Lección n°4: Ley de los grandes números para cadenas de Markov AMARUN 2025

En las lecciones anteriores estudiamos las propiedades básicas de las cadenas de Markov: la irreductibilidad, la recurrencia y la existencia de medidas invariantes. En esta parte veremos cómo esas nociones permiten describir el comportamiento a largo plazo de la cadena. Mostraremos que, bajo ciertas condiciones, las medias temporales de las observaciones convergen hacia su valor esperado con respecto a la medida invariante: es la **ley de los grandes números** o **teorema ergódico** para cadenas de Markov. Terminaremos analizando la *aperiodicidad* y la *convergencia al equilibrio*, que garantizan que la cadena se estabiliza en su distribución estacionaria.

1. Ley de los grandes números y teorema ergódico

Lema 1.1 (Número de pasadas y tiempo de retorno)**.** *Sea* $(X_n)_{n\in\mathbb{N}}$ *una* CM(E) *irreducible, y* $x \in E$*, entonces la frecuencia de pasadas en* x *antes del tiempo* n *verifica:*

$$\frac{\sum_{k=0}^{n-1} \mathbb{1}_{\{X_k=x\}}}{n} \xrightarrow[n\to\infty]{c.s.} \frac{1}{\mathbb{E}^x(T_x^*)} \in [0,1]$$

Demostración. Notemos $N_x^n := \sum_{k=0}^{n-1} \mathbb{1}_{\{X_k=x\}}$ el número de pasadas en x antes del tiempo n. Si x es transitorio, entonces c.s. el número de pasadas en x es finito, y por tanto $N_x^n/n \to 0$, mientras que $\mathbb{E}^x(T_x^*) = \infty$ pues T_x^* no es finito c.s.

Si x es recurrente, como el tiempo de alcance de x es finito c.s., podemos suponer que la cadena parte de x. Además, c.s., la sucesión estrictamente creciente $(T_x^n)_{n\geq 1}$ de los tiempos de retorno sucesivos en x está bien definida y tiende a ∞. Por convención, tenemos que $T_x^0 = 0$ y que $T_x^1 = T_x^*$. Las definiciones de N_x^n y T_x^n dan $T_x^{N_x^n} \leq n < T_x^{N_x^n+1}$, de donde:

$$\frac{T_x^{N_x^n}}{N_x^n} \leq \frac{n}{N_x^n} \leq \frac{T_x^{N_x^n+1}}{N_x^n+1}\frac{N_x^n+1}{N_x^n}$$

Por la propiedad de Markov fuerte, $(T_x^{n+1} - T_x^n)_{n\geq 1}$ está formada por v.a. i.i.d. de ley $\mathcal{L}(T_x \mid X_0 = x) = \mathcal{L}(T_x^2 - T_x^1)$, de donde:

$$\frac{T_x^n}{n} = \frac{T_x^1 + T_x^2 - T_x^1 + \cdots + T_x^n - T_x^{n-1}}{n} \xrightarrow[n\to\infty]{\text{c.s.}} \mathbb{E}^x(T_x^1)$$

por la ley fuerte de los grandes números. Por otra parte, la recurrencia y la irreductibilidad implican que

$$N_x^n \to \infty \text{ c.s.}$$

Estas dos convergencias y lo anterior implican que $n/N_x^n \to \mathbb{E}^x(T_x^1)$ c.s. □

Teorema 1.2 (Teorema ergódico de Chacon-Ornstein-Hopf [1])**.** *Sea* $(X_n)_{n\in\mathbb{N}}$ *una cadena de Markov sobre* E *recurrente irreducible, de ley inicial cualquiera. Entonces:*

(i) Para toda medida invariante μ *y para todas funciones* $f, g : E \to \mathbb{R}$ μ*-integrables con* $\mu g \neq 0$*, se tiene:*

$$\frac{f(X_0) + \cdots + f(X_{n-1})}{g(X_0) + \cdots + g(X_{n-1})} \xrightarrow[n\to\infty]{c.s.} \frac{\sum_{z\in E} f(z)\mu(z)}{\sum_{z\in E} g(z)\mu(z)} = \frac{\mu f}{\mu g} \in \mathbb{R}$$

(ii) Si la cadena es recurrente nula, entonces para toda función $f : E \to \mathbb{R}$ μ*-integrable:*

$$\frac{f(X_0) + \cdots + f(X_{n-1})}{n} \xrightarrow[n\to\infty]{c.s.} 0$$

(iii) Si la cadena es recurrente positiva, de ley invariante μ, entonces para toda $f : E \to \mathbb{R}$ μ-integrable:

$$\frac{f(X_0) + \cdots + f(X_{n-1})}{n} \xrightarrow[n\to\infty]{c.s.} \sum_{z\in E} f(z)\mu(z) \in \mathbb{R}$$

Demostración. Razonemos condicionalmente a $\{X_0 = x\}$, donde $x \in E$ es arbitrario. Los conjuntos c.s. construidos dependerán de x, pero su intersección sobre $x \in E$ será también c.s. pues E es a lo más numerable.

Para establecer (i), consideremos primero el caso $g = \mathbb{1}_{\{x\}}$ y $f \geq 0$. Como el núcleo es irreducible recurrente, el Teorema de *medidas invariantes para núcleos irreducibles (ver lección 3)* indica que todas las medidas invariantes son proporcionales, basta considerar una, por ejemplo $\mu = \mu_x$, la medida invariante que vale 1 en x estudiada en el lema de *numero de veces de pasada antes de regresar (ver lección 3)*. Procediendo similar a la demostración del Lema 1.1 se tiene entonces, con las mismas notaciones, con $T := T_x^*$ y $N^n := N_x^n$ que

$$\frac{f(X_0) + \cdots + f(X_{T^{N^n}-1})}{N^n} \leq \frac{f(X_0) + \cdots + f(X_n)}{N^n} \leq \frac{f(X_0) + \cdots + f(X_{T^{N^n+1}-1})}{N^n+1}\frac{N^n+1}{N^n}$$

La propiedad de Markov fuerte da $f(X_0) + \cdots + f(X_{T^m-1}) = Z_1 + \cdots + Z_m$ donde $(Z_k)_{k\geq 1}$ es una sucesión de v.a. i.i.d. Además $\mathbb{E}(Z_1) = \mathbb{E}(f(X_0) + \cdots + f(X_{T-1})) = \mu_x f$ por definición de μ_x y el teorema de Fubini-Tonelli. Como por otra parte $N^n \to \infty$ c.s., la ley fuerte de los grandes números para $(Z_k)_{k\geq 1}$ y el encuadramiento anterior dan:

$$R_n(f) := \frac{f(X_0) + \cdots + f(X_{T^{N^n}})}{N^n} \xrightarrow[n\to\infty]{\text{c.s.}} \mu_x f$$

Ahora, el resultado para todo $f \in L^1(\mu_x)$ se deduce poniendo $f = f^+ - f^-$ con $f^\pm := \text{máx}(0, \pm f)$. Finalmente, para todo $g \in L^1(\mu_x)$ tal que $\mu_x g \neq 0$, se tiene que $R_n(f)/R_n(g) \to \mu_x f/\mu_x g$ c.s. lo que es el resultado querido.

La propiedad (ii) y (iii) quedan como ejercicio. □

Definición 1.3 (Periodicidad)**.** Si $\mathbf{P}$ es un núcleo irreducible sobre E a lo más numerable, entonces existe un entero $d > 0$ llamado **período** y una partición $E = C_0 \cup \cdots \cup C_{d-1}$ tales que, con la convención $C_{nd+i} = C_i$, las propiedades siguientes se cumplen:

1. Si $\mathbf{P}^n(x, y) > 0$, entonces $(x, y) \in C_i \times C_{i+n}$ para algún i
2. Para todos $x, y \in C_i$, y todo i, existe r tal que $\mathbf{P}^{nd}(x, y) > 0$ para todo $n \geq r$
3. Para todo $z \in E$, $d = \text{mcd}\{n \geq 1 : \mathbf{P}^n(z, z) > 0\}$

En particular, el núcleo $\mathbf{P}^d$ posee d clases irreducibles, que son $C_0, \ldots, C_{d-1}$. Se dice que $\mathbf{P}$ es **aperiódico** cuando $d = 1$.

Definición 1.4 (Distancia en variación total)**.** Para medidas de probabilidad μ, ν sobre E:

$$\mathrm{d_{VT}}(\mu, \nu) = \frac{1}{2} \sup_{\|f\|_\infty \leq 1} |(\mu - \nu)f|$$

El prefactor $1/2$ asegura que los valores tomados están en $[0, 1]$, el valor 1 siendo alcanzado cuando μ y ν tienen soportes disjuntos. Es claro que en el conjunto de las medidas de probabilidad sobre E, la convergencia para esta distancia implica la convergencia estrecha, es decir la convergencia en ley.

Lema 1.5 (Contracción markoviana)**.** *Sea $\mathbf{P}$ un núcleo de transición sobre E. Entonces para todas medidas de probabilidad ν_1 y ν_2 sobre E:*

$$\mathrm{d_{VT}}(\nu_1\mathbf{P}, \nu_2\mathbf{P}) \leq \mathrm{d_{VT}}(\nu_1, \nu_2)$$

En particular, si μ es una ley invariante para $\mathbf{P}$ entonces para toda medida de probabilidad ν sobre E y todo $n \in \mathbb{N}$:

$$\mathrm{d_{VT}}(\nu\mathbf{P}^{n+1}, \mu) \leq \mathrm{d_{VT}}(\nu\mathbf{P}^n, \mu)$$

Demostración. Si $f : E \to [-1, +1]$, entonces $\mathbf{P}f : E \to [-1, +1]$, de donde:

$$|(\nu_1 - \nu_2)\mathbf{P}f| \leq \sup_{\|f\|_\infty \leq 1} |(\nu_1 - \nu_2)f| = 2\mathrm{d_{VT}}(\nu_1, \nu_2)$$

□

2. Recurrencia e invarianza en espacios de estado finitos

Teorema 2.1 (Recurrencia e invarianza en espacio de estado finito)**.** *Sea* $(X_n)_{n\in\mathbb{N}}$ *una* $\mathrm{CM}(E, v, \mathbf{P})$ *sobre* E *finito,* E_R *los estados recurrentes y* E_T *los estados transitorios. Entonces:*

1. *El conjunto* E_R *es no vacío, y todos los estados recurrentes son recurrentes positivos. Si* $\mathbf{P}$ *es irreducible, entonces es irreducible recurrente positivo.*

2. *La cadena es casi seguramente capturada por una clase de recurrencia:* $\mathbb{P}(T_{E_R} < \infty) = 1$*. Cuando todos los estados recurrentes son absorbentes, la cadena termina c.s. por ser constante.*

3. *Si* $(\mathbf{P}^n(x, \cdot))_{n\in\mathbb{N}}$ *converge, su límite es una ley de probabilidad invariante.*

4. *El espacio propio de* $(\mathbf{P}_{R_i\times R_i})^\top$ *asociado al valor propio 1 es de dimensión 1, y contiene un único vector a componentes positivas cuya suma vale 1, que es precisamente* μ_i*.*

5. *La matriz* $\mathbf{P}$ *es irreducible si y sólo si* $\mathbf{P}$ *admite una única ley invariante.*

6. *Una clase irreducible es cerrada si y sólo si es recurrente. Las clases irreducibles de las cuales ninguna flecha sale son recurrentes, mientras que las otras son transitorias. Un estado* x *es transitorio si y sólo si existe un estado* y *tal que* $x \to y$ *y* $y \nrightarrow x$*. La descomposición de* E *no depende más que de la conectividad del esqueleto del grafo de transiciones, es decir, de las flechas pero no de su peso.*

Demostración. Para probar 1. supongamos por contradicción que $E_R = \emptyset$, es decir, todos los estados son transitorios. Para cada estado $x \in E$, como es transitorio, tenemos $\mathbb{E}^x[N_x] < \infty$, donde N_x es el número total de visitas a x. Pero entonces:

$$\sum_{x\in E} \mathbb{E}^x[N_x] < \infty$$

Sin embargo, para cualquier ley inicial, la cadena debe visitar estados infinitas veces (pues el proceso continúa para siempre). Como E es finito, por el principio del palomar, al menos un estado debe ser visitado infinitas veces con probabilidad positiva, lo que contradice la transitoriedad. Más formalmente, si $N = \sum_{x\in E} N_x$ es el número total de visitas, tenemos $N = \infty$ casi seguramente, pero $\mathbb{E}[N] = \sum_{x\in E} \mathbb{E}[N_x] < \infty$ si todos los estados son transitorios, lo que es una contradicción.

Ahora, si x es recurrente, entonces su clase de comunicación $C(x)$ es finita (pues E es finito) y cerrada. En una clase finita y cerrada, todos los estados son recurrentes positivos.

Si $\mathbf{P}$ es irreducible, entonces E es una única clase de comunicación, que por lo anterior debe ser recurrente positiva.

Para demostrar 2., sea $x \in E$ cualquiera. Consideremos el número total de visitas a estados transitorios:

$$N_{E_T} = \sum_{n=0}^{\infty} \mathbb{1}_{\{X_n\in E_T\}}$$

Si la cadena no fuera capturada por E_R, entonces $N_{E_T} = \infty$ con probabilidad positiva. Pero para cada $y \in E_T$, como y es transitorio, tenemos $\mathbb{E}^x[N_y] < \infty$. Como E_T es finito:

$$\mathbb{E}^x[N_{E_T}] = \sum_{y\in E_T} \mathbb{E}^x[N_y] < \infty$$

Lo que implica que $N_{E_T} < \infty$ casi seguramente. Por tanto, casi seguramente, la cadena abandona E_T para siempre después de un tiempo finito, es decir, $T_{E_R} < \infty$ casi seguramente.

Si todos los estados recurrentes son absorbentes, entonces una vez que la cadena alcanza E_R, permanece en ese estado para siempre, por lo que la cadena termina siendo constante.

Para demostrar el tercer punto, supongamos que para cada $x \in E$, $\mathbf{P}^n(x, \cdot) \to \pi_x$ cuando $n \to \infty$, donde π_x es una medida de probabilidad. Entonces, para todo $y \in E$:

$$\mathbf{P}^{n+1}(x, y) = \sum_{z\in E} \mathbf{P}^n(x, z)\mathbf{P}(z, y) \to \sum_{z\in E} \pi_x(z)\mathbf{P}(z, y) = (\pi_x\mathbf{P})(y)$$

Por otro lado, $\mathbf{P}^{n+1}(x,y) \to \pi_x(y)$. Por lo tanto, $\pi_x = \pi_x \mathbf{P}$, es decir, π_x es invariante.

Para la cuarta Proposición, para una clase de recurrencia R_i, la matriz $\mathbf{P}_{R_i \times R_i}$ es irreducible y recurrente positiva. Por el teorema de Perron-Frobenius para matrices estocásticas irreducibles, el valor propio 1 es simple y el espacio propio izquierdo asociado es unidimensional, generado por un vector positivo que puede normalizarse para que sea una distribución de probabilidad. Este vector es precisamente μ_i.

6. y 7. se dejan como ejercicio. □

Teorema 2.2 (Espectro y periodicidad)**.** *Sea* $\mathbf{P}$ *irreducible sobre* E *finito. Entonces:*

1. *El período* d *siempre verifica* $d \leq |E|$.
2. *El espectro de* $\mathbf{P}$ *está incluido en* $\{z \in \mathbb{C} : |z| \leq 1\}$*, es invariante por* $z \mapsto \bar{z}$*, y contiene a 1.*
3. *Las siguientes propiedades son equivalentes:*
   a) $\mathbf{P}$ *es aperiódico.*
   b) *Existe un entero* r *tal que* $\mathbf{P}^n(x,y) > 0$ *para todos* $x, y \in E$ *y todo* $n \geq r$.
   c) *1 es el único valor propio de* $\mathbf{P}$ *de módulo 1.*
4. *Si* $\mathbf{P}$ *tiene período* d*, entonces su espectro contiene las raíces de la unidad* $\{\mathrm{e}^{2\pi \mathrm{k}\mathrm{i}/d} : 0 \leq k \leq d-1\}$.
5. *Si* $\mathbf{P}$ *es reversible, entonces su espectro es real.*

Demostración. Para verificar 2, notemos que el vector $(1,1,\ldots,1)^\top$ es vector propio derecho con valor propio 1. Ahora, sea λ un valor propio de $\mathbf{P}$ con vector propio $v \neq 0$. Sea x tal que $|v_x| = \text{máx}_i |v_i| > 0$. Entonces:

$$|\lambda||v_x| = |(\mathbf{P}v)_x| = \left|\sum_{y \in E} \mathbf{P}(x,y) v_y\right| \leq \sum_{y \in E} \mathbf{P}(x,y)|v_y| \leq |v_x| \sum_{y \in E} \mathbf{P}(x,y) = |v_x|$$

Luego $|\lambda| \leq 1$.

Por otro lado, como $\mathbf{P}$ es real, si λ es valor propio, entonces $\overline{\lambda}$ también lo es.

Verifiquemos las tres proposiciones equivalentes:

<u>(a) ⇒ (b):</u> Por el teorema de descomposición cíclica, si $\mathbf{P}$ es aperiódico ($d = 1$), entonces para cada $x \in E$ existe n_x tal que para todo $n \geq n_x$, $\mathbf{P}^n(x,x) > 0$. Sea $N_1 = \text{máx}_{x \in E}\, n_x$.

Además, por irreducibilidad, para cada par (x,y) existe $m_{x,y}$ tal que $\mathbf{P}^{m_{x,y}}(x,y) > 0$. Sea $N_2 = \text{máx}_{x,y \in E}\, m_{x,y}$.

Entonces, para $n \geq N_1 + N_2$ y para cualquier par (x,y), tenemos:

$$\mathbf{P}^n(x,y) \geq \mathbf{P}^{m_{x,y}}(x,y)\mathbf{P}^{n-m_{x,y}}(y,y) > 0$$

pues $n - m_{x,y} \geq N_1$. Luego (b) se cumple con $r = N_1 + N_2$.

<u>(b) ⇒ (c):</u> Supongamos que existe λ con $|\lambda| = 1$ y $\lambda \neq 1$ tal que $\mathbf{P}v = \lambda v$ para algún $v \neq 0$. Entonces $\mathbf{P}^n v = \lambda^n v$ para todo n.

Por (b), para $n \geq r$, $\mathbf{P}^n(x,y) > 0$ para todo x, y. Consideremos el vector $|v|$ con componentes $|v_i|$. Para cada $x \in E$:

$$|v_x| = |\lambda^n v_x| = |(\mathbf{P}^n v)_x| = \left|\sum_{y \in E} \mathbf{P}^n(x,y) v_y\right| \leq \sum_{y \in E} \mathbf{P}^n(x,y)|v_y| = (\mathbf{P}^n|v|)_x$$

Pero también tenemos:

$$\sum_{x \in E} |v_x| = \sum_{x \in E} (\mathbf{P}^n|v|)_x = \sum_{y \in E} |v_y|$$

pues $\mathbf{P}^n$ es estocástica. Por tanto, debemos tener igualdad en la desigualdad triangular para cada x, lo que implica que existe una constante α_n (dependiente de n) tal que para todo y, $v_y = \alpha_n |v_y|$ (es decir, todos los v_y tienen el mismo argumento).

Pero entonces $v = \alpha_n |v|$, y por tanto:

$$\mathbf{P}^n v = \alpha_n \mathbf{P}^n |v| = \alpha_n |v| = v$$

pues $\mathbf{P}^n|v| = |v|$ (ya que $\mathbf{P}^n$ es estocástica). Pero $\mathbf{P}^n v = \lambda^n v$, luego $\lambda^n v = v$, y como $v \neq 0$, tenemos $\lambda^n = 1$ para todo $n \geq r$. Esto implica $\lambda = 1$, contradicción.

(c) $\Rightarrow$ (a): Supongamos que $\mathbf{P}$ tiene período $d > 1$. Consideremos la descomposición cíclica $E = C_0 \cup \cdots \cup C_{d-1}$. Sea μ_0 la única ley invariante de la cadena restringida a C_0 con núcleo $\mathbf{P}^d|_{C_0}$ (que es irreducible y aperiódica en C_0).

Definamos el vector $v \in \mathbb{C}^E$ por:

$$v(x) = \mathrm{e}^{2\pi k \mathrm{i}/d} \mu_0(x) \quad \text{si } x \in C_k$$

donde extendemos μ_0 a las otras clases mediante $\mu_0 \mathbf{P}^k$. Entonces, para $x \in C_k$:

$$(\mathbf{P}v)(x) = \sum_{y \in E} \mathbf{P}(x,y) v(y) = \sum_{y \in C_{k+1}} \mathbf{P}(x,y) \mathrm{e}^{2\pi(k+1)\mathrm{i}/d} \mu_0(y) = \mathrm{e}^{2\pi k \mathrm{i}/d} \mathrm{e}^{2\pi \mathrm{i}/d} \sum_{y \in C_{k+1}} \mathbf{P}(x,y) \mu_0(y)$$

Pero $\sum_{y \in C_{k+1}} \mathbf{P}(x,y)\mu_0(y) = (\mu_0 \mathbf{P})(x) = \mu_0(x)$ por la estructura cíclica y la invarianza de μ_0 para $\mathbf{P}^d$. Luego,

$$(\mathbf{P}v)(x) = \mathrm{e}^{2\pi \mathrm{i}/d} v(x)$$

Por tanto, $\mathrm{e}^{2\pi \mathrm{i}/d}$ es valor propio de $\mathbf{P}$, de módulo 1 y diferente de 1 si $d > 1$, contradiciendo (c).

1., 4. y 5. se dejan como ejercicio. $\square$

3. Convergencia subexponencial al equilibrio y acoplamiento

Teorema 3.1 (Convergencia subexponencial al equilibrio)**.** *Sea $\mathbf{P}$ un núcleo de transición sobre E finito, irreducible y aperiódico, y sea μ su ley invariante. Entonces existe un real $0 < \rho < 1$ y un entero $r > 0$ que dependen de $\mathbf{P}$ tales que*

$$\max_{x \in E} \|\mathbf{P}^n(x, \cdot) - \mu\|_{\mathrm{VT}} \leq \rho^{\lfloor n/r \rfloor} \xrightarrow[n \to \infty]{} 0.$$

Demostración. Como $\mathbf{P}$ es irreducible y aperiódico sobre E finito, existe un entero r tal que:

$$\varepsilon := \min_{x,y \in E} \mathbf{P}^r(x,y) > 0$$

En particular, para todo $x, y \in E$:

$$\mathbf{P}^r(x,y) \geq \varepsilon \geq \varepsilon \mu(y)$$

pues $\mu(y) > 0$ para todo y (por irreducibilidad). Sea $\rho = 1 - \varepsilon$.

Definamos la matriz $\mathbf{M}$ cuyas filas son todas iguales a μ. Entonces podemos escribir:

$$\mathbf{P}^r = (1-\rho)\mathbf{M} + \rho \mathbf{Q}$$

donde $\mathbf{Q} = \frac{1}{\rho}(\mathbf{P}^r - (1-\rho)\mathbf{M})$. Notemos que $\mathbf{Q}$ es una matriz estocástica, pues:

- Sus entradas son no negativas: $\mathbf{P}^r(x,y) - (1-\rho)\mu(y) \geq \varepsilon - (1-\rho)\mu(y) = \rho\mu(y) \geq 0$
- Suma por filas 1: $\sum_y \mathbf{Q}(x,y) = \frac{1}{\rho}(\sum_y \mathbf{P}^r(x,y) - (1-\rho)\sum_y \mu(y)) = \frac{1}{\rho}(1 - (1-\rho)) = 1$

Además, $\mu\mathbf{Q} = \mu$ porque $\mu\mathbf{P}^r = \mu$ y $\mu\mathbf{M} = \mu$.

Ahora probemos por inducción que para todo $k \geq 1$:

$$\mathbf{P}^{rk} = (1-\rho^k)\mathbf{M} + \rho^k \mathbf{Q}^k$$

Para $k = 1$ es la definición. Supongamos cierto para k, entonces:

$$\mathbf{P}^{r(k+1)} = \mathbf{P}^{rk}\mathbf{P}^r = \left((1-\rho^k)\mathbf{M} + \rho^k \mathbf{Q}^k\right)((1-\rho)\mathbf{M} + \rho\mathbf{Q})$$

Desarrollando:

$$\mathbf{P}^{r(k+1)} = (1-\rho^k)(1-\rho)\mathbf{MM} + (1-\rho^k)\rho\mathbf{MQ} + \rho^k(1-\rho)\mathbf{Q}^k\mathbf{M} + \rho^{k+1}\mathbf{Q}^k\mathbf{Q}$$

Notemos que:

- $\mathbf{MM} = \mathbf{M}$ (pues cada fila de $\mathbf{MM}$ es $\mu\mathbf{M} = \mu$)
- $\mathbf{MQ} = \mathbf{M}$ (pues cada fila de $\mathbf{MQ}$ es $\mu\mathbf{Q} = \mu$)
- $\mathbf{Q}^k\mathbf{M} = \mathbf{M}$ (pues cada fila de $\mathbf{Q}^k\mathbf{M}$ es $\mu\mathbf{Q}^k\mathbf{M} = \mu\mathbf{M} = \mu$)

Luego:

$$\mathbf{P}^{r(k+1)} = [(1-\rho^k)(1-\rho) + (1-\rho^k)\rho + \rho^k(1-\rho)]\mathbf{M} + \rho^{k+1}\mathbf{Q}^{k+1}$$

Simplificando el coeficiente de $\mathbf{M}$:

$$(1-\rho^k)(1-\rho) + (1-\rho^k)\rho + \rho^k(1-\rho) = 1 - \rho^k - \rho + \rho^{k+1} + \rho - \rho^{k+1} + \rho^k - \rho^{k+1} = 1 - \rho^{k+1}$$

Por tanto:

$$\mathbf{P}^{r(k+1)} = (1-\rho^{k+1})\mathbf{M} + \rho^{k+1}\mathbf{Q}^{k+1}$$

Ahora, para cualquier $n \in \mathbb{N}$, escribamos $n = kr + i$ con $0 \leq i < r$. Entonces:

$$\mathbf{P}^n = \mathbf{P}^{kr}\mathbf{P}^i = \Big((1-\rho^k)\mathbf{M} + \rho^k\mathbf{Q}^k\Big)\,\mathbf{P}^i = (1-\rho^k)\mathbf{MP}^i + \rho^k\mathbf{Q}^k\mathbf{P}^i$$

Pero $\mathbf{MP}^i = \mathbf{M}$ (pues $\mu\mathbf{P}^i = \mu$). Luego:

$$\mathbf{P}^n - \mathbf{M} = \rho^k\left(\mathbf{Q}^k\mathbf{P}^i - \mathbf{M}\right)$$

Para cualquier $x, y \in E$:

$$|\mathbf{P}^n(x,y) - \mu(y)| = \rho^k|(\mathbf{Q}^k\mathbf{P}^i - \mathbf{M})(x,y)| \leq \rho^k\left(\mathbf{Q}^k\mathbf{P}^i(x,y) + \mu(y)\right)$$

Sumando sobre y:

$$\sum_{y\in E}|\mathbf{P}^n(x,y) - \mu(y)| \leq \rho^k\left(\sum_{y\in E}\mathbf{Q}^k\mathbf{P}^i(x,y) + \sum_{y\in E}\mu(y)\right) = 2\rho^k$$

Luego:

$$\|\mathbf{P}^n(x,\cdot) - \mu\|_{\mathrm{VT}} = \frac{1}{2}\sum_{y\in E}|\mathbf{P}^n(x,y) - \mu(y)| \leq \rho^k = \rho^{\lfloor n/r\rfloor}$$

□

Referencias

[1] Djalil Chafaï et al. *Processus stochastiques: Ébauche de notes de cours de M1*. Département de mathématiques et applications (DMA), École normale supérieure – PSL, Paris. Équipe pédagogique: Djalil Chafaï (cours) y Alexis Metz–Donnadieu (TD), revisado por Laëtitia Comminges y Arnaud Guyad. 2024.