace{\frac{1}{8}}_{110}, \underbrace{\frac{1}{8}}_{111} \Big) \tag{3.2}$$

Cadena de Macroestados o de Conteo (S_S)

Para simplificar el sistema numéricamente, podemos definir una variable macroscópica de interés: el número total de bolas presentes en el compartimento izquierdo, denotado como

$$Y_1 = \sum_{i=1}^{n} X_i$$

y el nuevo espacio reducido de estados es simplemente $S_S = \{0, 1, 2, 3\}$. Las leyes de transición para este conteo dependen directamente del número actual de bolas en L:

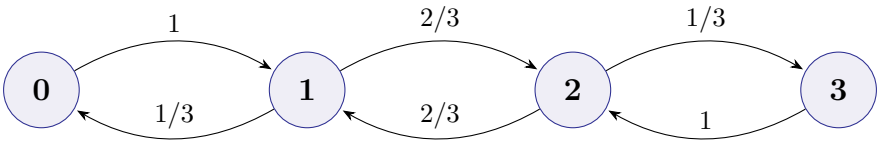
- Si hay k bolas en L, la probabilidad de elegir una de ellas y moverla a R (pasando al estado $k-1$) es
$$\frac{k}{n}$$
- La probabilidad de elegir una de las $n-k$ bolas de R y moverla a L (pasando al estado $k+1$) es
$$\frac{n-k}{n}$$

La matriz de transición colapsada P_S y su correspondiente distribución estacionaria π_S son:

$$P_S = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{array} \qquad \pi_S = \left(\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8}\right)$$

Esta estructura matemática modela una **birth-death chain**, caracterizada como una caminata aleatoria con sesgo hacia el centro dentro de un retículo entero finito. Presenta las siguientes propiedades cruciales:

- **Irreducible:** es posible acceder a cualquier estado desde cualquier otro en un número finito de pasos; el espacio de estados forma una única clase comunicante.
- **Periódica (Periodo $d = 2$):** es imposible permanecer en el mismo estado en un solo paso ($P(x, x) = 0, \forall x$). El sistema oscila de manera obligatoria entre estados pares e impares en cada transición temporal.



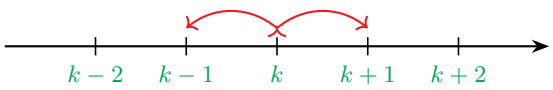
En este caso tenemos un sistema que no converge por oscilación, ya que el periodo es $d = 2$, lo cual es fácil de ver en la cadena de Markov de conteo. Por ejemplo, si comenzamos en el estado 0, en el siguiente paso solo podemos ir al estado 1, y desde allí solo podemos ir a los estados 0 o 2. Por lo tanto, no es posible permanecer en el mismo estado en un solo paso, lo que genera una oscilación entre estados pares e impares.

3.4.1. Birth-Death Chains

Definición 3.4.10. Un proceso estocástico se define como una **birth-death chain** si, estando en un estado discreto $k \in \mathbb{N}$ (o en el retículo entero $\mathbb{Z}$), las únicas transiciones permitidas con probabilidad no nula en el siguiente paso de tiempo son hacia sus vecinos inmediatos o quedarse quieto:

$$k \longrightarrow k-1 \quad \text{(Muerte)}, \qquad k \longrightarrow k \quad \text{(Permanencia)}, \qquad k \longrightarrow k+1 \quad \text{(Nacimiento)}$$

Cualquier otro movimiento de salto largo de largo alcance está estrictamente prohibido.



Seguiremos estudiando el modelo, para ello recordemos las probabilidades condicionales clásicas del modelo de conteo de Ehrenfest con n partículas, donde las fronteras de la urna actúan como barreras reflejantes puras:

$$P(Y_1(t+1) = k+1 \mid Y_1(t) = k) = \frac{n-k}{n} \qquad P(Y_1(t+1) = k-1 \mid Y_1(t) = k) = \frac{k}{n}$$

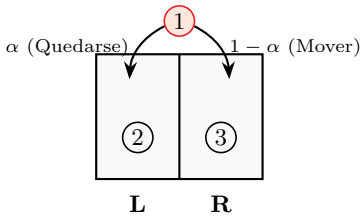
$$P(Y_1(t+1) = k \mid Y_1(t) = k) = 0 \qquad P(Y_1(t+1) = 1 \mid Y_1(t) = 0) = 1$$

$$P(Y_1(t+1) = n-1 \mid Y_1(t) = n) = 1$$

donde tenemos k partículas en la urna izquierda y $n-k$ partículas en la urna derecha.

Versión Perezosa (α-retrasada)

Para solucionar el problema que teníamos antes, es decir, que el sistema no converge por oscilación (periodo $d = 2$), introduciremos una probabilidad de "pereza" $\alpha \in (0, 1)$ para romper la periodicidad. En cada unidad de tiempo, lanzamos una moneda sesgada: con probabilidad α el sistema decide quedarse congelado (no hacer nada), y con probabilidad $1 - \alpha$ se elige una partícula al azar para cambiar de compartimento.



Parámetro de Pereza: $\alpha \in (0, 1)$

Matriz de Transición de Microestados Perezosa (P_I)

Al añadir la probabilidad α en la diagonal principal para representar el bucle de permanencia, la matriz de microestados distribuye el resto de la masa uniformemente entre las 3 partículas vecinas del hipercubo

$$P_I = \begin{array}{c} \\ 000 \\ 001 \\ 010 \\ 100 \\ 011 \\ 101 \\ 110 \\ 111 \end{array} \begin{array}{c} \begin{array}{cccccccc} 000 & 001 & 010 & 100 & 011 & 101 & 110 & 111 \end{array} \\ \begin{pmatrix} \alpha & \frac{1-\alpha}{3} & \frac{1-\alpha}{3} & \frac{1-\alpha}{3} & 0 & 0 & 0 & 0 \\ \frac{1-\alpha}{3} & \alpha & 0 & 0 & \frac{1-\alpha}{3} & \frac{1-\alpha}{3} & 0 & 0 \\ \frac{1-\alpha}{3} & 0 & \alpha & 0 & \frac{1-\alpha}{3} & 0 & \frac{1-\alpha}{3} & 0 \\ \frac{1-\alpha}{3} & 0 & 0 & \alpha & 0 & \frac{1-\alpha}{3} & \frac{1-\alpha}{3} & 0 \\ 0 & \frac{1-\alpha}{3} & \frac{1-\alpha}{3} & 0 & \alpha & 0 & 0 & \frac{1-\alpha}{3} \\ 0 & \frac{1-\alpha}{3} & 0 & \frac{1-\alpha}{3} & 0 & \alpha & 0 & \frac{1-\alpha}{3} \\ 0 & 0 & \frac{1-\alpha}{3} & \frac{1-\alpha}{3} & 0 & 0 & \alpha & \frac{1-\alpha}{3} \\ 0 & 0 & 0 & 0 & \frac{1-\alpha}{3} & \frac{1-\alpha}{3} & \frac{1-\alpha}{3} & \alpha \end{pmatrix} \end{array}$$

donde efectivamente cumple que las filas suman 1 como debe ser

$$\alpha + 3\left(\frac{1-\alpha}{3}\right) = \alpha + 1 - \alpha = 1$$

La consecuencia principal de esta matriz es la **aperiodicidad**, pues al forzar que los elementos de la diagonal sean estrictamente positivos ($P(x, x) = \alpha > 0$), la cadena rompe los ciclos rígidos de alternancia par/impar. La cadena de Markov ahora es **irreducible y aperiódica**, lo que garantiza, según el Teorema Ergódico, el cumplimiento del límite de convergencia asintótica de largo plazo.

Matriz de Transición de Macroestados Perezosa (P_S)

Colapsando el sistema al conteo masivo de partículas en el compartimento izquierdo, la matriz tridiagonal hereda el parámetro de retraso estructural:

$$P_S = \begin{pmatrix} \alpha & 1-\alpha & 0 & 0 \\ \frac{1-\alpha}{3} & \alpha & \frac{2(1-\alpha)}{3} & 0 \\ 0 & \frac{2(1-\alpha)}{3} & \alpha & \frac{1-\alpha}{3} \\ 0 & 0 & 1-\alpha & \alpha \end{pmatrix}$$

Simulación Asintótica

Si parametrizamos el modelo fijando un valor simétrico clásico de pereza $\alpha = \frac{1}{2}$, la matriz se simplifica aritméticamente. Al elevar la matriz estocástica a una potencia suficientemente grande ($n \gg 1$), el sistema converge de forma exacta hacia el equilibrio estacionario binomial independiente del estado inicial:

$$\begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/6 & 1/2 & 1/3 & 0 \\ 0 & 1/3 & 1/2 & 1/6 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix}^n \xrightarrow{n\to\infty} \begin{pmatrix} 1/8 & 3/8 & 3/8 & 1/8 \\ 1/8 & 3/8 & 3/8 & 1/8 \\ 1/8 & 3/8 & 3/8 & 1/8 \\ 1/8 & 3/8 & 3/8 & 1/8 \end{pmatrix}$$

Cada fila de la matriz límite se ha convertido en una copia exacta del vector estacionario

$$\pi_S = \left(\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8}\right)$$

validando numéricamente que la cadena de Markov perezosa ha alcanzado el equilibrio ergódico.

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:

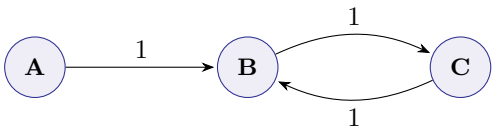
- **Estado Transitorio:** existe una probabilidad estrictamente positiva de que, una vez abandonado el estado, el proceso jamás regrese a él.
- **Estado Recurrente:** el proceso regresa al estado con probabilidad igual a la unidad de forma casi segura.
- **Conjunto Absorbente / Cerrado:** un subconjunto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

Definición 3.5.11. El **tiempo de parada** T_A para un conjunto absorbente $A \subset S$ se define como el primer instante de tiempo en que la cadena alcanza dicho conjunto

$$T_A = \text{ínf}\{t \geq 0 : X_t \in A\}$$

Ejemplo. Consideremos un espacio de tres estados $S = \{A, B, C\}$ definido por la siguiente matriz de transición P:

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$



Si el proceso inicia en el estado A, se genera la trayectoria determinista inalterable:

$$\boldsymbol{A \longrightarrow B \longrightarrow C \longrightarrow B \longrightarrow C \ldots}$$

donde tenemos

- **Estado Transitorio:** A es transitorio; una vez que salta a B, la probabilidad de regresar a A es 0.
- **Conjunto Absorbente:** el par $\{B, C\}$ constituye una clase cerrada o conjunto absorbente de estados. Individualmente, B y C son **estados recurrentes**.

En Data Science, suele interesar calcular el tiempo de parada

$$T_{\{B,C\}} = \text{ínf}\{t \geq 0 : X_t \in \{B, C\}\}$$

necesario para alcanzar dicha clase. Para esta matriz, partiendo de A, el tiempo es determinista

$$T_{\{B,C\}} = 1$$

Identidad (Estacionaria)

$$\begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}^*$$

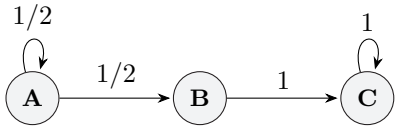


Figura 3.1: Cadena de Markov determinista con comportamiento absorbente y estacionario.

Trampa Absorbente dual

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

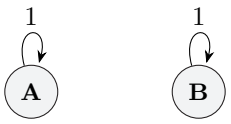


Figura 3.2: Cadena de Markov con dos estados absorbentes independientes.

3.5.1. Cadenas de Markov Deterministas

Las siguientes matrices ilustran comportamientos donde el azar es nulo (probabilidades puras 1 o 0), generando trayectorias cíclicas o estáticas rígidas

- **Identidad (Estacionaria):** la cadena permanece en el mismo estado indefinidamente. Ejemplo en Figura 3.1.
- **Trampa Absorbente dual:** dos estados absorbentes independientes, cada uno atrapando la cadena en su propio ciclo. Ejemplo en la Figura 3.2.
- **Oscilación Cíclica:** la cadena alterna entre dos estados de manera indefinida. Ejemplo en la Figura 3.3.
- **Permutación Cíclica (3 Estados):** la cadena recorre un ciclo de tres estados de forma repetitiva. Ejemplo en la Figura 3.4.

Definición 3.5.12. 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

Oscilación Cíclica

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

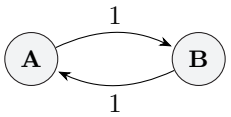


Figura 3.3: Cadena de Markov con oscilación cíclica entre dos estados.

Permutación Cíclica (3 Estados)

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

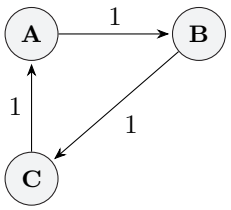


Figura 3.4: Cadena de Markov con permutación cíclica entre tres estados.

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 ($n \to \infty$), el sistema alcanzará un equilibrio perfecto y balanceado:

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

3.6. Class Structures

Definición 3.6.13 (Accesibilidad o Conducción). Se dice que el estado i **conduce al** estado j (notado como $i \to j$) si existe una probabilidad estrictamente positiva de visitar el estado j en algún momento, dado que el sistema comenzó en el estado i:

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

Definición 3.6.14 (Comunicación). Se dice que el estado i **se comunica con** el estado j (notado como $i \longleftrightarrow j$) si se cumplen simultáneamente que $i \to j$ y $j \to i$.

Teorema 3.6.3. *Para dos estados distintos $i \neq j$, las siguientes afirmaciones son equivalentes:*

1. $i \to j$.
2. *Existen estados intermedios* $i_1, i_2, \ldots, i_{n-1}$ *tales que* $P_{ii_1} P_{i_1 i_2} \ldots P_{i_{n-1} j} > 0$.
3. $P_{ij}^{(n)} > 0$ *para algún* $n \geqslant 1$.

Demostración. La equivalencia se sostiene mediante las siguientes relaciones fundamentales.

1) $\iff$ 3) Para cualquier paso fijo n, el evento $\{X_n = j\}$ es un subconjunto del evento general de "visitar j alguna vez". Por subaditividad (cota de Boole), podemos acotar superior e inferiormente:

$$P_{ij}^{(n)} \leqslant P(\exists m \geqslant 0 : X_m = j \mid X_0 = i) \leqslant \sum_{m=0}^{\infty} P_{ij}^{(m)}$$

Si la probabilidad central es > 0, la suma del extremo derecho debe tener al menos un término estrictamente positivo. Inversamente, si algún $P_{ij}^{(n)} > 0$, la probabilidad central es obligatoriamente mayor que cero.

2) $\iff$ 3) Por la expansión de las ecuaciones de Chapman-Kolmogorov, sabemos que:

$$P_{ij}^{(n)} = \sum_{i_1,\ldots,i_{n-1} \in S} P_{ii_1} P_{i_1 i_2} \ldots P_{i_{n-1} j}$$

Dado que todas las probabilidades de transición son no negativas ($P_{kl} \geqslant 0$), este sumatorio es estrictamente positivo si y solo si existe al menos un término (un camino específico) en el sumatorio cuyo producto sea estrictamente mayor que cero.

□

Proposición 3.6.4. *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 $i \to i$.

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

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

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

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

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

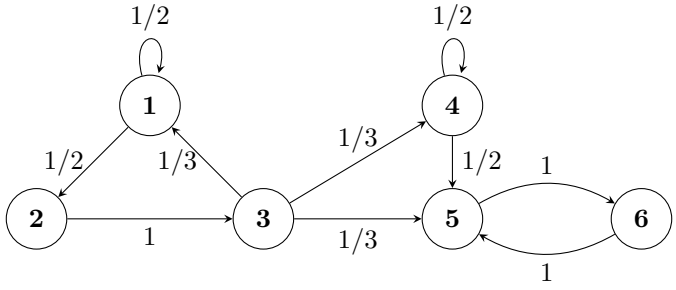
□

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**.

Ejemplo. Considere una cadena de Markov homogénea con espacio de estados $S = \{1, 2, 3, 4, 5, 6\}$ y matriz de probabilidad de transición P dada por la siguiente estructura matricial:

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 1/3 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

El comportamiento dinámico de las transiciones e interconexiones de la cadena se puede visualizar de forma equivalente mediante el siguiente grafo dirigido:



A partir de la relación de comunicación entre los estados del sistema, el espacio S se descompone de forma única en las siguientes clases de equivalencia (particiones):

- $C_1 = \{1, 2, 3\} \longrightarrow$ Clase abierta y transitoria.
- $C_2 = \{4\} \longrightarrow$ Clase abierta y transitoria.
- $C_3 = \{5, 6\} \longrightarrow$ Clase **cerrada** y recurrente.

Definición 3.6.15 (Clase Cerrada). Una clase de comunicación C se dice que es **cerrada** si para todo estado $i \in C$, la condición $i \to j$ implica obligatoriamente que $j \in C$. Es decir, una vez que el proceso entra en una clase cerrada, no puede salir de ella.

Definición 3.6.16 (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 3.6.17 (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 $(X_n)_{n\geqslant 0}$ con matriz de probabilidad de transición P que opera sobre el espacio de estados S.

Definición 3.6.18 (Tiempo de Llegada / Hitting Time). El **tiempo de primera llegada** a un subconjunto de estados $A \subseteq S$ es una variable aleatoria $H^A : \Omega \to \{0, 1, 2, \dots\} \cup \{\infty\}$ que cuantifica el número mínimo de pasos necesarios para intersectar la región A:

$$H^A(\omega) = \text{ínf}\{n \geqslant 0 : X_n(\omega) \in A\}$$

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

Definición 3.6.19 (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 3.6.20 (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:

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

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

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

Definición 3.6.21 (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.

Lema 3.6.5. *Sea T una variable aleatoria discreta bien definida que toma valores sobre el conjunto de los enteros no negativos ($\mathbb{N}_0$). Su esperanza matemática se puede calcular de manera alternativa a través de sus funciones de distribución acumulada complementaria mediante las identidades:*

$$E(T) = \sum_{n=0}^{\infty} nP(T = n) = \sum_{n=1}^{\infty} P(T \geqslant n)$$

Demostración. Expandiendo de forma explícita el operador del valor esperado lineal de la variable discreta por definición obtenemos:

$$E(T) = \sum_{n=0}^{\infty} n\, P(T = n)$$

Esta expresión se puede descomponer algebraicamente mediante una suma ordenada de probabilidades límite de cola superior como:

$$E(T) = \sum_{n=1}^{\infty} P(T \geqslant n) = P(T \geqslant 1) + P(T \geqslant 2) + P(T \geqslant 3) + \dots$$

Podemos validar analíticamente esta igualdad disponiendo los términos de las colas de probabilidad mediante un esquema de descomposición matricial infinita:

$$\begin{array}{lcccccc} P(T \geqslant 1) & = & P(T=1) & + & P(T=2) & + & P(T=3) + \dots \\ P(T \geqslant 2) & = & & & P(T=2) & + & P(T=3) + \dots \\ P(T \geqslant 3) & = & & & & & P(T=3) + \dots \end{array}$$

Efectuando la agregación y suma por columnas verticales en el esquema anterior, se recupera exactamente la ponderación lineal original: un término para $P(T = 1)$, dos términos para $P(T = 2)$, tres términos para $P(T = 3)$, y así sucesivamente, completando la demostración.

Por analogía con el caso de variables continuas no negativas, este resultado equivale a evaluar la siguiente integral de supervivencia generalizada:

$$E(T) = \int_0^{\infty} x f_T(x)\, dx = \int_0^{\infty} P(T > x)\, dx$$

ya que tenemos que

$$P(T > x) = \int_x^{\infty} f_T(t)\, dt \implies \int_0^{\infty} P(T > x)\, dx = \int_0^{\infty} \left(\int_x^{\infty} f_T(t)\, dt \right) dx$$

y por tanto

$$\int_0^{\infty} P(T > x)\, dx = \int_0^{\infty} \left(\int_0^{t} dx \right) f_T(t)\, dt = \int_0^{\infty} t f_T(t)\, dt = E(T)$$

□

Teorema 3.6.6. *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:*

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

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

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 $H^A \geqslant 1$. 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:
$$\begin{aligned} h_i^A = P_i(H^A < \infty) &= \sum_{j\in S} P_i(H^A < \infty, X_1 = j) \\ &\overset{(*)}{=} \sum_{j\in S} P(H^A < \infty \mid X_0 = i, X_1 = j)\, P(X_1 = j \mid X_0 = i) \\ &= \sum_{j\in S} (P_{ij}\, h_j^A) = \sum_{j\in S} P_{ij} h_j^A \end{aligned}$$
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:

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

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:

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

Dado que el vector de soluciones alternativo es estrictamente no negativo por hipótesis ($x \geqslant 0$) 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:

$$x_i \geqslant P_i(H^A \leqslant m) \quad \forall m \geqslant 1$$

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:

$$x_i \geqslant \lim_{m \to \infty} P_i(H^A \leqslant m) = P_i(H^A < \infty) = h_i^A$$

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

Teorema 3.6.7. *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:*

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

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

Demostración. Evaluamos el comportamiento del valor esperado analizando el origen del estado inicial de la cadena respecto al subconjunto A:

- Si $X_0 = i$ con $i \in A$, el tiempo de llegada se ejecuta instantáneamente de forma que $H^A = 0$. Por lo tanto, su valor esperado es nulo:

$$k_i^A = E_i[H^A] = 0$$

- Si $X_0 = i$ con $i \notin A$, el proceso consume obligatoriamente un paso temporal elemental, implicando que $H^A \geqslant 1$. Aplicando la propiedad de Markov espacial y condicionando a la trayectoria del primer salto del sistema obtenemos la relación:

$$E_i[H^A \mid X_1 = j] = 1 + E_j[H^A]$$

Desplegando el operador de la esperanza matemática mediante el teorema de la probabilidad total sobre el espacio exhaustivo del primer paso X_1, desarrollamos:

$$\begin{aligned} k_i^A = E_i(H^A) &= \sum_{j \in S} E_i(H^A \mid X_1 = j)\, P_i(X_1 = j) \\ &= \sum_{j \in S} \big(1 + E_j(H^A)\big)\, P_i(X_1 = j) \\ &= \sum_{j \in S} P_i(X_1 = j) + \sum_{j \in S} E_j(H^A)\, P_i(X_1 = j) \end{aligned}$$

Dado que $\sum_{j \in S} P_i(X_1 = j) = 1$, y reconociendo que para todo estado $j \in A$ se cumple la frontera $E_j(H^A) = k_j^A = 0$, el segundo sumatorio se restringe de forma natural a la región complementaria $j \notin A$:

$$k_i^A = 1 + \sum_{j \notin A} k_j^A \cdot P_{ij} = 1 + \sum_{j \notin A} P_{ij} k_j^A$$

Esto demuestra que el vector real k^A satisface estructuralmente la relación lineal de recurrencia (2).

Demostraremos ahora la minimalidad. Supongamos que existe un vector alternativo $y = (y_i)_{i \in S}$ que resuelve el sistema lineal (2) con componentes no negativas. Evaluamos por regiones:

- Si $i \in A$, por la definición de la frontera superior del sistema se cumple directamente que $y_i = 0 = k_i^A$.
- Si $i \notin A$, la ecuación lineal que gobierna la dinámica se escribe como:

$$y_i = 1 + \sum_{j \notin A} P_{ij} y_j$$

Procedemos a expandir la igualdad de manera iterativa sustituyendo la estructura analógica del término y_j dentro del sumatorio complementario:

$$\begin{aligned} y_i &= 1 + \sum_{j \notin A} P_{ij} \left(1 + \sum_{k \notin A} P_{jk} y_k \right) \\ &= 1 + \sum_{j \notin A} P_{ij} + \sum_{j \notin A} \sum_{k \notin A} P_{ij} P_{jk} y_k \end{aligned}$$

Identificando los términos mediante medidas de trayectoria finita, observamos que $1 = P_i(H^A \geqslant 1)$ y que la probabilidad de transición complementaria mapea exactamente la cola superior $\sum_{j \notin A} P_{ij} = P_i(X_1 \notin A) = P_i(H^A \geqslant 2)$:

$$y_i = P_i(H^A \geqslant 1) + P_i(H^A \geqslant 2) + \sum_{j \notin A} \sum_{k \notin A} P_{ij} P_{jk} y_k$$

Reiterando este procedimiento de sustitución inductiva a lo largo de un horizonte finito de m pasos, la expresión se generaliza como:

$$y_i = P_i(H^A \geqslant 1) + \cdots + P_i(H^A \geqslant m) + \sum_{j_1 \notin A} \cdots \sum_{j_m \notin A} P_{i j_1} P_{j_1 j_2} \ldots P_{j_{m-1} j_m} y_{j_m}$$

Debido a la hipótesis de no negatividad del vector ($y \geqslant 0$) junto con la naturaleza no negativa de las celdas de la matriz de transición P_{kl}, el término de la sumatoria múltiple residual es acotado inferiormente por cero. Esto nos permite fijar la desigualdad para cualquier entero finito m:

$$y_i \geqslant \sum_{l=1}^{m} P_i(H^A \geqslant l) \quad \forall m \geqslant 1$$

Aplicando el límite matemático cuando el horizonte de pasos tiende a infinito en ambos extremos de la desigualdad, y recurriendo a la identidad de la esperanza por colas acumuladas de supervivencia para variables enteras no negativas ($\sum_{l=1}^{\infty} P(T \geqslant l) = E(T)$), concluimos:

$$y_i \geqslant \lim_{m\to\infty} \sum_{l=1}^{m} P_i(H^A \geqslant l) = \sum_{l=1}^{\infty} P_i(H^A \geqslant l) = E_i(H^A) = k_i^A$$

Por lo tanto, se verifica formalmente que $y_i \geqslant k_i^A$ para todo estado $i \in S$, demostrando que el vector de tiempos medios k^A constituye la solución no negativa mínima del sistema lineal. □

3.6.2. Stopping Times & Strong Markov Property

Definición 3.6.22 (Tiempo de Parada / Stopping Time). Una variable aleatoria

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

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 $X_0, X_1, \dots, X_n$.

Ejemplo. A continuación, se analizan tres escenarios clásicos fundamentales para comprender la naturaleza de los tiempos de parada:

1. **El tiempo de primer paso a un estado j:** Definido matemáticamente como el primer instante de tiempo estrictamente positivo en el cual la cadena visita el estado objetivo j:

   $$\tau_j = \text{ínf}\{n \geqslant 1 : X_n = j\}$$

   Constituye un tiempo de parada bien definido puesto que el suceso se puede descomponer analíticamente a través de la historia observada hasta el paso n como:

   $$\{\tau_j = n\} = \{X_0 \neq j, X_1 \neq j, \dots, X_{n-1} \neq j, X_n = j\}$$

   Como esta condición solo inspecciona variables hasta el instante n, no requiere información del futuro.

2. **El tiempo de primera llegada H^A a un conjunto A:** Es un tiempo de parada dado que el suceso que describe que la primera colisión con la región A ocurra exactamente en el paso n se expresa de forma análoga mediante:

   $$\{H^A = n\} = \{X_0 \notin A, X_1 \notin A, \dots, X_{n-1} \notin A, X_n \in A\}$$

   Este evento es medible respecto a la información disponible en el instante n.

3. **El tiempo de última salida L^A a un conjunto A (No es tiempo de parada):** Definido como el instante temporal supremo en el cual el sistema interactúa por última vez con la región A:

   $$L^A = \sup\{n \geqslant 0 : X_n \in A\}$$

   Nota importante: El tiempo de última salida no constituye un tiempo de parada. Para poder afirmar con total certeza en el instante n que el proceso ha visitado el conjunto A por última vez, estaríamos obligados a conocer de antemano toda la trayectoria infinita futura del sistema $(X_{n+1}, X_{n+2}, \dots)$ para certificar que el proceso jamás regresará a A. Al depender intrínsecamente de información del futuro, viola la definición fundamental.

Teorema 3.6.8 (Strong Markov Property). *Sea $(X_n)_{n\geqslant 0}$ 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 $(X_n)_{n\geqslant 0}$.*

Entonces, condicionado al suceso $\{T < \infty\}$ y $\{X_T = i\}$, el proceso reiniciado $(X_{T+n})_{n\geqslant 0}$ 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 $(X_{T+n})_{n\geqslant 0}$ es estadísticamente independiente de la trayectoria histórica pasada $X_0, X_1, \dots, X_T$.

Demostración. No es relevante y es muy complicada.

□

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.

Observación. In continuous time, things can be trickier.

Definición 3.6.23 (Estados Recurrentes y Transitorios). Sea $(X_n)_{n\geqslant 0}$ 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:

$$T_i(\omega) = \text{ínf}\{n \geqslant 1 : X_n(\omega) = i\}, \qquad \text{bajo el convenio } \text{ínf}\,\emptyset = \infty$$

Definición 3.6.24 (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:

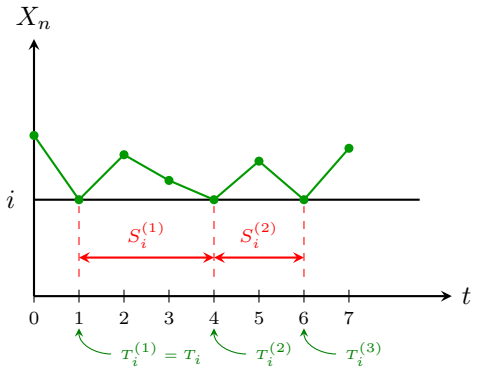
$$\begin{aligned} &T_i^{(0)}(\omega) = 0 \\ &T_i^{(1)}(\omega) = T_i(\omega) \\ &\text{Para todo } r \geqslant 1: \quad T_i^{(r+1)}(\omega) = \text{ínf}\{n > T_i^{(r)} : X_n(\omega) = i\} \end{aligned}$$

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:

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

A partir de estas definiciones temporales de trayectoria, podemos concluir la siguiente distinción intuitiva fundamental para el análisis del proceso:

- $T_i^{(n)}$ representa el **tiempo absoluto** (el valor del reloj global) en el que se efectúa la n-ésima visita al estado i.
- $S_i^{(n)}$ representa el **intervalo relativo** (el número de pasos netos gastados) transcurrido entre la visita $n-1$ y la visita n.



Lema 3.6.9. *Para todo índice de visita $r \geqslant 2$, condicionado a la ocurrencia en tiempo finito de la visita previa ($T_i^{(r-1)} < \infty$), el intervalo de tiempo relativo $S_i^{(r)}$ es estadísticamente* ***independiente*** *de la historia pasada del proceso $\{X_m : m \leqslant T_i^{(r-1)}\}$. Asimismo, su distribución satisface la identidad de homogeneidad:*

$$P\big(S_i^{(r)} = n \mid T_i^{(r-1)} < \infty\big) = P_i(T_i = n)$$

Demostración. □

Observación. Para entender que significa

$$P_i(T_i = n)$$

recordemos que la notación corta $P_i(\cdot)$ significa que obligamos a la cadena a empezar en el paso cero en el estado i ($X_0 = i$), y T_i es el tiempo del primer paso a ese estado i. Esa probabilidad significa que una cadena completamente nueva de fábrica, que arranca hoy en el estado i, tarde exactamente n pasos en regresar al estado i por primera vez.

Definición 3.6.25 (Número de Visitas a un Estado). Sea $(X_n)_{n\geqslant 0}$ 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:

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

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 3.6.26 (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 $T_i = \text{ínf}\{n \geqslant 1 : X_n = i\}$ representa el tiempo de primer paso al estado i.

Lema 3.6.10. *Para todo índice entero de visitas $r = 0, 1, 2, \ldots$, 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.

Demostración. □

3.7. Estados Recurrentes y Transitorios

Teorema 3.7.11 (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:

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

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:

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

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

$$E(X) = \sum_{k=1}^{\infty} k \cdot P(X = k) \implies E(X) \overset{(**)}{=} \sum_{k=1}^{\infty} \left(\sum_{r=0}^{k-1} 1 \right) P(X = k) = \sum_{k=1}^{\infty} \sum_{r=0}^{k-1} P(X = k) \implies$$

$$\implies E(X) = \sum_{r=0}^{\infty} \sum_{k=r+1}^{\infty} P(X = k) = \sum_{r=0}^{\infty} P(X > r)$$

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:

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

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 3.7.12. *Para todo estado $i \in S$ de una cadena de Markov, se verifican las siguientes equivalencias fundamentales sobre el horizonte temporal finito:*

1. *Un estado i es* ***recurrente*** *si y solo si $P_i(X_n = i \text{ para algún } n \geqslant 1) = 1$.*
2. *Un estado i es* ***transitorio*** *si y solo si $P_i(X_n = i \text{ para algún } n \geqslant 1) < 1$.*

Demostración. □

Teorema 3.7.13 (Propiedad de Solidaridad). *La recurrencia y la transitoriedad son propiedades de clase. Es decir, dada una clase de comunicación C, o bien todos los estados pertenecientes a C son transitorios, o bien todos los estados en C son recurrentes.*

Demostración. □

Teorema 3.7.14. *Toda clase de comunicación recurrente es matemáticamente cerrada.*

Demostración. □

Teorema 3.7.15. *Toda clase de comunicación que sea cerrada y finita es recurrente.*

Demostración. □

Demostración Alternativa Analítica. □

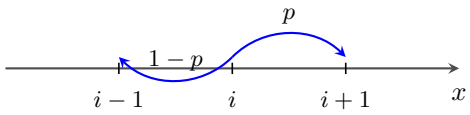
Corolario 3.7.15.1. *Si $(X_n)_{n\geqslant 0}$ es una cadena de Markov que opera sobre un espacio de estados de carácter finito ($|S| < \infty$), entonces se satisfacen de forma exacta las siguientes equivalencias estructurales fundamentales:*

1. *Una clase de comunicación es* ***cerrada*** *si y solo si es* ***recurrente****.*
2. *Una clase de comunicación* ***no es cerrada*** *si y solo si es* ***transitoria****.*

Ejemplo (Grafo Dinámico de la Marcha Aleatoria en $\mathbb{Z}$). Considere una marcha aleatoria simple que evoluciona sobre el conjunto infinito de los números enteros. Las probabilidades locales de transición asociadas a un estado genérico $i \in \mathbb{Z}$ vienen dadas, para un parámetro fijo $p \in (0, 1)$, por el siguiente comportamiento dinámico:

$$P_{i,i+1} = p, \qquad P_{i,i-1} = 1 - p$$

La representación esquemática del balance local de flujos de probabilidad de salto hacia los estados inmediatamente adyacentes se puede visualizar mediante el siguiente grafo unidimensional regulado en TikZ:



Teorema 3.7.16. *Sea $(Y_n)_{n\geqslant 0}$ una marcha aleatoria simple en $\mathbb{Z}$ con probabilidad de salto a la derecha dada por p. Se verifica la siguiente dicotomía fundamental para el espacio de estados:*

1. *Si $p \neq \frac{1}{2}$, entonces cada estado del espacio es* ***transitorio****.*
2. *Si $p = \frac{1}{2}$, entonces cada estado del espacio es* ***recurrente****.*

Demostración. □

Observación (Efecto de la Dimensión en la Recurrencia). Como conclusión estructural de este análisis de marchas aleatorias, se desprende que cualquier marcha sesgada (*biased*) es siempre transitoria, mientras que la marcha no sesgada o simétrica (*unbiased*) conserva su carácter recurrente sobre los retículos de dimensiones $d = 1$ y $d = 2$. No obstante, en virtud del Teorema de Polya, para cualquier retículo cúbico hiperbólico de dimensión tres o superior ($d \geqslant 3$), la marcha aleatoria simétrica se vuelve estrictamente **transitoria**.

Definición 3.7.27 (Variables de Ocupación antes del Retorno). Sea $(X_n)_{n\geqslant 0}$ 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:

$$T_k = \text{ínf}\{n \geqslant 1 : X_n = k\} \in \mathbb{N} \cup \{\infty\}$$

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 3.7.17 (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 $1 \leqslant m \leqslant T_k - 1$, por la propia definición matemática del tiempo de primer retorno ($T_k = \text{ínf}\{n \geqslant 1 : X_n = k\}$), 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 $\{m \leqslant T_k\}$ y explotando la linealidad del operador valor esperado, desarrollamos:

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

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

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

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

$$P_k(X_m = j,\, X_{m-1} = i,\, T_k \geqslant m) \overset{(*)}{=} P_k(X_m = j \mid X_{m-1} = i,\, T_k \geqslant m)\, P_k(X_{m-1} = i,\, T_k \geqslant m) = \\ \overset{(**)}{=} P_k(X_m = j \mid X_{m-1} = i)\, P_k(X_{m-1} = i,\, T_k \geqslant m)$$

y aplicando la homogeneidad temporal tenemos

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

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

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

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

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

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 $m, n \geqslant 0$ 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:

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

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

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

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:

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

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:

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

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 3.7.18 (Teorema del núcleo fundamental de medida invariante). *Sea $(X_n)_{n \geqslant 0}$ 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:*

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

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.

Demostración. □

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 3.7.18.1. *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:

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

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, $n \to \infty$).

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 nombrabamos 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 $(X_n)_{n \geqslant 0}$ 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:

$$T_i = \text{ínf}\{m \geqslant 1 : X_m = i\} \in \{1, 2, \dots\} \cup \{\infty\}$$

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
$$m_i = \sum_{n=1}^{\infty} n \cdot P_i(T_i = n) + \underbrace{\infty \cdot P_i(T_i = \infty)}_{\text{Miramos este término}} = (\text{Algo positivo}) + \infty(1 - h_i) \overset{(*)}{=} \infty$$
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 3.8.28 (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 3.8.29 (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:
$$m_i = E_i[T_i] = \infty \quad \text{e} \quad i \text{ es recurrente}$$

Teorema 3.8.19. *Sea $(X_n)_{n \geqslant 0}$ una cadena de Markov irreducible definida sobre un espacio de estados S. Entonces, las siguientes tres afirmaciones de estructura asintótica son matemáticamente equivalentes:*

1. *Al menos un estado del espacio S es positivo recurrente.*
2. *Todos los estados pertenecientes al espacio S son positivo recurrentes.*
3. *La cadena admite una única medida de probabilidad invariante $\lambda = (\lambda_i)_{i \in S}$.*

Si se verifica el cumplimiento de estas afirmaciones equivalentes, la masa de probabilidad de cada estado coincide de forma exacta con el inverso multiplicativo de su tiempo medio de primer retorno:
$$m_i = \frac{1}{\lambda_i} \qquad \forall i \in S \qquad (*)$$

Demostración. □

Observación. Toda cadena de Markov que sea simultáneamente irreducible y finita es de carácter **positivo recurrente**. Esto se deduce de manera directa puesto que, al ser cerrada y finita, posee una medida de probabilidad invariante que satisface de forma exacta la tercera afirmación equivalente del teorema fundamental de equivalencia.

Observación. Considere un paseo aleatorio simétrico estructurado sobre el conjunto infinito de los números enteros Z. Por criterios constructivos de sus trayectorias, este proceso es **irreducible** y **recurrente**.

Bajo estas hipótesis, el sistema admite una única medida invariante salvo por un factor de escala constante multiplicativo. Como ejemplo analítico, la medida que asigna de forma uniforme el valor 1 a cada estado $i \in Z$ satisface las ecuaciones de balance lineal ($\lambda = \lambda P$). Sin embargo, dado que el soporte es infinito enumerable, la masa total acumulada diverge:

$$\sum_{i \in Z} \lambda_i = \sum_{i \in Z} 1 = \infty$$

Al ser imposible normalizar la medida para construir una distribución de probabilidad legal sobre el espacio, se viola la condición del teorema de equivalencia. Por consiguiente, se concluye formalmente que todos los estados de un paseo aleatorio simétrico en Z son **nulo recurrentes**.

3.8.1. Convergencia a la Medida de Probabilidad Invariante

El estudio del comportamiento asintótico y la distribución de equilibrio a largo plazo del sistema se articula a través de dos herramientas fundamentales del análisis estocástico:

1) La **Ley Fuerte de los Grandes Números (LFGN)** adaptada para estructuras con dependencia de Markov, la cual gobierna la convergencia de los promedios temporales de ocupación de las trayectorias.

2) El **Teorema del Límite** para las probabilidades de transición en múltiples pasos, el cual describe el comportamiento asintótico de las celdas marginales de la matriz:

$$\lim_{m \to \infty} P^{(m)}(x, y) = \pi(y)$$

Teorema 3.8.20 (Ley Fuerte de los Grandes Números - Caso Clásico). *Sea $\{Y_i\}_{i \geqslant 1}$ 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:*

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

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

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

Observación. El cumplimiento de la tesis de la Ley Fuerte de los Grandes Números se extiende y conserva su validez analítica en escenarios donde el primer momento teórico diverge hacia el infinito, es decir, cuando $E[Y_1] = \infty$.

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

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

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

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

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

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

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

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

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

Esquema de implicaciones fundamentales de repaso:
Casi Seguro $\implies$ En Probabilidad $\implies$ En Distribución

Teorema 3.8.21 (Teorema Ergódico de Ocupación). *Considere una cadena de Markov irreducible $(X_n)_{n\geqslant 0}$ 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.):*

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

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:*

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

Demostración. □

Corolario 3.8.21.1. *Consideremos una cadena de Markov positivo recurrente $(X_n)_{n\geqslant 0}$ dotada de una probabilidad invariante $\pi = (\pi_i)_{i\in S}$, y sea $f : S \to R$ 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:*

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

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 çolaïnfinita 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:
$$\leq \sum_{i \in \mathcal{J}} \left| \frac{V_i(n)}{n} - \pi_i \right| + \sum_{i \notin \mathcal{J}} \frac{V_i(n)}{n} + \sum_{i \notin \mathcal{J}} \pi_i$$
Para eliminar la dependencia muestral de la variable aleatoria $V_i(n)$ en el término central de la cola, utilizamos el siguiente "truco.algebraico consistente en sumar y restar la medida π_i:
$$\sum_{i \notin \mathcal{J}} \frac{V_i(n)}{n} = \sum_{i \notin \mathcal{J}} \left(\frac{V_i(n)}{n} - \pi_i + \pi_i \right) \leq \sum_{i \notin \mathcal{J}} \left| \frac{V_i(n)}{n} - \pi_i \right| + \sum_{i \notin \mathcal{J}} \pi_i$$

Extendiendo de forma segura el rango del sumatorio de este valor absoluto intermedio a todo el conjunto indexado $\mathcal{J}$ para agruparlo con el primer término, consolidamos la cota superior fundamental de los apuntes:

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

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:

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

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 3.8.30 (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.

Ejercicio 3.8.1. Demuestre formalmente que un estado $i \in S$ es **aperiódico** si y solo si el máximo común divisor del conjunto de todos los instantes de retorno posibles es estrictamente igual a la unidad:

$$\text{mcd}\{m \geqslant 1 : P_{ii}^{(m)} > 0\} = 1$$

Lema 3.8.22 (Solidaridad de la Aperiodicidad). *Supongamos que la cadena de Markov $(X_n)_{n\geqslant 0}$ 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:

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

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 3.8.23 (Teorema de Convergencia Estacionaria). *Considere una cadena de Markov $(X_n)_{n\geqslant 0}$ 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:

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

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 geométricamente.

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:

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

esto lo vemos fácilmente con inducción como sigue

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

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

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

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

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

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

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

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:

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

Esta variable mapea el primer çhoque.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 $\widetilde{S} = S \times S$.

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:

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

- **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:

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

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 desarollo

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

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

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

□

Observación. El resultado del límite encierra una conclusión matemática profunda, la **pérdida de memoria histórica del proceso**. La ecuación afirma que, cuando el horizonte temporal tiende a infinito ($m \to \infty$), la probabilidad de encontrar a la cadena en el estado de destino j coincide exactamente con su peso estacionario λ_j, siendo completamente independiente del estado físico de partida i. El tiempo borra la huella de las condiciones de contorno iniciales.

Observación. El teorema se aplica de forma automática a **cualquier cadena de Markov que sea irreducible y opere sobre un espacio de estados finito** ($|S| < \infty$).

3.9. Ising Model

El soporte físico del modelo se formaliza a través de un grafo matemático no dirigido $G = (V, E)$, donde:

- V: Conjunto de vértices o nodos, que representan las posiciones fijas de los átomos en el retículo cristalino.
- E: Conjunto de aristas o enlaces, que definen las interacciones energéticas directas entre pares de átomos.

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Matriz de Adyacencia (A)

Variables de Espín y Acoplamientos

- **Variables de Estado Locales (S_i):** Para cada vértice $i \in V$, se asigna una variable aleatoria discreta binaria que modela el momento magnético intrínseco (espín):

$$S_i = \begin{cases} +1 & \text{(Espín orientado hacia arriba / Up)} \\ -1 & \text{(Espín orientado hacia abajo / Down)} \end{cases} \qquad \forall\, i \in V$$

- **Constantes de Acoplamiento (J_{ij}):** Asociadas a cada arista $(i, j) \in E$, cuantifican la energía de interacción elemental entre los espines involucrados. Se denominan acoplamientos magnéticos.

Hamiltoniano del Sistema (H)

Definición 3.9.31. El **hamiltoniano** representa el operador matemático que computa la energía total de una configuración dada del sistema $\underline{S} = (S_1, S_2, \ldots, S_N)$:

$$H(\underline{S}) = \sum_{\langle i,j \rangle} J_{ij} S_i S_j$$

Notación. El símbolo $\langle i, j \rangle$ restringe estrictamente el sumatorio a parejas de primeros vecinos, es decir, únicamente a aquellos pares de nodos conectados por una arista directa en el conjunto E ($A_{ij} = 1$).

Ejemplo. Estudiaremos el hamiltoniano de la Figura 3.5, que representa un grafo de 5 nodos con 6 aristas. La matriz de adyacencia A correspondiente se muestra en la sección previa.

Basándonos en las conexiones declaradas en la matriz de adyacencia A, el Hamiltoniano se expande linealmente en sus 6 términos activos de interacción:

$$H(\underline{S}) = J_{12}S_1S_2 + J_{13}S_1S_3 + J_{23}S_2S_3 + J_{24}S_2S_4 + J_{34}S_3S_4 + J_{45}S_4S_5$$

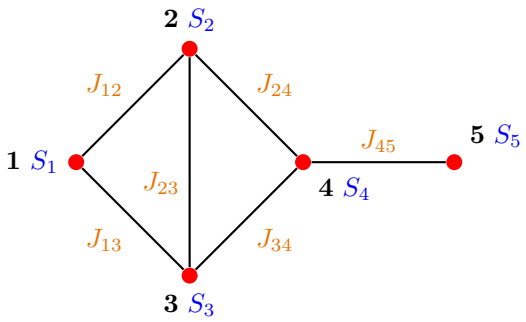


Figura 3.5: Grafo del Modelo de Ising correspondiente a la imagen.

Espacio de Estados y Variantes del Modelo

Supongamos el caso particular donde todos los espines adoptan una orientación unánime hacia abajo, correspondiente al vector de estado $\underline{S} = (-1, -1, -1, -1, -1)$. La energía de este estado se reduce a la suma directa de sus acoplamientos. Por ejemplo, para el ejemplo antes visto tendríamos

$$H(-1,-1,-1,-1,-1) = J_{12}(-1)(-1) + J_{13}(-1)(-1) + \cdots = J_{12} + J_{13} + J_{23} + J_{24} + J_{34} + J_{45}$$

Si imponemos una condición homogénea donde $J_{ij} = -J$ (siendo $J > 0$ un valor real fijo positivo) para todo par conectado, se desprenden las siguientes categorizaciones

1. **Modelo de Ising Ferromagnético Tradicional:** Si los acoplamientos son constantes e idénticos ($J_{ij} = -J$), el signo negativo favorece energéticamente el alineamiento paralelo de los espines vecinos:
$$H(\underline{S}) = -J \sum_{\langle i,j \rangle} S_i S_j$$

2. **Vidrio de Espín de Ising (*Spin Glass*):** Si las constantes de acoplamiento J_{ij} dejan de ser fijas y se modelan como variables aleatorias independientes (ej. distribuciones Gaussianas), el sistema introduce desorden y frustración magnética.

Análisis Termodinámico y Unidades Energéticas

La probabilidad de que el sistema exhiba una configuración geométrica concreta $\underline{S}$ en el equilibrio térmico viene dictada por la micro-medida canónica de Gibbs-Boltzmann:

$$P(\underline{S}) \propto e^{-\beta H(\underline{S})}, \qquad \text{donde } \beta = \frac{1}{k_B T} \text{ (parámetro de temperatura inversa)}$$

- k_B: Constante de Boltzmann, cuyo valor experimental es $1{,}38 \cdot 10^{-23}$ Joules $\cdot$ Kelvin^{-1}.
- T: Temperatura absoluta escalada en Kelvin ($T = t(^\circ\text{C}) + 273{,}15$).

Ejemplo. Ejemplo de cálculo para Temperatura Ambiente ($T = 300$ K $\approx 27^\circ$C). La escala fundamental de energía térmica fluctuante ($\mathcal{E}$) se calcula como:

$$\mathcal{E} = k_B T = (1{,}38 \cdot 10^{-23}\ \text{J} \cdot \text{K}^{-1}) \cdot (300\ \text{K}) = 4{,}14 \cdot 10^{-21}\ \text{Joules}$$

Haciendo uso de la equivalencia fundamental del electrón-voltio ($1\ \text{J} \approx 6{,}242 \cdot 10^{18}$ eV), transformamos la escala de energía:

$$\mathcal{E} \approx 4{,}14 \cdot 10^{-21} \cdot (6{,}242 \cdot 10^{18})\ \text{eV} \approx 0{,}0258\ \text{eV} = 25{,}8\ \text{meV}$$

Notación (Comportamiento Asintótico de la Temperatura). La interacción entre la energía del Hamiltoniano y la agitación térmica del entorno viene gobernada de forma exclusiva por el producto adimensional βJ. El comportamiento del sistema en los extremos térmicos dicta la física de las transiciones de fase:

- **Límite de Baja Temperatura ($T \to 0^+$):**

$$T \to 0^+ \implies \beta = \frac{1}{k_B T} \to \infty \implies \beta J \to \infty$$

En este régimen, el factor de Boltzmann $e^{-\beta H(\underline{S})}$ penaliza drásticamente cualquier estado excitado de energía alta. La medida de probabilidad colapsa casi con certeza (*c.s.*) sobre las configuraciones que minimizan el Hamiltoniano (estados fundamentales ordenados o ferromagnéticos, donde todos los espines apuntan en la misma dirección).

- **Límite de Alta Temperatura ($T \to \infty$):**

$$T \to \infty \implies \beta = \frac{1}{k_B T} \to 0 \implies \beta J \to 0$$

Al anularse el argumento del exponente térmico, el factor de Boltzmann tiende a la unidad ($e^0 = 1$) para las 2^n configuraciones del espacio de estados $\mathcal{S}$. La medida de probabilidad se vuelve uniforme sobre todo el soporte discreto:

$$P(\underline{S}) \to \frac{1}{2^n} \quad \forall\, \underline{S} \in \mathcal{S}$$

El sistema alcanza el desorden absoluto (régimen paramagnético), donde la entropía satura el sistema y destruye cualquier correlación magnética entre primeros vecinos.

Función de Partición y Deducción de la Energía Interna

Para normalizar la medida de probabilidad de Gibbs sobre todo el espacio de configuraciones discretas, se define la **Función de Partición** (*Zustandssumme*):

$$Z(\beta) = \sum_{\underline{S}} e^{-\beta H(\underline{S})} \qquad \implies \qquad P(\underline{S}) = \frac{e^{-\beta H(\underline{S})}}{Z(\beta)}$$

Efectuamos la derivada analítica de la función de partición respecto a la variable macroscópica β:

$$\frac{dZ(\beta)}{d\beta} = \frac{d}{d\beta} \sum_{\underline{S}} e^{-\beta H(\underline{S})} = \sum_{\underline{S}} \frac{d}{d\beta} e^{-\beta H(\underline{S})} = \sum_{\underline{S}} -H(\underline{S}) e^{-\beta H(\underline{S})} = -\sum_{\underline{S}} H(\underline{S}) e^{-\beta H(\underline{S})}$$

Por definición de la Mecánica Estadística, el valor esperado de la energía del sistema constituye la **Energía Interna Termodinámica (U)**:

$$U = \langle H \rangle = \sum_{\underline{S}} H(\underline{S}) \cdot P(\underline{S}) = \frac{\sum_{\underline{S}} H(\underline{S}) e^{-\beta H(\underline{S})}}{Z(\beta)}$$

Aislamos el término del sumatorio algebraico multiplicando por el signo opuesto, lo que nos permite conectar la esperanza con la derivada deducida previamente:

$$U = -\frac{1}{Z(\beta)} \left(-\sum_{\underline{S}} H(\underline{S}) e^{-\beta H(\underline{S})} \right) = -\frac{1}{Z(\beta)} \frac{dZ(\beta)}{d\beta}$$

Aplicando la regla de la cadena para la derivada de un logaritmo natural ($\frac{d}{dx} \ln(f(x)) = \frac{f'(x)}{f(x)}$), compactamos la expresión analítica final para la cantidad termodinámica fundamental U:

$$\boxed{U = \langle H \rangle = -\frac{d \ln Z(\beta)}{d\beta}}$$

Potenciales Termodinámicos e Identidades Fundamentales

En el ensamble canónico (sistemas a volumen y temperatura constante en contacto con un baño térmico), el estado de equilibrio no se alcanza minimizando la energía interna U, sino minimizando el potencial de Helmholtz.

Definición 3.9.32 (Energía Libre de Helmholtz (F)). La **Energía Libre de Helmholtz** representa la fracción de energía interna del sistema que está disponible para realizar trabajo útil tras descontar el coste del desorden térmico. Se define analíticamente a partir de la función de partición mediante la identidad:

$$F(\beta) = -\frac{1}{\beta} \ln Z(\beta) = -k_B T \ln Z$$

Proposición 3.9.24 (Transformada de Legendre y Entropía). *La conexión macroscópica clásica entre la Energía Interna ($\mathcal{E} \equiv U = \langle H \rangle$), la temperatura y la Entropía (S) se formaliza a través de la relación fundamental:*

$$F = \mathcal{E} - TS$$

Esta ecuación es la Transformada de Legendre que permite transicionar del colectivo microcanónico (donde domina el recuento de microestados W vía $S = k_B \ln W$) al colectivo canónico regido por $Z(\beta)$.

Operador de Magnetización (M) y Simetría del Sistema

Para caracterizar cuantitativamente el orden magnético global de una configuración, introducimos la observable macroscópica **magnetización total** ($m(\underline{s})$).

Definición 3.9.33. La **magnetización total** de una configuración $\underline{s}$ se define como la suma algebraica de los espines individuales:

$$m(\underline{s}) = \sum_{i=1}^{N} S_i$$

Al calcular el valor esperado o primer momento estadístico de la magnetización ($E[M]$) promediado bajo la medida canónica de Gibbs, el sistema arroja una anulación simétrica absoluta debido a la isotropía espacial del Hamiltoniano:

$$E(M) = \sum_{\underline{s} \in \mathcal{S}} \frac{m(\underline{s}) e^{-\beta H(\underline{s}) Z}}{Z(\beta)} = 0$$

Observación. A pesar de que el promedio simétrico es nulo ($E[M] = 0$), el valor esperado del módulo o magnitud absoluta de la magnetización es estrictamente positivo, $\mathbb{E}[|M|] \neq 0$. Esto denota que el cristal posee momentos magnéticos locales reales, pero carece de una dirección preferencial espontánea global en el ensamble canónico puro sin campo externo.

Ejemplo. Consideremos un sistema ferromagnético simple constituido por un grafo cerrado de tres vértices interconectados en forma de triángulo ($N = 3$), visto en la Figura 3.6. El Hamiltoniano macroscópico del sistema, bajo constantes de acoplamiento homogéneas $J > 0$, responde a la sumatoria sobre primeros vecinos:

$$H(\underline{S}) = -J(S_1 S_2 + S_2 S_3 + S_1 S_3)$$

El espacio de configuraciones posee un tamaño discreto de $2^3 = 8$ estados posibles. Al evaluar los espines, la energía se discretiza en dos únicos autoestados:

- Energía $+J$: Asociada a 6 configuraciones frustradas o no alineadas.
- Energía $-3J$: Asociada a 2 estados perfectamente alineados.

En la Tabla 3.7 podemos ver la energía para todo el espacio de estados de S.
Tomando como base las degeneraciones obtenidas en el recuento microcanónico de la tabla (2 estados estables y 6 degenerados), la **Función de Partición** $Z(\beta)$ se condensa analíticamente como:

$$Z(\beta) = \sum_{\underline{s} \in \mathcal{S}} e^{-\beta H(\underline{s})} = 2e^{-\beta(-3J)} + 6e^{-\beta(+J)} = 2e^{3\beta J} + 6e^{-\beta J}$$

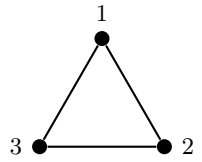


Figura 3.6: Grafo triangular para el Modelo de Ising.

S_1	S_2	S_3	Producto de Pares Varianza	Suma del Paréntesis	Energía $H(\underline{S})$
-1	-1	-1	$(+1)+(+1)+(+1)$	$+3$	$-3J$
$+1$	-1	-1	$(-1)+(+1)+(-1)$	-1	$+J$
-1	$+1$	-1	$(-1)+(-1)+(+1)$	-1	$+J$
-1	-1	$+1$	$(+1)+(-1)+(-1)$	-1	$+J$
$+1$	$+1$	-1	$(+1)+(-1)+(-1)$	-1	$+J$
$+1$	-1	$+1$	$(-1)+(-1)+(+1)$	-1	$+J$
-1	$+1$	$+1$	$(-1)+(+1)+(-1)$	-1	$+J$
$+1$	$+1$	$+1$	$(+1)+(+1)+(+1)$	$+3$	$-3J$

Figura 3.7: Tabla de Energías para el Modelo de Ising Triangular.

A partir de la función de partición canónica, derivamos la **Energía Libre de Helmholtz (F)** del cristal:

$$F = -\frac{1}{\beta}\log Z(\beta) = -\frac{1}{\beta}\log\left(2e^{3\beta J} + 6e^{-\beta J}\right)$$

Calculamos el valor esperado de la energía o **Energía Interna Termodinámica ($E \equiv \langle H \rangle$)** ejecutando el operador diferencial canónico $-\frac{1}{Z}\frac{dZ}{d\beta}$:

$$E = -\frac{1}{Z(\beta)}\frac{dZ(\beta)}{d\beta} = -\frac{1}{Z(\beta)}\left(2\cdot 3J\cdot e^{3\beta J} + 6\cdot(-J)\cdot e^{-\beta J}\right) = -\frac{6Je^{3\beta J} - 6Je^{-\beta J}}{2e^{3\beta J} + 6e^{-\beta J}}$$

Factorizando y simplificando el signo negativo del numerador, el promedio de energía se compacta como:

$$E = 6J\,\frac{e^{-\beta J} - e^{3\beta J}}{6e^{-\beta J} + 2e^{3\beta J}}$$

Evaluemos el comportamiento asintótico de la energía interna bajo los dos extremos de la escala térmica:

$$\lim_{\beta J\to 0} E = 6J\,\frac{1-1}{6+2} = 0 \qquad \text{[Límite de Alta Temperatura } T\to\infty\text{]}$$

$$\lim_{\beta J\to+\infty} E = \lim_{\beta J\to+\infty} 6J\,\frac{\cancel{e^{-\beta J}} - e^{3\beta J}}{\cancel{6e^{-\beta J}} + 2e^{3\beta J}} = 6J\left(\frac{-e^{3\beta J}}{2e^{3\beta J}}\right) = -3J \qquad \left[\text{Límite de Cero Absoluto } T\to 0^{+}\right]$$

Haciendo uso de la ecuación puente de la entropía deducida por la transformada de Legendre ($\frac{S}{k_B} = \beta(E-F)$), extendemos la función funcional completa del desorden:

$$\frac{S}{k_B} = \beta\left[6J\,\frac{e^{-\beta J} - e^{3\beta J}}{6e^{-\beta J} + 2e^{3\beta J}} + \frac{1}{\beta}\log\left(2e^{3\beta J} + 6e^{-\beta J}\right)\right] = 6\beta J\,\frac{e^{-\beta J} - e^{3\beta J}}{6e^{-\beta J} + 2e^{3\beta J}} + \log\left(2e^{3\beta J} + 6e^{-\beta J}\right)$$

Nota Analítica: Al evaluar de forma directa el límite de baja temperatura ($\beta J\to\infty$), la estructura colapsa en una indeterminación de tipo $[-\infty+\infty]$. Procedemos a su resolución aplicando el comportamiento asintótico de los términos exponenciales dominantes:

1. **Aproximación del término logarítmico (F):**

$$\log\left(2e^{3\beta J} + 6\cancelto{0}{e^{-\beta J}}\right) \approx \log\left(2e^{3\beta J}\right) = \log(2) + \log\left(e^{3\beta J}\right) = \log(2) + 3\beta J$$

2. **Aproximación del término fraccionario (E):**

$$6\beta J \frac{\cancelto{0}{e^{-\beta J}} - e^{3\beta J}}{\cancelto{0}{6e^{-\beta J}} + 2e^{3\beta J}} \approx 6\beta J \left(\frac{-e^{3\beta J}}{2e^{3\beta J}} \right) = -3\beta J$$

Sustituyendo ambas simplificaciones asintóticas en la ecuación de la entropía, los infinitos lineales de la temperatura inversa se cancelan de forma exacta, resolviendo la indeterminación:

$$\lim_{\beta J \to \infty} \frac{S}{k_B} = -3\beta J + \log(2) + 3\beta J = \boxed{\log(2)}$$

El resumen de los límites de la entropía queda como

$$\lim_{\beta J \to 0} \frac{S}{k_B} = \log(8) \quad \text{(Saturación de desorden: Equiprobabilidad en los 8 microestados)}$$

$$\lim_{\beta J \to \infty} \frac{S}{k_B} = \log(2) \quad \text{(Orden latente: El sistema se congela en los 2 estados fundamentales)}$$

3.10. Reversible Markov Chain

Recordemos la definición de cadena de Márkov reversible.

Definición 3.10.34 (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 ($\pi = \pi P$).

3.10.1. Acoplamiento de la Dinámica de Markov al Modelo de Ising

Para resolver el Modelo de Ising mediante simulación, definimos una cadena de Markov donde cada estado en un instante temporal t es una configuración microscópica completa de espines espaciales $\mathcal{S}(t) = \underline{s}$. La probabilidad de transición condicional para transicionar hacia un estado de prueba consecutivo $\underline{s}'$ se denota como:

$$P(\mathcal{S}(t+1) = \underline{s}' \mid \mathcal{S}(t) = \underline{s}) = P(\underline{s}' \mid \underline{s}) = p_{\underline{s}\,\underline{s}'}$$

Exigimos que este proceso estocástico artificial satisfaga el balance detallado con respecto a la medida canónica invariante de Gibbs-Boltzmann, denotada aquí como $\mu(\underline{s})$:

$$P(\underline{s}' \mid \underline{s})\mu(\underline{s}) = P(\underline{s} \mid \underline{s}')\mu(\underline{s}') \quad \forall\, (\underline{s}, \underline{s}') \in \mathcal{S} \times \mathcal{S}$$

Sustituyendo explícitamente la estructura analítica de la distribución de Gibbs de tus apuntes en ambos miembros de la ecuación, cancelamos algebraicamente el factor de normalización común correspondiente a la función de partición $Z(\beta)$:

$$P(\underline{s}' \mid \underline{s}) \frac{e^{-\beta H(\underline{s})}}{\cancel{Z(\beta)}} = P(\underline{s} \mid \underline{s}') \frac{e^{-\beta H(\underline{s}')}}{\cancel{Z(\beta)}}$$

Agrupando los términos de transiciones probabilísticas en el miembro izquierdo y las funciones exponenciales de energía en el derecho, aislamos el cociente de probabilidades de transición:

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

Donde $\Delta H = H(\underline{s}') - H(\underline{s})$ define el balance diferencial de energía macroscópica asociado al salto microscópico cristalino.

3.10.2. Metropolis Algorithm

Para garantizar de manera unívoca el cumplimiento de la relación cinemática

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

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

Teorema 3.10.25 (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 ($S_i \to -S_i$) 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 sobrecoste 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 Continuous Time Random Walks

Definición 1. Definimos un **paseo aleatorio discreto (P.A.)** (random walk) mediante una sucesión $\{X_i\}_{i=1}^{\infty}$ de variables aleatorias independientes e idénticamente distribuidas (i.i.d.). Estas variables cuantifican la magnitud y dirección de cada salto espacial individual:

- **Soporte algebraico:** las variables X_i toman valores en $\mathbb{R}$, es decir, no están restringidas a valores no negativos, permitiendo desplazamientos bidireccionales en el espacio.
- **Naturaleza analítica:** las variables X_i pueden ser continuas, caracterizadas mediante funciones de densidad de probabilidad absolutamente continuas con respecto a la medida de Lebesgue sobre la recta real.

Observación. Se dice random walk discreto, ya que el proceso se define a través de una sucesión discreta de saltos espaciales, aunque cada salto individual pueda tomar valores continuos. En contraste, un proceso de difusión continuo (como el Movimiento Browniano) se define mediante una trayectoria continua en el tiempo y el espacio.

A partir de dicha sucesión, la posición acumulada del proceso tras efectuar exactamente m saltos espaciales viene gobernada por la variable aleatoria suma Y_m

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

Modelado bajo Distribuciones Gaussianas

Supongamos el caso particular en el que los saltos espaciales individuales siguen una distribución normal o Gaussiana clásica, parametrizada por su media μ y su desviación estándar σ:

$$X_i \sim N(\mu, \sigma)$$

La función de densidad de probabilidad (f.d.p.) del primer salto elemental X_1 adopta la estructura analítica:

$$f_{X_1}(u) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\dfrac{(u-\mu)^2}{2\sigma^2}}$$

Debido a la propiedad de estabilidad de las variables Gaussianas independientes, la función de densidad de la posición acumulada Y_m tras m saltos se determina rigurosamente mediante el operador de convolución de orden m, denotado por $f_{X_1}^{*m}(u)$:

$$f_{Y_m}(u) = f_{X_1}^{*m}(u) = \frac{1}{\sqrt{2\pi m}\sigma}\, e^{-\dfrac{(u-m\mu)^2}{2m\sigma^2}}$$

Calcularemos ahora la **esperanza matemática** y la **varianza** de la variable aleatoria suma Y_m haciendo uso de la linealidad del operador de esperanza matemática y de la propiedad de aditividad

de la varianza para variables aleatorias independientes, los dos primeros momentos teóricos de la suma se expanden linealmente con el número de saltos m:

$$\mathbb{E}[Y_m] = \mathbb{E}\left[\sum_{i=1}^{m} X_i\right] = m\,\mathbb{E}[X_1] = m\mu$$

$$\mathrm{Var}[Y_m] = \mathrm{Var}\left[\sum_{i=1}^{m} X_i\right] = m\,\mathrm{Var}[X_1] = m\sigma^2$$

Observación. La identidad fundamental $f_{Y_m}(u) = f_{X_1}^{*m}(u)$ constituye la **fórmula general para la función de densidad de probabilidad de un solo punto** (one-point probability density function) del proceso suma Y_m. Esta relación de convolución iterada es un principio universal para la suma de variables independientes, independientemente de si la ley base es Gaussiana o no.

Para transformar el paseo aleatorio clásico en un proceso indexado en tiempo continuo (CTRW), es imperativo introducir una dimensión temporal estocástica que desacople el número de saltos del tiempo cronológico.

Definición 2. Sea $\{J_i\}_{i=1}^{\infty}$ una sucesión de variables aleatorias estrictamente positivas ($J_i > 0$) e idénticamente distribuidas (i.i.d.), caracterizadas por una función de densidad común $f_{J_1}(u)$. Estas variables modelan los **tiempos de espera** (waiting times) o retardos que experimenta la partícula inmóvil antes de ejecutar el siguiente salto espacial.

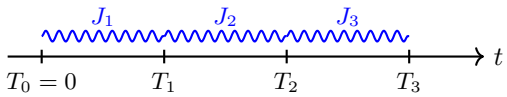
Observación. Lo que realmente denota cada variable aleatoria J_i es el tiempo de retención o confinamiento que la partícula permanece en su posición actual antes de efectuar el salto espacial X_i. En otras palabras, J_i representa el intervalo de tiempo transcurrido entre el $(i-1)$-ésimo y el i-ésimo salto del proceso.

Definición 3. El **instante temporal exacto del m-ésimo salto del proceso** viene determinado por la variable aleatoria de tiempo acumulado T_m:

$$T_m = \sum_{i=1}^{m} J_i \qquad \text{sujeto a la condición de contorno fija: } T_0 = 0 \quad \text{c.s. (casi seguro)}$$

La variable T_m representa analíticamente un **proceso de renovación** sobre la línea temporal continua, porque cada vez que la partícula ejecuta un salto, el cronómetro local de espera se destruye y se genera uno nuevo (J_{i+1}) con la misma distribución. Es como si el sistema "olvidara" el cansancio acumulado y volviera a nacer o a renovarse en cada parada. Los instantes $\{T_1, T_2, T_3, \dots\}$ se denominan científicamente épocas o tiempos de arribo de la renovación.

Podemos ver una representación gráfica de la evolución temporal del proceso de renovación en el siguiente diagrama, donde se ven reflejadas las naturalezas de ambas variables aleatorias J_i y T_m



Ejemplo. Sea J_1 una variable aleatoria continua que define el primer tiempo de espera o confinamiento de una partícula en un estado. Supongamos que se rige por una distribución exponencial de parámetro de tasa $\lambda > 0$:

$$J_1 \sim \mathrm{Exp}(\lambda) \implies f_{J_1}(u) = \lambda e^{-\lambda u}, \quad u \geq 0$$

Si consideramos una sucesión $\{J_i\}_{i=1}^{\infty}$ de variables aleatorias independientes e idénticamente distribuidas (i.i.d.) bajo esta misma ley, el instante exacto en el que se ejecuta el m-ésimo salto del proceso viene determinado por la variable aleatoria de tiempo acumulado

$$T_m = \sum_{i=1}^{m} J_i$$

Debido a la independencia de los sumandos, la función de densidad de probabilidad (f.d.p.) de T_m se calcula rigurosamente mediante la convolución analítica iterada de orden m de la densidad exponencial base, lo que da origen a la **Distribución de Erlang**:

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

La distribución de Erlang constituye un caso especial de la distribución Gamma de parámetros continuos cuando el parámetro de forma m se restringe al conjunto de los números enteros positivos ($m \in \mathbb{N}^+$), lo cual permite sustituir la función Gamma abstracta por el operador factorial clásico en el denominador

$$\Gamma(m) = \int_0^\infty t^{m-1} e^{-t}\, dt = (m-1)!$$

Paralelismo dinámico: Este esquema temporal es análogo al comportamiento de transición de las Cadenas de Markov (C.M.) convencionales. No obstante, se introduce un cambio de paradigma crítico: los tiempos de retención inter-estado J_i pasan a ser **variables aleatorias continuas**, mientras que en las Cadenas de Markov en tiempo discreto clásicas estaban confinados de forma estricta a distribuciones de soporte **discreto** (distribución geométrica).

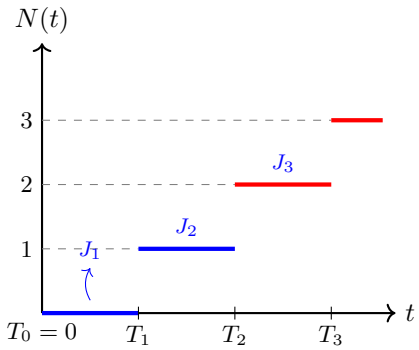
Definición 4. Mediante una operación de inversión lógica del paseo aleatorio temporal, definimos el **proceso de conteo** $N(t)$ indexado en tiempo continuo. Este operador estocástico de valor entero computa el número acumulado (o máximo) de eventos o saltos espaciales completados antes o exactamente en el hito temporal continuo $t \geq 0$:

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

Definición 5. Un **proceso de renovación** es una secuencia de eventos o hitos temporales $\{T_n\}_{n\geq 0}$ donde los tiempos transcurridos entre sucesos consecutivos, definidos como $J_i = T_i - T_{i-1}$ para $i \geq 1$, forman una sucesión de variables aleatorias que cumple rigurosamente las siguientes condiciones:

- Las variables aleatorias $J_1, J_2, J_3, \ldots$ son independientes e idénticamente distribuidas (i.i.d.).
- El soporte de su función de distribución común $F_J(u)$ está restringido a la semirrecta positiva no nula, es decir:
$$P(J_i > 0) = 1 \quad \forall\, i \in \mathbb{N}^+$$

Observación. El proceso de conteo continuo $(N(t))_{t\geq 0}$ que contabiliza cuántos de estos eventos de renovación han acontecido hasta el instante t se denomina, por extensión, el proceso de conteo asociado a dicha renovación.



Propiedades de las Trayectorias muestrales:
Por naturaleza analítica de los procesos de conteo acumulativos, las trayectorias de $N(t)$ satisfacen la condición matemática de tipo **càdlàg** (*continue à droite et limité à gauche*):

Las trayectorias son **continuas por la derecha**

y poseen **límites por la izquierda**.

Físicamente, esto garantiza que en el instante preciso en el que ocurre una renovación u ocurrencia ($t = T_m$), el valor de la variable de conteo absorbe instantáneamente el salto entero superior, reflejando el estado actual del cronómetro.

Definición 6 (Proceso Compuesto / CTRW). El **paseo aleatorio en tiempo continuo generalizado** se construye formalmente como un proceso compuesto $(Y(t))_{t\geq 0}$, subordinando la cadena

de saltos espaciales discretos Y_m al reloj estocástico gobernado por el proceso de conteo de renovaciones $N(t)$:

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

donde X_i denota la magnitud aleatoria continua del i-ésimo de desplazamiento espacial.

Observación. Las sucesiones discretas independientes de posiciones espaciales acumuladas $(Y_m)_{m\geq 0}$ y de épocas temporales de arribo de renovación $(T_m)_{m\geq 0}$ constituyen de forma aislada, por su diseño recursivo de adición, dos estructuras de Cadenas de Markov discretas en su índice de pasos m.

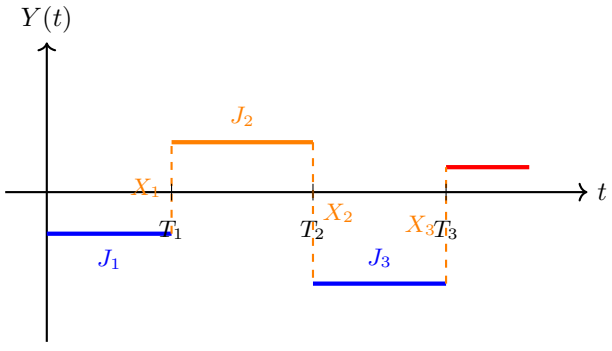
Proposición 1. *Si la cadena de posiciones en el espacio $(Y_m)_{m\geq 0}$ y la cadena de tiempos de renovación $(T_m)_{m\geq 0}$ son mutuamente independientes en el sentido estocástico (lo que se denota formalmente mediante la relación de ortogonalidad probabilística $(Y_m)_{m\geq 0} \perp\!\!\!\perp (T_m)_{m\geq 0}$), entonces el vector de estado bidimensional acoplado (T_m, Y_m) hereda y preserva de forma estricta la propiedad de Markov sobre su espacio muestral ordenado.*

Observación. A pesar del carácter markoviano del par coordenado indexado por pasos, el proceso final compuesto expandido en el tiempo continuo $(Y(t))_{t\geq 0}$ **no es un proceso de Markov en general**. Al exhibir tiempos de espera arbitrarios J_i, la probabilidad de transición futura depende explícitamente de la memoria del sistema, concretamente del tiempo que lleva la partícula esperando en su posición actual, lo que introduce una dependencia temporal no markoviana.

Esta nota indica que si entramos en el laboratorio y vemos una partícula inmóvil en la casilla $x = 5$. Si queremos calcular la probabilidad de que salte en el próximo segundo, necesitmos hacernos una pregunta crucial: ¿Cuánto tiempo lleva esa partícula durmiendo en esa casilla?

- Si la partícula acaba de aterrizar en la casilla hace un microsegundo, su temporizador J_i está entero, por lo que es muy poco probable que salte inmediatamente.
- Si la partícula lleva tres horas atrapada en esa misma casilla, su temporizador de espera J_i está a punto de agotarse, por lo que la probabilidad de que salte de forma inminente es altísima.

Aquí tenemos la ruptura de Markov, para calcular el futuro, ya no basta con saber el estado presente (la posición $x = 5$); necesitas conocer la "edad" o el "tiempo de residencia" que la partícula acumula en ese lugar. Podemos ver una representación de como avanzaría $Y(t)$ a lo largo del tiempo, con sus periodos de confinamiento y sus saltos espaciales, en el siguiente diagrama:



4.1. Espacio de Probabilidad Filtrado

Definición 4.1.7 (Espacio Filtrado). La evolución formal de un proceso estocástico en tiempo continuo requiere la introducción de una estructura tetra-dimensional denominada **espacio de probabilidad filtrado** o espacio adaptado:

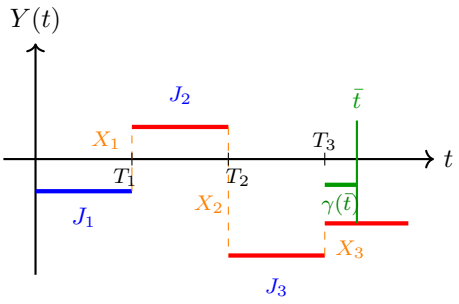
$$(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$$

donde la componente $\{\mathcal{F}_t\}_{t\geq 0}$ constituye una **filtración**, definida analíticamente como una colección o sucesión continua indexada y no decreciente de σ-álgebras contenidas en el espacio medible global $\mathcal{F}$:

$$\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F} \quad \forall\, 0 \leq s \leq t$$

Informalmente, la σ-álgebra $\mathcal{F}_t$ representa el flujo acumulado de información histórica e instrumental disponible y conocida que "influye" y condiciona dinámicamente el comportamiento del proceso hasta el instante cronológico t.

Típicamente en la teoría de procesos de saltos, hacemos uso de la **filtración natural** del sistema, la cual es generada única y exclusivamente por la historia endógena y la trayectoria pasada del propio proceso, es decir, es la que se genera si tu única fuente de información es mirar la trayectoria del propio objeto. No tienes información externa (como el clima o interferencias), solo sabes dónde ha estado la partícula y cuánto ha tardado en moverse.



Evaluemos el proceso compuesto en una ventana temporal fija e intermedia denotada por $\bar{t}$. De acuerdo con el grafo adjunto, en dicho instante se verifican los siguientes observables algebraicos:

$$Y(\bar{t}) = \sum_{i=1}^{3} X_i$$

$$\mathcal{F}_{\bar{t}} = \sigma\left(J_1, J_2, J_3, X_1, X_2, X_3, \gamma(\bar{t})\right)$$

$$\gamma(\bar{t}) = \bar{t} - T_3$$

La información acumulada en la σ-álgebra contiene de forma estricta las longitudes de los saltos, los intervalos previos y la fracción de tiempo consumida en el estado actual.

4.1.1. Definición Clásica de la Propiedad de Markov ($s > t$)

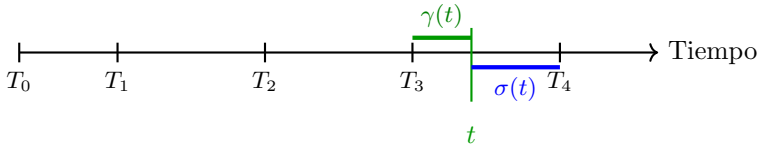
Un proceso estocástico ideal $(X_t)_{t\geq 0}$ satisface la propiedad de falta de memoria si la distribución predictiva condicional del futuro dada la filtración histórica colapsa en la información del presente, es decir

$$P(X_s \in B \mid \mathcal{F}_t) = P(X_s \in B \mid X_t = x)$$

lo que viene a ser

- El miembro izquierdo ($P(X_s \in B \mid \mathcal{F}_t)$): calcula la probabilidad de que el proceso en el futuro (s) aterrice en un estado B, condicionada a toda la bolsa de información histórica acumulada hasta el presente ($\mathcal{F}_t$) (es decir, sabiendo dónde estuvo la partícula en el segundo 1, en el segundo 2, a qué hora dio cada salto, etc.).
- El miembro derecho ($P(X_s \in B \mid X_t = x)$): calcula esa misma probabilidad futura, pero condicionada únicamente a saber el valor exacto en el instante presente ($X_t = x$), ignorando por completo el pasado.

A continuación comentaremos como podemos trabajar con el modelo dado, razonando el porque en parte puede actuar como un proceso markoviano. Para ello usaremos el siguiente diagrama de base



donde T_3 es el último tiempo de arribo de renovación que ocurrió antes o exactamente en el instante t, y T_4 es el próximo tiempo de arribo que ocurrirá después de t.

Para analizar el flujo de transición del proceso compuesto $Y(t) \longrightarrow Y(s)$ para un horizonte $s > t$, es imperativo segmentar el intervalo de confinamiento actual mediante dos variables temporales complementarias:

- $\gamma(t) = t - T_3$: Es una variable aleatoria determinista respecto a $\mathcal{F}_t$. Representa la **edad** o el tiempo transcurrido desde el último punto de renovación del sistema hasta el presente t. **Es una variable conocida.**
- $\sigma(t) = T_4 - t$: Es una variable aleatoria predictiva fuera del alcance de $\mathcal{F}_t$. Representa el **tiempo de vida residual** o remanente que la partícula pasará confinada antes de ejecutar el siguiente salto. **Es una variable desconocida.**

Debido a que el tiempo de vida residual $\sigma(t)$ se encuentra condicionado intrínsecamente por la edad acumulada $\gamma(t)$, la probabilidad de transición futura depende de la memoria de residencia, lo que clasifica a este sistema como un **PROCESO SEMI-MARKOVIANO**.

Proposición 4.1.2 (Vectorización de Markov). *A pesar de que el proceso compuesto espacial $Y(t)$ es inherentemente no markoviano por sí solo, la propiedad de falta de memoria se restaura de forma estricta expandiendo el espacio de estados. El par ordenado bidimensional constituido por la posición y la edad:*

$$(Y(t), \gamma(t))$$

*constituye un **proceso de Markov** puro en tiempo continuo.*

Proposición 4.1.3 (Condición de Markovianidad Pura). *El proceso compuesto unidimensional $Y(t)$ preserva la propiedad de Markov clásica si y solo si los tiempos de inter-arribo o espera J_i se distribuyen bajo una ley exponencial de tasa λ:*

$$Y(t) \text{ es un proceso de Markov} \iff J_i \sim Exp(\lambda) \quad \forall\, i \in \mathbb{N}^+$$

4.2. One-point Distribution of $Y(t)$

Definición 4.2.8 (Distribución Unipuntual). La función $q(t, u)$ representa la **función de densidad de probabilidad (f.d.p.) de un solo punto** (*one-point PDF*) del proceso compuesto $Y(t)$. Físicamente, determina la probabilidad de encontrar la partícula en la posición espacial exacta $u \in \mathbb{R}$ en un instante de tiempo fijo $t \geq 0$.

En el espacio real (t, u), el cálculo de $q(t, u)$ es algebraicamente impracticable debido a que el acoplamiento del CTRW genera infinitas integrales de convolución iteradas. Para resolver el sistema de forma inmediata, se proyecta el proceso hacia un **dominio dual transformado** aplicando simultáneamente:

- **Transformada de Fourier** sobre la variable espacial: $u \longrightarrow k$ (número de onda).
- **Transformada de Laplace** sobre la variable temporal: $t \longrightarrow s$ (frecuencia compleja).

Teorema 4.2.4 (Flujo Algorítmico del CTRW Desacoplado). *La transición de la complejidad infinitesimal del mundo real hacia la simplificación algebraica del espacio dual se sintetiza a través del siguiente mapa de ruta:*

Mundo Real (t, u)	$\xrightarrow{\mathcal{F}[\cdot] \ e \ \mathcal{L}[\cdot]}$	***Espacio Dual*** (s, k)
Densidad de Saltos $f_{X_1}(u)$	$\longrightarrow$	*Función Característica* $\hat{f}_{X_1}(k) = \int_{-\infty}^{\infty} f_{X_1}(u) e^{iku}\, du$
Densidad de Esperas $f_{J_1}(t)$	$\longrightarrow$	*Transformada de Laplace* $\widetilde{f}_{J_1}(s) = \int_0^{\infty} f_{J_1}(t) e^{-st}\, dt$
Supervivencia Temporal $\overline{F}_{J_1}(t)$	$\longrightarrow$	*Transformada de Supervivencia* $\widetilde{\overline{F}}_{J_1}(s) = \frac{1-\widetilde{f}_{J_1}(s)}{s}$
Densidad de Ocupación $q(t, u)$	$\longrightarrow$	***Solución Algebraica*** $\widetilde{\hat{q}}(s, k)$

Teorema 4.2.5 (Ecuación de Montroll-Weiss). *Bajo la hipótesis de desacoplamiento estricto (independencia mutua entre longitudes de salto y tiempos de espera), la solución exacta de la f.d.p. unipuntual en el espacio dual transformado de Fourier-Laplace colapsa de forma cerrada en la siguiente expresión fundamental:*

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

Protocolo de Uso en Problemas Prácticos

Paso 1: Obtener las transformadas base de los datos del problema: $\hat{f}_{X_1}(k)$ y $\widetilde{f}_{J_1}(s)$.

Paso 2: Sustituir directamente dichos términos en la fracción de Montroll-Weiss.

Paso 3: Aplicar la transformada inversa de Laplace ($\mathcal{L}^{-1}$) y la inversa de Fourier ($\mathcal{F}^{-1}$) sobre el resultado para regresar al espacio real y obtener la solución explícita $q(t,u)$.

4.3. Caso Particular: Tiempos de Espera Exponenciales

Retomamos la Ecuación de Montroll-Weiss para estudiar qué ocurre cuando los tiempos de espera J_i siguen la ley más simple posible: la exponencial. Este caso es especialmente relevante porque es el **único** que devuelve al CTRW la propiedad de Markov pura (sin necesidad de aumentar el espacio de estados con la edad $\gamma(t)$).

Ejemplo (Distribución del número de saltos $N(t)$). Sea $J_1 \sim \text{Exp}(\lambda)$. Sus transformadas de Laplace son

$$\widetilde{f}_{J_1}(s) = \frac{\lambda}{\lambda + s}, \qquad \widetilde{\overline{F}}_{J_1}(s) = \frac{1}{\lambda + s}.$$

Como $\widetilde{P}(n,s) = \left[\widetilde{f}_{J_1}(s)\right]^n \widetilde{\overline{F}}_{J_1}(s)$, sustituyendo se obtiene

$$\widetilde{P}(n,s) = \frac{\lambda^n}{(\lambda + s)^{n+1}}.$$

Invirtiendo la transformada de Laplace (reconociendo la forma de la densidad Gamma/Erlang desplazada), se llega a

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

Observación. El hecho de que $N(t)$ sea Poisson no es casual: es la **única** elección de f_{J_1} compatible con la propiedad de falta de memoria de los tiempos de espera. Esto convierte a $N(t)$ en un **proceso de Poisson**, que cumple

$$t_1 < t_2 \implies N(t_2) - N(t_1) \overset{d}{=} N(t_2 - t_1), \qquad N(t_2) - N(t_1) \perp\!\!\!\perp N(t_1),$$

es decir, tiene **incrementos independientes y estacionarios**. Esta propiedad es la que define a un **proceso de Lévy**.

4.4. Límite Difusivo: de la Ecuación de Renovación a la Ecuación del Calor

Existe una versión de la f.d.p. unipuntual escrita directamente en el espacio real (t,x), conocida como **ecuación de renovación semi-Markoviana**:

$$\mu(t,x) = \overline{F}_{J_1}(t)\,\delta(x) + \int_{-\infty}^{+\infty} dx' \int_0^t dt'\; f_{X_1}(x - x')\, f_{J_1}(t - t')\, \mu(x', t').$$

El primer término representa la probabilidad de que la partícula aún **no haya saltado** (sigue en $x = 0$); el segundo es la convolución espacio-temporal con todas las trayectorias posibles que ya dieron al menos un salto.

Observación. Aplicando Fourier-Laplace a esta ecuación se recupera, de forma equivalente, la propia ecuación de Montroll-Weiss:

$$\widetilde{\hat{\mu}}(s,k) = \frac{1-\widetilde{f}_{J_1}(s)}{s} \cdot \frac{1}{1-\hat{f}_{X_1}(k)\,\widetilde{f}_{J_1}(s)}.$$

Idea del límite hidrodinámico/difusivo: si los saltos espaciales tienen media cero y varianza finita, y los tiempos de espera tienen media finita, entonces para $k, s \to 0$ (es decir, observando el proceso a escalas grandes de espacio y tiempo) las transformadas admiten el desarrollo de Taylor

$$\hat{f}_{X_1}(k) \approx 1 - k^2 + \text{o.t.s.}, \qquad \widetilde{f}_{J_1}(s) \approx 1 - s + \text{o.t.s.}$$

Sustituyendo y reescalando $k \to c^{1/2}k$, $s \to cs$ con $c \to \infty$, el denominador $1 - \hat{f}_{X_1}\widetilde{f}_{J_1}$ se comporta como $c(k^2 + s)$, de modo que

$$\widetilde{\hat{\mu}}(s,k) \xrightarrow[c\to\infty]{} \widetilde{\hat{u}}(s,k) = \frac{1}{k^2+s}.$$

Teorema 4.4.6 (Ecuación de Difusión como límite del CTRW). *La relación* $\widetilde{\hat{u}}(s,k) = \dfrac{1}{k^2+s}$ *es equivalente, deshaciendo las transformadas, a invertir Fourier en k y Laplace en s sobre la ecuación algebraica* $k^2\hat{u} + s\hat{u} = 1$*, lo que corresponde al* ***problema de Cauchy para la ecuación de difusión****:*

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

cuya solución es el conocido ***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 aplicado al CTRW: bajo escalas grandes, **cualquier** CTRW desacoplado con saltos de varianza finita y esperas de media finita converge a un Movimiento Browniano, independientemente de la forma exacta de f_{X_1} y f_{J_1} (sólo dependen de ellas a través de sus dos primeros momentos).

Con esto cerramos el bloque de CTRW/renovación.

5. Procesos Gaussianos

Este bloque de la teoría se centrará en **variables y vectores Gaussianos**, herramienta fundamental para el estudio del Movimiento Browniano que aparecerá como límite en la sección anterior.

Definición 1 (Variable Aleatoria Normal). Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad. Una v.a. $X : \Omega \to \mathbb{R}$ es **normal o Gaussiana**, $X \sim N(\mu, \sigma^2)$, si admite densidad

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}, \qquad \mu \in \mathbb{R},\ \sigma > 0.$$

Se demuestra que necesariamente $\mu = E[X]$ y $\sigma^2 = \mathrm{Var}(X)$.

Definición 2 (Centrado y estandarización). Una v.a. Z es **centrada** si $E[Z] = 0$. Se dice que $X \sim N(0,1)$ es **normal estándar**, donde la estandarización

$$Z = \frac{X - E[X]}{\sigma} \sim N(0,1) \qquad \Longleftrightarrow \qquad X = \sigma Z + \mu \sim N(\mu, \sigma^2)$$

permite reducir cualquier cálculo sobre una normal genérica a uno sobre la normal estándar.

Proposición 1 (Propiedades de las v.a. normales). *Sea $X \sim N(\mu, \sigma^2)$. Entonces:*

1. ***Función característica:***

$$\hat{f}_X(k) = E[e^{ikX}] = e^{i\mu k - \frac{1}{2}\sigma^2 k^2}.$$

2. ***Momentos*** *(caso centrado $X \sim N(0, \sigma^2)$):*

$$E[X^n] = \begin{cases} \dfrac{(2k)!}{2^k k!}\,\sigma^{2k}, & n = 2k \\ 0, & n = 2k+1 \end{cases}, \qquad k = 0, 1, 2, \ldots$$

   es decir, todos los momentos impares se anulan por simetría y los pares se expresan en función de σ.

3. ***Concentración:*** *para $\kappa > 0$,*

$$P(|X - \mu| \le \kappa\sigma) = 2\Phi(\kappa) - 1,$$

   con Φ la f.d.a. de $Z \sim N(0,1)$. Esto cuantifica cómo se concentra la masa de probabilidad alrededor de la media en unidades de desviación típica.

4. ***Colas:*** *para $Z \sim N(0,1)$ y $x > 0$,*

$$\frac{1}{\sqrt{2\pi}}\left(\frac{1}{x} - \frac{1}{x^3}\right) e^{-x^2/2} \le P(Z \ge x) \le \frac{1}{\sqrt{2\pi}}\frac{1}{x} e^{-x^2/2},$$

   lo que muestra el conocido decaimiento super-rápido (gaussiano) de las colas.

Teorema 2 (Combinación lineal de normales). *Sean $X \sim N(\mu_1, \sigma_1^2)$ e $Y \sim N(\mu_2, \sigma_2^2)$ independientes, y $a \in \mathbb{R}$. Entonces*

$$X + Y \sim N(\mu_1 + \mu_2,\ \sigma_1^2 + \sigma_2^2), \qquad aX \sim N(a\mu_1,\ a^2\sigma_1^2).$$

Observación (Idea de la demostración). Por independencia, la función característica de la suma es el producto de las funciones características:

$$\hat{f}_{X+Y}(k) = \hat{f}_X(k)\hat{f}_Y(k) = e^{i(\mu_1+\mu_2)k - \frac{1}{2}(\sigma_1^2+\sigma_2^2)k^2},$$

que es exactamente la función característica de $N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$. Como la función característica determina unívocamente la ley, se concluye. El caso aX es análogo evaluando $\hat{f}_X(ak)$.

Proposición 3 (Estabilidad de la normalidad bajo límites). *Sea $(X_n)_{n\geq 0}$ una sucesión de v.a. normales independientes, $X_n \sim N(\mu_n, \sigma_n^2)$, que converge en distribución a una v.a. X. Entonces:*

1. *El límite es también normal, $X \sim N(\mu, \sigma^2)$, con $\mu = \lim_n \mu_n$ y $\sigma^2 = \lim_n \sigma_n^2$.*
2. *Si además (X_n) converge a X en probabilidad, entonces converge también en L^p para todo $1 \leq p < \infty$.*

5.1. Vectores Aleatorios Gaussianos

Definición 5.1.3 (Vector Gaussiano). Sea $X = (X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$. Decimos que X es un **vector Gaussiano** si **toda** combinación lineal de sus componentes es una v.a. normal:

$$\sum_{i=1}^{n} c_i X_i \text{ es normal}, \qquad \forall\, c_1, \ldots, c_n \in \mathbb{R}.$$

Observación. Si X es Gaussiano, cada componente X_i es necesariamente normal (tomando $c = e_i$). **El recíproco es falso**: que todas las marginales sean normales no implica que el vector sea Gaussiano.

Ejemplo (Contraejemplo del recíproco). Sea $X \sim N(0,1)$ y $a > 0$ fijo. Definimos

$$Y = \begin{cases} X & \text{si } |X| \leq a \\ -X & \text{si } |X| > a. \end{cases}$$

Se puede comprobar que $Y \sim N(0,1)$ (es normal), pero el par (X, Y) **no** es un vector Gaussiano (por ejemplo $X + Y$ no es normal, ya que vale $2X$ en $|X| \leq a$ y 0 fuera).

Definición 5.1.4 (Media y matriz de covarianzas). Sea $X = (X_1, \ldots, X_n)$ un vector Gaussiano. Se definen:

$$E[X] = \big(E[X_1], \ldots, E[X_n]\big) \quad \text{(vector de medias)},$$

$$C(X) = \big(\mathrm{Cov}(X_i, X_j)\big)_{i,j=1}^{n} \quad \text{(matriz de covarianzas)}, \quad \mathrm{Cov}(X_i, X_j) = E[X_i X_j] - E[X_i]E[X_j].$$

Un vector Gaussiano se dice **centrado** si $E[X] = 0$.

Teorema 5.1.4 (Función característica de un vector Gaussiano). *Sea $X = (X_1, \ldots, X_n)$ un vector Gaussiano con media m y matriz de covarianzas $C(X)$. Entonces, para todo $x \in \mathbb{R}^n$,*

$$\hat{f}_X(x) = \exp\left(i\langle m, x\rangle - \frac{1}{2}\langle x, C(X)x\rangle \right).$$

Observación (Idea de la demostración). Para $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$, por definición de vector Gaussiano la combinación lineal $\langle a, X\rangle = \sum_i a_i X_i$ es una v.a. normal, con

$$E[\langle a, X\rangle] = \langle a, m\rangle, \qquad \mathrm{Var}(\langle a, X\rangle) = \langle a, C(X)a\rangle.$$

Su función característica (escalar) es entonces $e^{i\langle a,m\rangle - \frac{1}{2}\langle a, C(X)a\rangle}$. Tomando $a = x$ y evaluando en el punto 1 se obtiene exactamente $\hat{f}_X(x)$.

Proposición 5.1.5 (Independencia y covarianza nula). *Sean X, Y dos v.a. normales (conjuntamente Gaussianas). Entonces*

$$X \perp\!\!\!\perp Y \quad \Longleftrightarrow \quad Cov(X, Y) = 0.$$

Observación (Idea de la demostración). ($\Rightarrow$) Es válido en general: independencia implica $E[XY] = E[X]E[Y]$, luego $\mathrm{Cov}(X,Y) = 0$.
($\Leftarrow$) Aquí se usa que el vector es Gaussiano: si $\mathrm{Cov}(X,Y) = 0$, la matriz de covarianzas del vector (X,Y) es diagonal, $C = \mathrm{diag}(\sigma_X^2, \sigma_Y^2)$. Sustituyendo en la función característica del vector,

$$\hat{f}_{(X,Y)}(u) = e^{-\frac{1}{2}(u_1^2\sigma_X^2 + u_2^2\sigma_Y^2)} = \hat{f}_X(u_1)\,\hat{f}_Y(u_2),$$

y como la función característica conjunta factoriza como producto de las marginales, X e Y son independientes.

Teorema 5.1.6 (Densidad Gaussiana multivariante). *Sea $X = (X_1, \ldots, X_n)$ un vector Gaussiano en $\mathbb{R}^n$ con media m y matriz de covarianzas $C(X)$* ***invertible****. Entonces X admite densidad de probabilidad*

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

Observación. Esta fórmula generaliza directamente la densidad normal univariante: cuando $n = 1$ se recupera

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

La condición de invertibilidad de C es necesaria: si C es singular, el vector vive en realidad en un subespacio de dimensión menor que n y no existe densidad respecto a la medida de Lebesgue en $\mathbb{R}^n$.

5.2. Multivariant Gaussian Distribution

Definición 5.2.5 (Función de Densidad de Probabilidad Multivariante). Sea $\underline{X} = (X_1, \ldots, X_n)$ un vector aleatorio Gaussiano en el espacio euclídeo $\mathbb{R}^n$, caracterizado por un vector de valores esperados (medias) $m = \mathbb{E}[\underline{X}]$ y una matriz de covarianzas $C(\underline{X})$. Asumiendo que la matriz de covarianzas es regular y admite inversa (es decir, existe su matriz de precisión $C^{-1}(\underline{X})$), entonces la **función de densidad de probabilidad (f.d.p.)** del vector $\underline{X}$ viene definida de forma explícita por la expresión:

$$f_{\underline{X}}(x) = \frac{1}{(2\pi)^{n/2}\,(\det(C))^{1/2}}\, e^{-\frac{1}{2}\langle x-m,\, C^{-1}(x-m)\rangle}$$

donde $\langle\cdot,\cdot\rangle$ denota el producto escalar euclídeo estándar en $\mathbb{R}^n$, el cual parametriza la forma cuadrática de la distancia de Mahalanobis.

Demostración. □

Lema 5.2.7 (Isomorfismo Estocástico Gausiano e Inversión Lineal). *Si el vector aleatorio $\underline{X}$ es Gaussiano con vector de medias m y una matriz de covarianzas $C(\underline{X})$ simétrica y definida positiva (y por ende, invertible), se cumple de forma unívoca que la transformación lineal normalizada:*

$$C^{-1/2}(\underline{X} - m) \sim \mathcal{N}(0, I_n) \quad \textit{constituye un vector Gaussiano estándar (v.G.e.).}$$

Inversamente, si partimos de un vector Gaussiano estándar nativo denotado como $Z \sim \mathcal{N}(0, I_n)$, la transformación afín mapeada por la raíz cuadrada matricial y el desplazamiento del promedio genera el vector original:

$$\underline{X} = m + C^{1/2}(Z) \quad \textit{el cual recupera las propiedades } \mathbb{E}[\underline{X}] = m \textit{ y } Var(\underline{X}) = C(\underline{X}).$$

$$\textit{Representación sintética de síntesis: } \underline{X} = C^{1/2}(Z) + m \quad \text{con } Z \sim \mathcal{N}(0, I_n).$$

Etapa 2: Aplicación del Teorema del Cambio de Variable mediante Medida de Cajas
Definamos un dominio canónico E como cualquier hiperrectángulo o caja n-dimensional acotada inmersa en el espacio medible $\mathbb{R}^n$:

$$E = \bigotimes_{i=1}^{n} [a_i, b_i]$$

Determinar que el vector de interés se localiza en el interior de dicha región equivale a restringir cada componente marginal a sus intervalos independientes correspondientes:

$$\underline{X} \in E \equiv X_1 \in [a_1, b_1], \dots, X_n \in [a_n, b_n]$$

Modelemos la medida de probabilidad de este suceso utilizando la relación de la transformación lineal deducida en el Lema ($\underline{X} = C^{1/2}(Z) + m$). Al operar de forma algebraica con los límites del conjunto y aplicar el formalismo clásico de integración multidimensional bajo cambios de variable coordenados, se desarrolla la siguiente cadena de identidades:

$$\mathbb{P}(\underline{X} \in E) = \mathbb{P}\left(C^{1/2}(Z) + m \in E\right) = \mathbb{P}\left[Z \in C^{-1/2}(E - m)\right] = \int_{C^{-1/2}(E-m)} f_Z(u)\, du =$$

Introduciendo la sustitución geométrica del cambio de base lineal $y = C^{1/2}u + m$, el diferencial de volumen en la medida de Lebesgue se escala a través del determinante de la matriz Jacobiana del sistema, es decir, $dy = \det(C^{1/2})du = \sqrt{\det(C)}\, du$. Ajustando consecuentemente el dominio y el argumento de la f.d.p. estándar de Z, el sumatorio se transforma en:

$$= \int_E f_Z\left(C^{-1/2}(y - m)\right) \frac{dy}{\sqrt{\det(C)}} = \int_E \frac{1}{(2\pi)^{n/2}\sqrt{\det(C)}} \exp\left[-\frac{1}{2}\left\|C^{-1/2}(y - m)\right\|^2\right] dy =$$

Por último, aplicando las propiedades de la transposición de operadores matriciales y la linealidad del producto escalar, la norma cuadrada del argumento linealizado se reestructura como una forma bilineal gobernada por la matriz de precisión inversa:

$$\left\|C^{-1/2}(y - m)\right\|^2 = \langle C^{-1/2}(y-m),\, C^{-1/2}(y-m)\rangle = \langle y-m,\, \left(C^{-1/2}\right)^T C^{-1/2}(y-m)\rangle = \langle y-m,\, C^{-1}(y-m)\rangle$$

Sustituyendo este núcleo en la integral del volumen medible, concluimos la densidad de probabilidad espacial:

$$= \int_E \frac{1}{(2\pi)^{n/2}\sqrt{\det(C)}} e^{-\frac{1}{2}\langle y-m,\, C^{-1}(y-m)\rangle}\, dy$$

Conclusión del Teorema:
Puesto que la relación probabilística se verifica formalmente para cualquier caja elemental $E \subset \mathbb{R}^n$, el integrando resultante se identifica de manera unívoca como el núcleo de la densidad conjunta del sistema. Así, queda demostrado con total rigor que la f.d.p. del vector Gaussiano generalizado $\underline{X}$ es:

$$f_{\underline{X}}(x) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(C)}} \exp\left[-\frac{1}{2}\langle x - m,\, C^{-1}(x - m)\rangle\right]$$

Ejemplo (Deducción Explícita para la Dimensión $n = 2$). Desarrollemos rigurosamente la función de densidad de probabilidad (f.d.p.) de un vector Gaussiano multivariante en el caso bidimensional. Fijemos el orden del sistema en $n = 2$, definiendo el vector aleatorio de componentes continuas $\underline{X}$ y su correspondiente vector de evaluación espacial x como:

$$\underline{X} = (X_1, X_2) \in \mathbb{R}^2, \qquad x = (x_1, x_2) \in \mathbb{R}^2$$

Supongamos que el vector de medias o valores esperados del sistema viene dado por $m = (m_1, m_2) = (\mathbb{E}[X_1], \mathbb{E}[X_2])$.

1. Parametrización y Determinante de la Matriz de Covarianza
La matriz de covarianza simétrica y definida positiva C se construye a partir de las varianzas marginales y las covarianzas cruzadas de las componentes del vector:

$$C = \begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) \end{pmatrix} = \begin{pmatrix} \sigma_{X_1}^2 & \sigma_{X_1X_2} \\ \sigma_{X_1X_2} & \sigma_{X_2}^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} \\ \sigma_1\sigma_2\rho_{12} & \sigma_2^2 \end{pmatrix}$$

Donde introducemos las simplificaciones estadísticas estandarizadas para las desviaciones típicas ($\sigma_{X_1} = \sigma_1$, $\sigma_{X_2} = \sigma_2$), la covarianza ($\sigma_{X_1X_2} = \sigma_{12}$) y el coeficiente de correlación lineal de Pearson ρ_{12}, definido axiomáticamente como:

$$\rho_{12} = \rho_{X_1X_2} = \frac{\sigma_{X_1X_2}}{\sigma_{X_1}\sigma_{X_2}} = \frac{\sigma_{12}}{\sigma_1\sigma_2}$$

Calculemos el determinante de la matriz de covarianza C para estructurar el factor de normalización de la densidad:

$$\det(C) = \det\begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} \\ \sigma_1\sigma_2\rho_{12} & \sigma_2^2 \end{pmatrix} = \sigma_1^2\sigma_2^2 - (\sigma_1\sigma_2\rho_{12})^2 = \sigma_1^2\sigma_2^2(1-\rho_{12}^2)$$

Extrayendo la raíz cuadrada del determinante, obtenemos el coeficiente del denominador de la constante de normalización:

$$\sqrt{\det(C)} = \sigma_1\sigma_2\sqrt{1-\rho_{12}^2}$$

2. Inversión Algebraica de la Matriz de Covarianza (C^{-1})
Para hallar la matriz inversa de precisión o concentración, planteamos la identidad fundamental del producto matricial $C \cdot C^{-1} = I_2$:

$$\begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} \\ \sigma_1\sigma_2\rho_{12} & \sigma_2^2 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Esta multiplicación matricial da origen a un sistema lineal acoplado de cuatro ecuaciones algebraicas con cuatro incógnitas (a, b, c, d):

$$\begin{cases} 1)\ \sigma_1^2 a + \sigma_1\sigma_2\rho_{12}c = 1 \\ 2)\ \sigma_1^2 b + \sigma_1\sigma_2\rho_{12}d = 0 \\ 3)\ \sigma_1\sigma_2\rho_{12}a + \sigma_2^2 c = 0 \\ 4)\ \sigma_1\sigma_2\rho_{12}b + \sigma_2^2 d = 1 \end{cases}$$

• **Resolución de la segunda columna (b, d):**
Aislando el coeficiente b a partir de la relación lineal de la ecuación 2):

$$b = -\frac{\sigma_1\sigma_2\rho_{12}}{\sigma_1^2}d = -\frac{\sigma_2}{\sigma_1}\rho_{12}d$$

Sustituyendo directamente esta equivalencia en la ecuación 4) para romper el acoplamiento:

$$\sigma_1\sigma_2\rho_{12}\left(-\frac{\sigma_2}{\sigma_1}\rho_{12}d\right) + \sigma_2^2 d = 1 \implies -\sigma_2^2\rho_{12}^2 d + \sigma_2^2 d = 1 \implies \sigma_2^2(1-\rho_{12}^2)d = 1$$

$$\implies \boxed{d = \frac{1}{\sigma_2^2(1-\rho_{12}^2)}}$$

Sustituyendo de vuelta el valor obtenido para d en la expresión despejada de b, deducimos:

$$b = -\frac{\sigma_2}{\sigma_1}\rho_{12}\left(\frac{1}{\sigma_2^2(1-\rho_{12}^2)}\right) \implies \boxed{b = -\frac{\rho_{12}}{\sigma_1\sigma_2(1-\rho_{12}^2)}}$$

• **Resolución de la primera columna (a, c):**
Aislando de forma homóloga el coeficiente c a partir de la relación lineal de la ecuación 3):

$$c = -\frac{\sigma_1\sigma_2\rho_{12}}{\sigma_2^2}a = -\frac{\sigma_1}{\sigma_2}\rho_{12}a$$

Sustituyendo de manera análoga esta expresión en la ecuación 1):

$$\sigma_1^2 a + \sigma_1\sigma_2\rho_{12}\left(-\frac{\sigma_1}{\sigma_2}\rho_{12}a\right) = 1 \implies \sigma_1^2 a - \sigma_1^2\rho_{12}^2 a = 1 \implies \sigma_1^2(1-\rho_{12}^2)a = 1$$

$$\implies \boxed{a = \frac{1}{\sigma_1^2(1-\rho_{12}^2)}}$$

Sustituyendo el valor obtenido para a en la expresión despejada de c, deducimos (lo que confirma la simetría $b = c$):

$$c = -\frac{\sigma_1}{\sigma_2}\rho_{12}\left(\frac{1}{\sigma_1^2(1-\rho_{12}^2)}\right) \implies \boxed{c = -\frac{\rho_{12}}{\sigma_1\sigma_2(1-\rho_{12}^2)}}$$

Agrupando los cuatro componentes elementales en la estructura matricial original, la inversa de la matriz de covarianza C^{-1} queda determinada de forma exacta como:

$$C^{-1} = \begin{pmatrix} \frac{1}{\sigma_1^2(1-\rho_{12}^2)} & -\frac{\rho_{12}}{\sigma_1\sigma_2(1-\rho_{12}^2)} \\ -\frac{\rho_{12}}{\sigma_1\sigma_2(1-\rho_{12}^2)} & \frac{1}{\sigma_2^2(1-\rho_{12}^2)} \end{pmatrix}$$

---

3. Desarrollo de la Forma Cuadrática de la Distancia
Planteemos la acción de la matriz de precisión C^{-1} operando sobre el vector de desviaciones centrado $(x - m)$:

$$C^{-1}(x-m) = \begin{pmatrix} \frac{1}{\sigma_1^2(1-\rho_{12}^2)} & -\frac{\rho_{12}}{\sigma_1\sigma_2(1-\rho_{12}^2)} \\ -\frac{\rho_{12}}{\sigma_1\sigma_2(1-\rho_{12}^2)} & \frac{1}{\sigma_2^2(1-\rho_{12}^2)} \end{pmatrix} \begin{pmatrix} x_1 - m_1 \\ x_2 - m_2 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{\sigma_1^2(1-\rho_{12}^2)}(x_1-m_1) - \frac{\rho_{12}}{\sigma_1\sigma_2(1-\rho_{12}^2)}(x_2-m_2) \\ \frac{1}{\sigma_2^2(1-\rho_{12}^2)}(x_2-m_2) - \frac{\rho_{12}}{\sigma_1\sigma_2(1-\rho_{12}^2)}(x_1-m_1) \end{pmatrix}$$

Para construir el argumento exponencial de la densidad multivariante, ejecutamos el producto escalar interno $\langle x-m,\, C^{-1}(x-m)\rangle$, premultiplicando el vector columna anterior por el vector fila de desviaciones $(x_1 - m_1,\, x_2 - m_2)$:

$$\langle x-m,\, C^{-1}(x-m)\rangle =$$

$$= (x_1-m_1,\, x_2-m_2) \begin{pmatrix} \frac{1}{\sigma_1^2(1-\rho_{12}^2)}(x_1-m_1) - \frac{\rho_{12}}{\sigma_1\sigma_2(1-\rho_{12}^2)}(x_2-m_2) \\ \frac{1}{\sigma_2^2(1-\rho_{12}^2)}(x_2-m_2) - \frac{\rho_{12}}{\sigma_1\sigma_2(1-\rho_{12}^2)}(x_1-m_1) \end{pmatrix} =$$

Expandiendo el álgebra distributiva del producto y agrupando los dos términos cruzados idénticos en un único factor multiplicativo doble, la forma cuadrática se reduce a la expresión escalar continua:

$$= \frac{1}{\sigma_1^2(1-\rho_{12}^2)}(x_1-m_1)^2 - 2\frac{\rho_{12}}{\sigma_1\sigma_2(1-\rho_{12}^2)}(x_1-m_1)(x_2-m_2) + \frac{1}{\sigma_2^2(1-\rho_{12}^2)}(x_2-m_2)^2$$

4. Formulación Exclícita Final de la Densidad Bivariante
Sustituyendo el término del determinante extraído $\sqrt{\det(C)}$ en la constante multiplicativa externa $\frac{1}{(2\pi)^{n/2}\sqrt{\det(C)}}$ para $n = 2$, y el polinomio escalar expandido en la componente interna del exponente $-\frac{1}{2}\langle\cdot\rangle$, la función de densidad de probabilidad conjunta bivariante $f_{X_1,X_2}(x_1, x_2)$ se escribe analíticamente como:

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{12}^2}} \exp\Bigg(-\frac{(x_1-m_1)^2}{2\sigma_1^2(1-\rho_{12}^2)} - \frac{(x_2-m_2)^2}{2\sigma_2^2(1-\rho_{12}^2)} + \\ + \frac{\rho_{12}}{\sigma_1\sigma_2(1-\rho_{12}^2)}(x_1-m_1)(x_2-m_2)\Bigg)$$

5.3. Existencia de Procesos Gaussianos

Hasta ahora hemos trabajado con vectores Gaussianos de dimensión finita. El siguiente paso es preguntarnos si, dadas una función de medias y una función de covarianzas razonables", existe realmente un **proceso** Gaussiano (infinitas variables, una por cada $t \geq 0$) que las tenga como tales.

Definición 5.3.6 (Función simétrica y positiva definida). Una función $K : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}$ es:

- **Simétrica** si $K(t,s) = K(s,t)$ para todo s, t.
- **Positiva definida** si para todo $n \in \mathbb{N}$, todo $0 \leq t_1 < \ldots < t_n$ y todo $c_1, \ldots, c_n \in \mathbb{R}$ se tiene $\sum_{i,j} c_i c_j K(t_i, t_j) \geq 0$.

Definición 5.3.7 (Proceso Gaussiano, media y covarianza). Sea $(X_t)_{t\geq 0}$ un proceso Gaussiano en $(\Omega, \mathcal{F}, P)$ (es decir, todo vector finito $(X_{t_1}, \ldots, X_{t_n})$ es Gaussiano). Se definen:

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

El proceso es **centrado** si $m(t) = 0\ \forall t$; en general podemos centrarlo con $Y_t = X_t - m(t)$.

Proposición 5.3.8. *La función de covarianza $K(s,t)$ de un proceso Gaussiano es siempre simétrica y positiva definida (es consecuencia directa de que $Var(\sum_i c_i X_{t_i}) \geq 0$).*

Teorema 5.3.9 (Existencia, vía Kolmogorov). *Sean $m : \mathbb{R}^+ \to \mathbb{R}$ una función arbitraria y $K : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}$ simétrica y positiva definida. Entonces existe un espacio de probabilidad $(\Omega, \mathcal{F}, P)$ y un proceso Gaussiano $(X_t)_{t\geq 0}$ sobre él tal que $E[X_t] = m(t)$ y $Cov(X_s, X_t) = K(s,t)$.*

Observación (Idea de la demostración). Basta construir, para cada conjunto finito de tiempos $S = \{t_1, \ldots, t_K\}$, la matriz $K^S = (K(t_i, t_j))_{i,j}$, que es simétrica y positiva definida y por tanto define un vector Gaussiano X^S válido. Como estas matrices son compatibles entre sí (coinciden en las entradas comunes cuando $S \subset S'$), se cumplen las **condiciones de consistencia de Kolmogorov**, y el **Teorema de Extensión de Kolmogorov** garantiza que existe una única medida de probabilidad sobre el proceso completo $(X_t)_{t\geq 0}$ con esas distribuciones finito-dimensionales. Este resultado es, esencialmente, el que prueba la **existencia** de cualquier proceso Gaussiano (en particular, del Movimiento Browniano).

5.4. Estacionariedad

Definición 5.4.8 (Estacionariedad). Un proceso real $(X_t)_{t\geq 0}$ es:

- **(Estrictamente) estacionario** si para todo $n \in \mathbb{N}$, todo $0 \leq t_1 < \ldots < t_n$ y todo $s > 0$, los vectores $(X_{t_1}, \ldots, X_{t_n})$ y $(X_{t_1+s}, \ldots, X_{t_n+s})$ tienen la misma distribución.
- **Débilmente estacionario** si $E[X_t] = E[X_{t+s}]$ y $\mathrm{Cov}(X_{t_1}, X_{t_2}) = \mathrm{Cov}(X_{t_1+s}, X_{t_2+s})$ para todo $t_1, t_2, s \geq 0$.

Observación. Definiendo la **función de autocovarianza** $c(t, t+s) = \mathrm{Cov}(X_t, X_{t+s})$, la estacionariedad débil equivale a que c dependa **solo de la diferencia** s entre los tiempos, no de su valor absoluto: $c(u, u+v) = c(0, v)$.

Teorema 5.4.10. *Si $(X_t)_{t\geq 0}$ es un proceso **Gaussiano**, entonces es estrictamente estacionario si y solo si es débilmente estacionario.*

Observación. Esta equivalencia es **especial de los procesos Gaussianos**: fuera del contexto Gaussiano, en general la estacionariedad débil no implica la fuerte (la débil solo controla los dos primeros momentos, mientras que para procesos Gaussianos la distribución completa queda determinada por la media y la covarianza).

5.5. Propiedades de las Trayectorias

Definición 5.5.9 (Indistinguibilidad y versiones). Sean $(X_t)_{t\geq 0}$ e $(Y_t)_{t\geq 0}$ procesos reales en $(\Omega, \mathcal{F}, P)$.

- Son **indistinguibles** si existe un conjunto de medida nula N tal que $X_t(\omega) = Y_t(\omega)$ para todo $\omega \notin N$ y **todo** $t \geq 0$ simultáneamente (condición muy fuerte: un único N vale para todos los tiempos a la vez).
- Son **versiones** (o modificaciones) la una de la otra si para cada $t \geq 0$ existe un conjunto de medida nula N_t (que puede depender de t) tal que $X_t(\omega) = Y_t(\omega)$ para todo $\omega \notin N_t$.

Definición 5.5.10 (Continuidad de Hölder). Una función $\rho : \mathbb{R}^+ \to \mathbb{R}$ es **localmente Hölder-continua** en $t \geq 0$ con exponente $\gamma > 0$ si existen $\varepsilon, c > 0$ tales que $|\rho(t) - \rho(s)| \leq c|t-s|^\gamma$ para todo s con $|t-s| < \varepsilon$.

Teorema 5.5.11 (Kolmogorov-Chentsov). *Sea $(X_t)_{t\geq 0}$ un proceso real en $(\Omega, \mathcal{F}, P)$. Supongamos que existen $\alpha \geq 1,\ \beta, c > 0$ tales que*

$$E\big[|X_t - X_s|^\alpha\big] \leq c\,|t-s|^{1+\beta} \qquad \forall\, s,t \geq 0.$$

Entonces, para cada $T > 0$, (X_t) tiene una versión $(Y_t)_{t\in[0,T]}$ con trayectorias γ-Hölder continuas, para cualquier $0 \leq \gamma < \beta/\alpha$.

Teorema 5.5.12 (Corolario para procesos Gaussianos centrados). *Sea $(X_t)_{t\geq 0}$ un proceso Gaussiano centrado con función de covarianza $K(s,t)$. Si existen $\beta, c > 0$ tales que*

$$K(t,t) + K(s,s) - 2K(t,s) \leq c\,|t-s|^\beta \qquad \forall\, s,t \geq 0,$$

entonces para todo $\gamma < \beta/2$ el proceso tiene una versión con trayectorias γ-Hölder continuas.

Observación (Idea de la demostración). El término de la izquierda es exactamente $E[(X_t - X_s)^2] = \mathrm{Var}(X_t - X_s)$. Usando que los momentos pares de una Gaussiana centrada se expresan en función de la varianza, $E[(X_t - X_s)^{2n}] = \frac{(2n)!}{n!2^n}\big[\mathrm{Var}(X_t - X_s)\big]^n \leq c'|t-s|^{(n\beta-1)+1}$, y aplicando Kolmogorov-Chentsov con $\alpha = 2n$ para cada n, se obtiene Hölder continuidad para exponente $\gamma < \dfrac{n\beta - 1}{2n}$; optimizando en $n \to \infty$ se llega a $\gamma < \beta/2$.

Definición 5.5.11 (Recordatorio: Propiedad de Markov). Un proceso $(X_t)_{t\geq 0}$ tiene la **propiedad de Markov** si para todo $n \in \mathbb{N}$, todo $t_1 < \ldots < t_n$ y todo $x, x_1, \ldots, x_{n-1} \in \mathbb{R}$,

$$P(X_{t_n} \leq x \mid X_{t_1} = x_1, \ldots, X_{t_{n-1}} = x_{n-1}) = P(X_{t_n} \leq x \mid X_{t_{n-1}} = x_{n-1}).$$

5.6. RKHS y Expansión de Karhunen-Loève

Estos dos resultados se enuncian a nivel de cultura general (no se exige reproducir las demostraciones, que son largas): permiten representar cualquier proceso Gaussiano centrado como una "serie de Fourier.ªleatoria.

Teorema 5.6.13 (Reproducing Kernel Hilbert Space, RKHS). *Sea $K : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}$ una función de covarianza continua. Existe un espacio de Hilbert $\mathcal{H}_K$ de funciones reales continuas, con producto escalar $(\cdot,\cdot)_{\mathcal{H}_K}$, tal que*

$$K(s,\cdot) \in \mathcal{H}_K \quad \forall s \geq 0, \qquad \big(g, K(s,\cdot)\big)_{\mathcal{H}_K} = g(s) \quad \forall\, g \in \mathcal{H}_K,\ s \geq 0$$

*(la llamada **propiedad reproductora**: evaluar g en s equivale a proyectarla escalarmente sobre $K(s,\cdot)$).*

Teorema 5.6.14 (Expansión de Karhunen-Loève). *Sea $(X_t)_{t\geq 0}$ un proceso Gaussiano centrado con covarianza continua K, definido en $(\Omega, \mathcal{F}, P)$. Entonces admite la representación, en el sentido de $L^2(\Omega, P)$,*

$$X_t = \sum_{j=1}^{\infty} \varphi_j(t)\, Z_j, \qquad \text{donde } \varphi_j \text{ son funciones continuas en } \mathbb{R}^+ \text{ y } Z_j \overset{i.i.d.}{\sim} N(0,1).$$

Observación. La idea es construir un isomorfismo entre $\mathcal{H}_K$ y el subespacio de $L^2(\Omega, P)$ generado por las X_t; al expresar $K(s,\cdot)$ en una base ortonormal (φ_n) de $\mathcal{H}_K$, las variables Z_j resultantes son combinaciones lineales de Gaussianas no correlacionadas, y por tanto **independientes** entre sí.

5.7. Movimiento Browniano

5.7.1. Motivación: límite difusivo de un CTRW

Recordamos el modelo de paseo aleatorio en tiempo continuo con $J_i \sim \text{Exp}(\lambda)$ y $X_i \sim N(0, \delta^2)$. Calculando la f.d.p. de $Y(t)$ y pasando al espacio de Laplace-Fourier (vía la fórmula tipo Montroll-Weiss), se obtiene la aproximación

$$\widehat{p}(u,s) \approx \frac{1}{s + u^2\left(\frac{1}{\lambda\delta^2}\right)^{-1}} \cdots$$

y, tomando el **límite de escala** $\delta \to 0,\ \lambda \to \infty$ manteniendo $\delta^2\lambda = 1$ (el "movimiento Browniano emerge cuando los saltos son infinitesimales pero infinitamente frecuentes"), se llega exactamente al núcleo de la ecuación de difusión:

$$\widehat{p}(u,s) \longrightarrow \widehat{u}(u,s) = \frac{1}{s+u^2} \quad \Longrightarrow \quad u(x,t) = \frac{1}{\sqrt{2\pi t}}\, e^{-\frac{x^2}{2t}}.$$

Esta densidad Gaussiana es la que caracteriza al Movimiento Browniano en cada instante t fijo.

5.7.2. Definición y propiedades fundamentales

Definición 5.7.12 (Movimiento Browniano Estándar, MBE). Un **Movimiento Browniano estándar** (también llamado proceso de Wiener) es un proceso Gaussiano $(B_t)_{t\geq 0}$ con

$$E[B_t] = 0 \ \ \forall t \geq 0, \qquad \text{Cov}(B_s, B_t) = s \wedge t = \text{mín}(s,t).$$

Observación. La función $K(s,t) = s \wedge t$ es simétrica y positiva definida, así que el Teorema de Existencia (Kolmogorov) garantiza directamente que el MBE **existe**.

Proposición 5.7.15 (Distribución de los incrementos). *Para $0 \leq s < t$, el incremento $B_t - B_s$ es Gaussiano con*

$$E[B_t - B_s] = 0, \qquad \mathit{Var}(B_t - B_s) = E[B_t^2] - 2\,\mathit{Cov}(B_t, B_s) + E[B_s^2] = t - 2s + s = t - s,$$

*es decir, $B_t - B_s \sim N(0, t-s)$: la varianza del incremento depende solo de la **longitud** del intervalo de tiempo.*

Teorema 5.7.16. *$(B_t)_{t\geq 0}$ admite una versión con trayectorias **Hölder-continuas** (consecuencia directa del corolario de Kolmogorov-Chentsov, usando que $K(t,t) + K(s,s) - 2K(s,t) = |t-s|$, es decir $\beta = 1$, lo que da continuidad Hölder para cualquier exponente $\gamma < 1/2$).*

Teorema 5.7.17 (Propiedades características del MBE). *El Movimiento Browniano Estándar cumple:*

(i) $P(B_0 = 0) = 1$.

(ii) $t \mapsto B_t(\omega)$ *es continua para P-casi todo ω.*

*(iii) Para toda partición $0 = t_0 < t_1 < \ldots < t_n$, los incrementos $(B_{t_i} - B_{t_{i-1}})_{1\leq i\leq n}$ son **independientes** y $B_{t_i} - B_{t_{i-1}} \sim N(0, t_i - t_{i-1})$ (**incrementos independientes y estacionarios**).*

Observación (Idea de la demostración). (i) y (ii) son consecuencia directa de la definición. Para (iii), la distribución de cada incremento ya se obtuvo arriba; la independencia se prueba viendo que, para $r < s < t$, el vector $(B_s - B_r,\ B_t - B_s)$ es Gaussiano con $\text{Cov}(B_s - B_r,\ B_t - B_s) = 0$ (cálculo directo usando $\text{Cov}(B_u, B_v) = u \wedge v$), y por la Proposición de independencia-covarianza-nula vista antes, covarianza cero entre Gaussianas conjuntas implica independencia.

Proposición 5.7.18 (Propiedades de simetría e invarianza del MBE). *Sea $(B_t)_{t\geq 0}$ un MBE. Entonces:*

(i) (Traslación espacial) $(x + B_t)_{t\geq 0}$ es un Movimiento Browniano que parte de x.

(ii) (Reflexión) $(-B_t)_{t\geq 0} \sim (B_t)_{t\geq 0}$.

(iii) (Escalado Browniano) Para todo $c > 0$, $(\sqrt{c}\, B_{t/c})_{t\geq 0} \sim (B_t)_{t\geq 0}$.

(iv) (Inversión temporal) $Z_t := t\, B_{1/t}$ (con $Z_0 := 0$) satisface $(Z_t)_{t\geq 0} \sim (B_t)_{t\geq 0}$.

(v) (Reversión temporal en $[0, t]$) $(B_t)_t \sim (B_t - B_{t-\cdot})$.

Observación. Todas se demuestran comprobando que el proceso transformado vuelve a ser Gaussiano, centrado, y con la misma función de covarianza $s\wedge t$ (por ejemplo en (iii): $\mathrm{Cov}(\sqrt{c}B_{s/c}, \sqrt{c}B_{t/c}) = c\,(s/c \wedge t/c) = s \wedge t$). La propiedad (iv) es la que permite extender resultados de comportamiento en $t \to 0$ a resultados en $t \to \infty$, y viceversa.

Proposición 5.7.19 (Propiedad de Markov del MBE). *El Movimiento Browniano Estándar es un proceso de Markov.*

Observación (Idea de la demostración). Se define $R(s,t) = \dfrac{K(s,t)}{K(s,s)} = \dfrac{s\wedge t}{s}$, y se comprueba que para $0 \leq r < s < t$ se cumple la **factorización** $R(r,t) = R(r,s)R(s,t)$ (de hecho los tres valen 1 en este caso). Esta condición algebraica es exactamente el **criterio de Doob** para procesos Gaussianos, que garantiza la propiedad de Markov.

5.8. Propiedad de Martingala

Definición 5.8.13 (Filtración natural). Para un proceso $(X_t)_{t\geq 0}$, la **filtración natural** $\mathcal{F}_t^X = \sigma(X_s,\, 0 \leq s \leq t)$ es la familia (no decreciente) de σ-álgebras generadas por la historia del proceso hasta el instante t.

Definición 5.8.14 (Martingala). Sea (X_t) un proceso sobre el espacio filtrado $(\Omega, \mathcal{F}, \mathcal{F}_t^X, P)$. Un proceso (Y_t) es una **martingala** respecto a $(\mathcal{F}_t^X)$ si:

(i) Es **adaptado** (Y_t es $\mathcal{F}_t^X$-medible) e **integrable**: $E[|Y_t|] < \infty\ \forall t \geq 0$.

(ii) $E[Y_t \mid \mathcal{F}_s^X] = Y_s$ para todo $s < t$ (la mejor predicción del futuro, dado el presente, es el valor actual).

Lema 5.8.20 (Ya demostrado). *Sea (B_t) un MBE con filtración natural $\mathcal{F}_t^B$. Para $0 \leq s \leq t$, el incremento $B_t - B_s$ es independiente de $\mathcal{F}_s^B$, y por tanto $E[B_t - B_s \mid \mathcal{F}_s^B] = E[B_t - B_s]$.*

Teorema 5.8.21. *Sea $(B_t)_{t\geq 0}$ un MBE. Entonces tanto (B_t) como $(B_t^2 - t)$ son $\mathcal{F}_t^B$-martingalas.*

Observación (Idea de la demostración). Ambos procesos son adaptados (son funciones medibles de B_t) e integrables ($E[|B_t|] \leq \sqrt{2t/\pi} < \infty$ y $E[|B_t^2 - t|] \leq 2t < \infty$). Para la propiedad de martingala de B_t: usando el lema anterior,

$$E[B_t \mid \mathcal{F}_s^B] = E[B_t - B_s \mid \mathcal{F}_s^B] + B_s = E[B_t - B_s] + B_s = 0 + B_s = B_s.$$

Para $B_t^2 - t$, escribiendo $B_t = (B_t - B_s) + B_s$ y usando de nuevo la independencia del incremento respecto a $\mathcal{F}_s^B$:

$$E[B_t^2 \mid \mathcal{F}_s^B] = E[(B_t - B_s)^2] + B_s^2 + 2B_s \underbrace{E[B_t - B_s \mid \mathcal{F}_s^B]}_{=0} = (t - s) + B_s^2,$$

luego $E[B_t^2 - t \mid \mathcal{F}_s^B] = B_s^2 - s$.

Teorema 5.8.22 (Caracterización de Lévy del Movimiento Browniano). *Sea $(X_t)_{t\geq 0}$ un proceso real con filtración natural $(\mathcal{F}_t^X)$. Si:*

(i) $X_0 = 0$ *c.s.,*

(ii) $t \mapsto X_t$ *es continua,*

(iii) tanto (X_t) *como* $(X_t^2 - t)$ *son* $\mathcal{F}_t^X$*-martingalas,*

entonces $(X_t)_{t\geq 0}$ *es un Movimiento Browniano Estándar.*

Observación. Este teorema es muy potente en la práctica: **es el recíproco** del resultado anterior, y permite **verificar** que un proceso dado (por ejemplo, construido como solución de una ecuación diferencial estocástica) es un Movimiento Browniano sin tener que comprobar directamente la definición Gaussiana, sino solo estas tres condiciones (más fáciles de chequear en la práctica).

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:** $\pi(\underline{s}) = \dfrac{e^{-\beta H(\underline{s})}}{Z(\beta)}$, 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 6.3.1 (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

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

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

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

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:

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

Por tanto:

$$\pi(x)P(x, y) = \pi(x)q(x, y) \cdot 1 = \pi(x)q(x, y),$$

$$\pi(y)P(y, x) = \pi(y)q(y, x) \cdot \frac{\pi(x)q(x, y)}{\pi(y)q(y, x)} = \pi(x)q(x, y).$$

como ambos lados coinciden, entonces se cumple el balance detallado. □

Corolario 6.4.1.1. *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 6.5.2 (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:*

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

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 + Irreducibleza 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 convergecia 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 6.6.3 (Recordatorio: Teorema Ergódico / Corolario de promedios temporales). *Si la cadena es positivo recurrente con invariante π, entonces para toda función acotada $f : S \to \mathbb{R}$:*

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

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

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 empecemos, 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` $\pi \to$ `Cadena de Markov diseñada (Metropolis-Hastings)` $\to$ `Simulación larga` $\to$ `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.