UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

REPRESENTACIONES DE GRUPOS Y CONVERGENCIA AL ESTADO ESTACIONARIO DE CADENAS DE MARKOV

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

MATEMÁTICO

P R E S E N T A :

ALONSO BARANDA LOZADA

DIRECTORA DE TESIS:

DRA. ELIANE REGINA RODRIGUES

2012

Datos del alumno
Apellido paterno
Apellido materno
Nombre
Teléfono
Universidad Nacional Autónoma de México
Facultad de Ciencias
Carrera
Número de cuenta

Datos del tutor
Grado
Nombres
Apellido paterno
Apellido materno

Datos del sinodal 1
Grado
Nombre
Apellido paterno
Apellido materno

Datos del sinodal 2
Grado
Nombres
Apellido paterno
Apellido materno

Datos del sinodal 3
Grado
Nombre
Apellido paterno
Apellido materno

Datos del sinodal 4
Grado Dr
Nombre
Apellido paterno
Apellido materno

Datos del trabajo escrito
Título
Número de páginas
Año

Datos del alumno
Baranda
Lozada
Alonso
56558741
Universidad Nacional Autónoma de México
Facultad de Ciencias
Matemáticas
304519033

Datos del tutor
Dra
Eliane Regina
Rodrigues
Rodrigues

Datos del sinodal 1
Dra
Eugenia
O'Reilly
Regueiro

Datos del sinodal 2
Dr
Gerónimo Francisco
Uribe
Bravo

Datos del sinodal 3
Dra
Ana
Meda
Guardiola

Datos del sinodal 4
Dr
Juan
González
Hernández

Datos del trabajo escrito
Representaciones de grupos y convergencia
al estado estacionario de cadenas de Markov
73
2012

Agradecimientos

Haber cursado la licenciatura de matemáticas en la UNAM, es en gran parte, lo que ha hecho de mí la persona que soy. Estoy enormemente agradecido por todo lo que me ha dado nuestra máxima casa de estudios. México necesita de matemáticos para resolver muchos de los problemas actuales que tenemos. Exhorto a cualquier persona con fuerza de voluntad e interesada en la razón, a estudiar matemáticas. No se arrepentirán.

Jamás hubiera logrado concluir la carrera sin el amor y la ayuda de mi madre Teresa ¡Te quiero mami!

Quiero agradecer a mi padre Joaquín por todos sus valiosos consejos, verás que todo saldra bien y que pronto te recuperarás.

A mi tutora, la Dra. Eliane Rodrigues, por haberme aceptado como alumno y haberme propuesto este estupendo tema. ¡Gracias Eliane!

A mis sinodales, Gerónimo Uribe Bravo, Eugenia O'Reilly-Regueiro, Ana Meda Guardiola y Juan González Hernández, quienes mejoraron el contenido de este trabajo de una manera asombrosa.

Agradezco también a todos mis amigos, profesores, compañeros del equipo de fútbol y demás conocidos, por los muchos gratos momentos tanto dentro como fuera de la escuela. En particular agradezco a Lau, Paria, Alex, Chars, Alan, Johnatan, José Luis, Ofelia, Álvaro, Teo, Pato, Juan, Germán, Fernando, Manolo, Andrea, Panda, Viry, Naim, Caro, Oli, Esther, Eduardo, Adolfo, José, Jaime, Poeta, Osvaldo, Alejandra, Ameyalli, Ángel, Chico, Cony, Cruz & la Maestra, etc.

Dedico esta tesis a todas aquellas personas que no han podido completar los trámites de titulación ¡Ánimo!

II

Índice general

Introducción

Es razonable pensar que para encontrar propiedades sobre procesos estocásticos definidos en estructuras algebraicas, es conveniente estudiar éstas por separado. A esas estructuras las podemos estudiar de muchas maneras, en particular, a través de ciertos homomorfismos. En nuestro caso las estructuras algebraicas serán los grupos finitos. El estudio lo haremos a través de las representaciones lineales de grupos finitos (las cuales son homomorfismos), y los procesos estocásticos serán las caminatas aleatorias en un grupo finito G generadas por cualquier medida de probabilidad Q definida en G.

En un curso de licenciatura de procesos estocásticos se demuestra que cualquier cadena de Markov (cualquier caminata aleatoria es una cadena de Markov) aperiódica, irreducible y positiva recurrente tiene una única distribución estacionaria. Si además suponemos que el espacio de estados es finito entonces podemos asegurar que, sin importar cuál sea la distribución inicial X_0, la distribución de la cadena convergerá a la distribución estacionaria cuando $n \to \infty$. Muchas veces dicha distribución estacionaria es la distribución uniforme. Este trabajo es un intento de obtener cotas para estimar el tiempo necesario para alcanzar ese estado estacionario, es decir, acotamos el tiempo requerido para que la cadena tenga una distribución aproximadamente uniforme.

En el capítulo 1 se aborda la teoría elemental de las representaciones lineales de grupos finitos. El capítulo 2 es de referencia, y en él se aborda brevemente el tema de cadenas de Markov. En el capítulo 3 estudiamos las caminatas aleatorias en $\mathbb{Z}$ y finalizamos el capítulo dando un teorema sobre un criterio de recurrencia para caminatas aleatorias en $\mathbb{Z}$. En el capítulo 4 (que es la aplicación del capítulo 1) se estudian las caminatas aleatorias definidas en un grupo finito G generadas por cualquier medida de probabilidad Q definida en G. Además se obtienen cotas (superiores e inferiores) acerca del tiempo que le toma a la cadena aproximarse al estado estacionario. En el apéndice A se introducen los tiempos de paro y las martingalas. El apéndice B es acerca de teoría de grupos y álgebra lineal. El trabajo finaliza con un glosario de simbología, seguido por la bibliografía.

Capítulo 1

Representaciones lineales

En las secciones 1.1 y 1.2 de este capítulo desarrollaremos la teoría elemental de las representaciones lineales de grupos finitos. Introduciremos la noción de carácter de una representación y se verá qué relación guarda con la representación, además de otros resultados. En la sección 1.3 expondremos los conceptos de transformación de Fourier asociada a una representación y el de convolución de dos medidas de probabilidad definidas en un grupo finito. De esto último se obtendrán algunos resultados que serán de vital importancia en el capítulo 4. Los requisitos para este capítulo son: teoría de grupos finitos, álgebra lineal y probabilidad, todos a un nivel elemental. La bibliográfia de este capítulo es [1], [7] y [8].

1.1. Representaciones lineales de grupos finitos

Definición 1.1. Sean G un grupo, V un espacio vectorial y $\mathrm{GL}(V)$ el conjunto de automorfismos de V. Una **representación lineal de G en V** es un homomorfismo $\rho : G \to \mathrm{GL}(V)$. A la dimensión de V la llamaremos el **grado de la representación** y la denotaremos por d_ρ.

Así, una representación lineal es un homomorfismo de G en el conjunto de automorfismos de V. Siempre consideraremos grupos finitos y espacios vectoriales complejos o reales de dimensión finita a menos que sea especificado.

Observación. Denote por $\mathrm{GL}(n, F)$ al grupo de matrices invertibles de $n \times n$ sobre F. A veces convendrá trabajar con matrices. Teniendo en cuenta que $\mathrm{GL}(V) \simeq \mathrm{GL}(n, F)$ y tomando una base β de V podemos considerar el homomorfismo de grupos $\rho' : G \to \mathrm{GL}(n, F)$ dado por $\rho'(g) = [\rho(g)]_\beta$, que

por conveniencia, seguiremos llamando ρ. Es importante señalar que el homomorfismo depende de la base, y siempre que podamos utilizaremos bases canónicas.

Ejemplo 1.1. Sea G un grupo finito, consideremos $\rho : G \to \mathrm{GL}(1, \mathbb{C})$ dada por $\rho(g) = 1$ para todo $g \in G$, entonces ρ es la representación trivial de G.

Ejemplo 1.2. Sea $\rho : G \to \mathrm{GL}(V)$ una representación lineal de G en V y sea $H \leq G$ cualquier subgrupo de G. Entonces $\eta = \rho|_H$ es una representación lineal de H en V. Note que $\eta : H \to \mathrm{GL}(V)$. Si h_0, $h_1 \in H$, es claro que $h_0 h_1 \in H$ y

$$\eta(h_0 h_1) = \rho(h_0 h_1) = \rho(h_0)\rho(h_1) = \eta(h_0)\eta(h_1).$$

Por lo tanto $\eta = \rho|_H$ es una representación lineal de H en V.

Ejemplo 1.3. Sea G un grupo finito de orden n, entonces una representación compleja de G de grado $d_\rho = 1$ es un homomorfismo $\rho : G \to \mathrm{GL}(1, \mathbb{C})$ tal que para todo $g \in G$ se tiene $\rho(g)^n = \rho(g^n) = 1$, así $\rho(g)$ es una de las n raíces n-ésimas de la unidad, es decir, $\rho(g) = (e^{\frac{2\pi i}{n}})^{m_g}$, donde m_g es un entero no negativo que depende de g.

Ejemplo 1.4. Sean $\mathbb{Z}_3$ el grupo de enteros módulo 3 y $\rho : \mathbb{Z}_3 \to \mathrm{GL}(\mathbb{R}^2)$ dada por

$$\rho([n]) = \begin{cases} Id & \text{si } [n] = [0], \\ rot(\frac{2\pi}{3}) & \text{si } [n] = [1], \\ rot(\frac{4\pi}{3}) & \text{si } [n] = [2], \end{cases}$$

donde $rot(\theta)$ es la rotación por θ grados, entonces ρ es una representación de grado $d_\rho = 2$. Notemos que ni $rot(\frac{2\pi}{3})$ ni $rot(\frac{4\pi}{3})$ son diagonalizables.

Ejemplo 1.5. Sean $G = S_n$ (el grupo de todas las permutaciones en $\{1, 2 \ldots, n\}$), β la base canónica de $\mathbb{R}^n$, y $\rho : S_n \to \mathrm{GL}(n, \mathbb{R})$ dada por

$$\rho(\pi)(e_i) = e_{\pi(i)}.$$

Sean $\pi, \sigma \in S_n$ entonces tenemos que

$$\rho(\pi \circ \sigma)e_i = e_{\pi\circ\sigma(i)} = \rho(\pi)e_{\sigma(i)} = \rho(\pi)\rho(\sigma)e_i.$$

Así, ρ es una representación de grado $d_\rho = n$. Aterricemos este ejemplo para $n = 3$. Sean $\pi = (12)$, $\sigma = (23)$[1] ambas permutaciones en S_3. Entonces

$$\rho(\pi)e_1 = e_2,\ \rho(\pi)e_2 = e_1,\ \rho(\pi)e_3 = e_3;$$

[1]Escribiremos las permutaciones como producto de ciclos disjuntos, cuya forma es única, salvo por el orden de los productos. Los ciclos de longitud 1 los omitiremos. Así tenemos que $\pi(1) = 2$, $\pi(2) = 1$ y $\pi(3) = 3$, es decir $\pi = (12)(3) = (12)$. Mientras que $\pi\sigma(1) = 2$, $\pi\sigma(2) = 3$ y $\pi\sigma(3) = 1$, es decir, $\pi\sigma = (12)(23) = (123)$.

$$\rho(\sigma)e_1 = e_1,\ \rho(\sigma)e_2 = e_3,\ \rho(\sigma)e_3 = e_2.$$

De esta forma,

$$\rho(\pi) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix},\ \rho(\sigma) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Multiplicando las matrices obtenemos

$$\rho(\pi)\rho(\sigma) = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = \rho(123) = \rho(\pi\sigma)$$

como era de esperarse. Dicha representación es llamada **representación de permutaciones**.

Ejemplo 1.6. Sean G cualquier grupo finito de orden n, V espacio vectorial de dimensión n sobre $\mathbb{R}$ o $\mathbb{C}$, y $\beta = \{v_g\}_{g\in G}$ base de V, entonces $\rho : G \to \mathrm{GL}(V)$ dada por

$$\rho(h)v_g = v_{hg}$$

es una representación de grado $d_\rho = n$. Para cada $h \in G$, $\rho(h)$ simplemente permuta la base β. Verifiquemos que ρ es homomorfismo. Sean g, $h \in G$ y $v_k \in \beta$ entonces,

$$\rho(gh)v_k = v_{(gh)k} = v_{g(hk)} = \rho(g)v_{hk} = \rho(g)\rho(h)v_k.$$

ρ es llamada **representación regular** y la denotaremos por ρ_R.

Observación. Note que para todo $g \in G$, la matriz asociada a la base $\beta = \{v_g\}_{g\in G}$ es una matriz de ceros y unos. Abusando de notación tenemos que $[\rho_R(g)]_\beta(ij) = \delta_{(gi)j}$ donde

$$\delta_{(gi)j} = \begin{cases} 1 & \text{si } gi = j, \\ 0 & \text{en otro caso,} \end{cases}$$

es la delta de Kronecker.

Definición 1.2. Sean $\rho : G \to \mathrm{GL}(V)$ una representación con $d_\rho < \infty$ y W un subespacio de V. Diremos que W es **invariante o estable bajo** G, si para todo $g \in G$ y $w \in W$ se tiene que $\rho(g)w \in W$.

Observación. Es inmediato que los subespacios $W = V$ y $W = \{0\}$ son invariantes.

Si W es invariante bajo G, entonces para todo $g \in G$ se tiene que $\rho(g)|_W : W \to W$ es un isomorfismo (ya que es inyectivo y W tiene dimensión finita).

Definición 1.3. Si W es invariante bajo G entonces, la restricción a W, $\rho_W : G \to \mathrm{GL}(W)$ es una representación de G en W, y decimos que ρ_W es una **subrepresentación** de ρ.

Definición 1.4. Si los únicos subespacios invariantes de V son V y $\{0\}$ diremos que la representación ρ es **irreducible**. En particular toda representación de grado $d_\rho = 1$ es irreducible.

Ejemplo 1.7. Sean $\rho : S_3 \to \mathrm{GL}(3, \mathbb{R})$ la representación de permutaciones y $W = \{(x, y, z) \in \mathbb{R}^3 : x + y + z = 0\}$. Sean $\pi \in S_3$ y $w = (x, y, z) \in W$, entonces tenemos que

$$\rho(\pi)(w) = \rho(\pi)(xe_1 + ye_2 + ze_3) = xe_{\pi(1)} + ye_{\pi(2)} + ze_{\pi(3)}$$

lo que implica que la suma de las entradas de $\rho(\pi)(w)$ es igual a la suma de las entradas de w, las cuales suman cero. Por tanto, W es invariante bajo S_3. Mostraremos que la subrepresentación $\rho_W : S_3 \to \mathrm{GL}(W)$ definida por $\rho_W(\pi) = \rho(\pi)|_W$ es irreducible. Sean

$$w_1 = e_1 - e_2 = (1, -1, 0);$$

$$w_2 = e_2 - e_3 = (0, 1, -1).$$

Tome $\beta = \{w_1, w_2\}$, entonces β es base de W. Para ver esto considere lo siguiente, denote por $\langle\beta\rangle$ el subespacio generado por β. Si $u \in \langle\beta\rangle$ entonces existen a, $b \in \mathbb{R}$ tales que $u = (a, b - a, -b)$ y por tanto $u \in W$, es decir, $\langle\beta\rangle \subseteq W$. Por otro lado, si $w = (x, y, z) \in W$, entonces $z = -x - y$ y

$$(x, y, z) = (x, y, -x - y) = xw_1 + (x + y)w_2.$$

Por lo tanto $W \subseteq \langle\beta\rangle$. Así β es base de W y $dim(W) = 2$.

Supongamos que existe W' tal que $\{0\} \lneq W' \lneq W$ y que además es estable bajo ρ_W, y por tanto, invariante. Sea $w' = (x, y, z) \in W'$ distinto de cero y supongamos que $x \neq 0$ (los casos $y \neq 0$ o $z \neq 0$ son análogos). De esta forma tendríamos que $(1, \frac{y}{x}, \frac{z}{x}) \in W'$. Permutando las primeras dos entradas $(\frac{y}{x}, 1, \frac{z}{x}) \in W'$, y al restarlos obtendríamos $(1 - \frac{y}{x}, \frac{y}{x} - 1, 0) \in W'$. Por un lado si $\frac{y}{x} \neq 1$, entonces $w_1 \in W'$. Permutando la primera y tercera entrada llegaríamos a $w_2 \in W'$, una contradicción.

Por otro lado, si $\frac{y}{x} = 1$, entonces $(1, 1, -2) \in W'$. Permutando las últimas dos entradas tendríamos que $(1, -2, 1) \in W'$ y restandolos obtendríamos

$(0, 3, -3) \in W'$, es decir, $w_2 \in W'$. Permutando la primera y tercera entrada llegaríamos a que $w_1 \in W'$, otra contradicción.

Calculemos las matrices $\rho_W(12)$ y $\rho_W(132)$. Entonces tenemos que

$$\rho_W(12)w_1 = -w_1,\ \rho_W(12)w_2 = w_1 + w_2;$$

$$\rho_W(132)w_1 = -(w_1 + w_2),\ \rho_W(132)w_2 = w_1$$

de modo que al expresarlas en la base $\beta = \{w_1, w_2\}$ obtenemos

$$\rho_W(12) = \begin{pmatrix} -1 & 1 \\ 0 & 1 \end{pmatrix},\ \rho_W(132) = \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix}.$$

Definición 1.5. Sea V un espacio vectorial y $W \leq V$. Decimos que W_0 es un complemento de W si $W_0 \leq V$ y además se cumple $W \cap W_0 = \{0\}$, $V = W + W_0$. En este caso escribiremos $V = W \oplus W_0$.

El siguiente teorema nos dice por qué es importante encontrar subespacios invariantes.

Teorema 1.1. *Sea $\rho : G \to \mathrm{GL}(V)$ una representación con $d_\rho < \infty$ y $W \leq V$ invariante. Entonces existe un complemento W_0 invariante bajo G.*

Demostración. Sea $\langle \cdot, \cdot \rangle_1$ un producto interior en V. Definimos $\langle \cdot, \cdot \rangle_0$ un nuevo producto interior en V, de tal forma que para todo $u, v \in V$,

$$\langle u, v \rangle_0 := \sum_{g \in G} \langle \rho(g)u, \rho(g)v \rangle_1 .$$

Es fácil ver que $\langle \cdot, \cdot \rangle_0$ es producto interior, por ejemplo

$$\begin{aligned} \langle u + w, v \rangle_0 &= \sum_{g \in G} \langle \rho(g)(u + w), \rho(g)v \rangle_1 \\ &= \sum_{g \in G} \langle \rho(g)u + \rho(g)w, \rho(g)v \rangle_1 \\ &= \sum_{g \in G} \langle \rho(g)u, \rho(g)v \rangle_1 + \sum_{g \in G} \langle \rho(g)w, \rho(g)v \rangle_1 \\ &= \langle u, v \rangle_0 + \langle w, v \rangle_0 . \end{aligned}$$

Las otras propiedades también son fáciles de verificar.

Mostraremos que

$$W_0 = W^\perp = \{x \in V : \langle x, w \rangle_0 = 0 \text{ para todo } w \in W\}$$

es el subespacio buscado. Observe que $\langle \cdot, \cdot \rangle_0$ es invariante, es decir, para todo $g \in G$ y $u, v \in V$ se cumple

$$\begin{aligned}\langle \rho(g)u, \rho(g)v \rangle_0 &= \sum_{s \in G} \langle \rho(s)\rho(g)u, \rho(s)\rho(g)v \rangle_1 \\ &= \sum_{s \in G} \langle \rho(sg)u, \rho(sg)v \rangle_1 \\ &= \langle u, v \rangle_0 .\end{aligned}$$

Queremos mostrar que para todo $g \in G$ y $w_0 \in W_0$ se tiene que $\rho(g)w_0 \in W_0$. Es decir, queremos ver que para todo $v \in W$ se tiene que $\langle \rho(g)w_0, v \rangle_0 = 0$. Como $\rho_W(g)$ es isomorfismo, existe $z \in W$ tal que $\rho_W(g)z = \rho(g)z = v$, y por tanto,

$$\langle \rho(g)w_0, v \rangle_0 = \langle \rho(g)w_0, \rho(g)z \rangle_0 = \langle w_0, z \rangle_0 = 0.$$

De forma que W_0 es invariante bajo G y por construcción, $V = W \oplus W_0$. □

Observaciones. *1.* Sea $\rho : G \to \mathrm{GL}(V)$ una representación, y sean W y W_0 subespacios G-invariantes tales que $V = W \oplus W_0$, entonces todo $v \in V$ se puede escribir de manera única como $v = w + w_0$, con $w \in W$ y $w_0 \in W_0$. De manera que para todo $g \in G$ obtenemos

$$\rho(g)v = \rho(g)(w + w_0) = \rho_W(g)w + \rho_{W_0}(g)w_0$$

con $\rho_W(g)w \in W$ y $\rho_{W_0}(g)w_0 \in W_0$ (pues W y W_0 son G-invariantes). En este caso diremos que ρ es la suma directa de ρ_W y ρ_{W_0} y escribiremos $\rho = \rho_W \oplus \rho_{W_0}$. Si α y β son bases de W y W_0 respectivamente entonces $\gamma = \alpha \cup \beta$ es base de V y para todo $g \in G$ tenemos

$$[\rho(g)]_\gamma = \begin{pmatrix} [\rho_W(g)]_\alpha & 0 \\ 0 & [\rho_{W_0}(g)]_\beta \end{pmatrix}.$$

Este resultado lo podemos generalizar de la siguiente forma. Sean $W_1, \ldots, W_n \leq V$, decimos que V es la suma directa de los W_i denotado $V = W_1 \oplus \cdots \oplus W_n$ si $V = W_1 + \cdots + W_n$ y $W_i \cap \sum_{j \neq i} W_j = \{0\}$ para $1 \leq i \leq n$. Si V es la suma directa de los W_i y cada W_i es G-invariante entonces podemos escribir

$$\rho = \rho_{W_1} \oplus \cdots \oplus \rho_{W_n}. \tag{1.1}$$

Si β_i es base de W_i para $1 \leq i \leq n$, entonces $\beta = \cup_{i=1}^n \beta_i$ es base de V, y para todo $g \in G$ tenemos

$$[\rho(g)]_\beta = \begin{pmatrix} [\rho_{W_1}(g)]_{\beta_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & [\rho_{W_n}(g)]_{\beta_n} \end{pmatrix}. \tag{1.2}$$

2. El complemento G-invariante del teorema 1.1 no es necesariamente único. Dada $\sigma \in S_n$, definimos

$$sgn(\sigma) = \begin{cases} 1 & \text{si } \sigma \text{ es par} \\ -1 & \text{si } \sigma \text{ es impar} \end{cases}$$

donde el ser par significa que se puede escribir como un número par de transposiciones (el hecho de que sgn esté bien definido y que $sgn(\sigma\pi) = sgn(\sigma)sgn(\pi)$ para todo $\sigma, \pi \in S_n$ es un tema fundamental de teoría de grupos). Definimos $\rho : S_n \to \mathrm{GL}(\mathbb{R}^2)$ dada por

$$\rho(\sigma)(x, y) = sgn(\sigma)(x, y).$$

Entonces ρ es una representación llamada **representación alternante**. Si $W = \{(x, x) : x \in \mathbb{R}\}$ entonces W es estable bajo S_n. Tomando

$$W_1 = \{(x, -x) : x \in \mathbb{R}\},\ W_2 = \{(x, -2x) : x \in \mathbb{R}\}$$

tenemos que W_1 y W_2 son estables bajo S_n, $\mathbb{R}^2 = W \oplus W_1 = W \oplus W_2$ y sin embargo $W_1 \neq V_2$.

Veamos cómo es el subespacio W_0 que construye el teorema 1.1. Sea $\langle \cdot, \cdot \rangle_1$ el producto interior usual de vectores en $\mathbb{R}^2$. Tenemos que

$$W = \{\overline{x} = (x, x) : x \in \mathbb{R}\},$$

en consecuencia

$$W_0 = \{\overline{y} = (u, v) \in \mathbb{R}^2 : \langle \overline{x}, \overline{y} \rangle_0 = 0,\ \text{para todo } \overline{x} \in W\}.$$

Resolviendo, obtenemos

$$\begin{aligned} 0 &= \langle \overline{x}, \overline{y} \rangle_0 \\ &= \sum_{g \in S_n} \langle \rho(g)\overline{x}, \rho(g)\overline{y} \rangle_1 \\ &= \frac{n!}{2} \langle \overline{x}, \overline{y} \rangle_1 + \frac{n!}{2} \langle -\overline{x}, -\overline{y} \rangle_1 \\ &= n! \langle \overline{x}, \overline{y} \rangle_1 = n!(xu + xv). \end{aligned}$$

Por lo tanto $u = -v$ y concluimos

$$W_0 = \{(w, -w) : w \in \mathbb{R}\} = W_1.$$

3. El teorema 1.1 puede ser falso para grupos infinitos. Sean $G = (\mathbb{R}, +)$, $V = \{f : \mathbb{R} \to \mathbb{R} : f(x) = ax + b,\ a, b \in \mathbb{R}\}$ y consideremos $\rho : \mathbb{R} \to \mathrm{GL}(V)$ dado por

$$\rho(r)f(x) = f(r + x).$$

Entonces ρ es una representación de grado $d_\rho = \infty$. Veamos por ejemplo que ρ es homomorfismo, sean $r, t, x \in \mathbb{R}$ arbitrarios, entonces

$$\begin{aligned}\rho(r + t)f(x) &= f(r + t + x)\\ &= ax + at + ar + b\\ &= \rho(r)(ax + (at + b))\\ &= \rho(r)\rho(t)f(x).\end{aligned}$$

Sea

$$W = \{f \in V : f(x) = b,\ b \in \mathbb{R}\}.$$

Entonces W es un subespacio de V invariante bajo $\mathbb{R}$ (ya que $\rho(r)f = f \in W$) y

$$W_0 = \{f \in V : f(x) = ax,\ a \in \mathbb{R}\}$$

es el único subespacio de V tal que $V = W \oplus W_0$. Sin embargo, W_0 no es invariante ($\rho(r)f = f + ar \notin W_0$).

4. Sea $\beta = \{e_1, \ldots, e_n\}$ una base ortonormal de V con respecto al producto interior $\langle \cdot, \cdot \rangle_0$ que definimos en el teorema 1.1. Entonces para todo i, j y $g \in G$ se tiene

$$\langle \rho(g)e_i, \rho(g)e_j \rangle_0 = \langle e_i, e_j \rangle_0 = \delta_{ij}. \tag{1.3}$$

Así los $\rho(g)e_i$ forman una base ortonormal de V para todo $g \in G$. Recuerde que una matriz A es **unitaria** si $AA^* = Id = A^*A$ (donde $A^* = \overline{A^t}$). Considere $A = [\rho(g)]_\beta$, entonces afirmamos que A es una matriz unitaria para todo $g \in G$. Para convencernos de esto, sean

$1 \leq j \leq n$ y $\rho(g)e_j = \sum_{i=1}^{n} a_{ij}e_i$. Entonces

$$\begin{aligned} 1 &= \langle \rho(g)e_j, \rho(g)e_j \rangle_0 \\ &= \left\langle \sum_{i=1}^{n} a_{ij}e_i, \sum_{i=1}^{n} a_{ij}e_i \right\rangle_0 \\ &= \sum_{i=1}^{n} \langle a_{ij}e_i, a_{ij}e_i \rangle_0 \\ &= \sum_{i=1}^{n} a_{ij}\overline{a_{ij}} \langle e_i, e_i \rangle_0 \\ &= \sum_{i=1}^{n} \overline{a_{ij}}a_{ij} = (A^*A)_{jj}. \end{aligned}$$

La primera, tercera y quinta igualdad se dan por la ecuación (1.3). Las demás igualdades se dan por definición y por las propiedades que cumple cualquier producto interior. De la misma manera se muestra que si $i \neq j$, entonces $0 = (A^*A)_{ij}$, por lo tanto $A^*A = Id$. Ahora, A, $A^* \in \mathrm{GL}(n, \mathbb{C})$ y como $\mathrm{GL}(n, \mathbb{C})$ es grupo, tenemos que el inverso es único, así debe ser el caso que $A^* = A^{-1}$. De modo que siempre que se quiera se podrán utilizar matrices unitarias.

El siguiente teorema nos dice cómo se descompone cualquier representación de un grupo finito de grado $d_\rho < \infty$.

Teorema 1.2 (Maschke)**.** *Toda representación de un grupo finito es suma directa de representaciones irreducibles.*

Demostración. Lo haremos por inducción sobre el grado de la representación. Sea $\rho : G \to \mathrm{GL}(V)$. Si $d_\rho = 1$ entonces, ρ es irreducible, y no hay nada que probar.

Supongamos que $d_\rho = n + 1$ y que el resultado es cierto para $m \leq n$. Tenemos dos casos. El primero es que ρ sea irreducible en donde no habría nada que demostrar. El segundo es que ρ no sea irreducible, así existe un subespacio W estable bajo G y por el teorema 1.1 existe un complemento de W, digamos W_0, estable bajo G. Por tanto $\rho = \rho_W \oplus \rho_{W_0}$.

Finalmente, como las dimensiones de W y W_0 son menores ó iguales a n podemos aplicar la hipótesis de inducción y en consecuencia, tanto ρ_W como ρ_{W_0} son sumas de irreducibles lo que implica que ρ también lo es. □

Observación. El teorema 1.2 nos dice que toda representación ρ de G se puede escribir en la forma de las ecuaciones (1.1) y (1.2).

Definición 1.6. Sean $\rho_1 : G \to \mathrm{GL}(V_1)$ y $\rho_2 : G \to \mathrm{GL}(V_2)$ dos representaciones. Decimos que ρ_1 y ρ_2 son **isomorfas** si existe un isomorfismo lineal $f : V_1 \to V_2$ tal que para todo $g \in G$ se cumple $f \circ \rho_1(g) = \rho_2(g) \circ f$.

Ejemplo 1.8. Sean $\rho_1 : S_n \to \mathrm{GL}(\mathbb{R})$ la representación trivial ($\rho_1(\sigma)x = x$) y $\rho_2 : S_n \to \mathrm{GL}(\langle e_1 + \cdots + e_n \rangle)$ la representación de permutaciones ($\rho_2(\sigma)e_i = e_{\sigma(i)}$) restringida a $V = \langle e_1 + \cdots + e_n \rangle$. Es claro que $\langle e_1 + \cdots + e_n \rangle$ es ρ_2-invariante ($\rho_2(\sigma)x(e_1 + \cdots + e_n) = x(e_1 + \cdots + e_n) \in \langle e_1 + \cdots + e_n \rangle$). Sea $f : \mathbb{R} \to \langle e_1 + \cdots + e_n \rangle$ definida por

$$f(x) = x(e_1 + \cdots + e_n).$$

Entonces

$$f \circ \rho_1(\sigma)(x) = x(e_1 + \cdots + e_n) = \rho_2(\sigma)(x(e_1 + \cdots + e_n)) = \rho_2(\sigma) \circ f(x)$$

lo que implica que ρ_1 y ρ_2 son isomorfas.

Lema 1.3 (Schur)**.** *Sean $\rho_1 : G \to \mathrm{GL}(V_1)$ y $\rho_2 : G \to \mathrm{GL}(V_2)$ dos representaciones irreducibles de G. Sea $f : V_1 \to V_2$ lineal, tal que para todo $g \in G$, se cumple $f \circ \rho_1(g) = \rho_2(g) \circ f$, entonces*

(i) *Si ρ_1 y ρ_2 no son isomorfas, entonces $f = 0$.*

(ii) *Si $\rho_1 = \rho_2$ (en particular $V_1 = V_2$), entonces $f = \lambda Id$.*

Demostración. (i) Mostraremos que si $f \neq 0$ entonces f es isomorfismo y por consiguiente ρ_1 y ρ_2 son isomorfas. Afirmamos que tanto $Nuc(f)$ como $Im(f)$ son subespacios invariantes de V_1 y V_2 respectivamente. Para ver esto, sean $g \in G$, $x \in Nuc(f)$ y $f(z) \in Im(f)$ arbitrarios, entonces

$$f(\rho_1(g)(x)) = \rho_2(g)(f(x)) = \rho_2(g)(0) = 0;$$
$$\rho_2(g)(f(z)) = f(\rho_1(g)(z)),$$

lo que demuestra la afirmación. Si $f \neq 0$, entonces como ρ_1 y ρ_2 son irreducibles concluimos que $Nuc(f) = \{0\}$ e $Im(f) = V_2$ y por tanto f es isomorfismo.

(ii) Suponga que $\rho_1 = \rho_2$ y $V_1 = V_2$. Sea λ un valor propio[2] de f. Entonces $h = f - \lambda Id$ cumple

$$\begin{aligned} h \circ \rho_1(g) &= f \circ \rho_1(g) - \lambda\rho_1(g) \\ &= \rho_2(g) \circ f - \lambda\rho_2(g) \\ &= \rho_2(g) \circ (f - \lambda Id) \\ &= \rho_2(g) \circ h. \end{aligned}$$

[2]Aquí asumimos que $\rho_1 = \rho_2$ es una representación compleja, esto garantiza que exista $\lambda \in \mathbb{C}$ valor propio de f.

Por lo tanto, para todo $g \in G$, h cumple

$$h \circ \rho_1(g) = \rho_2(g) \circ h$$

y en consecuencia, $Nuc(h)$ e $Im(h)$ son invariantes. Como $Nuc(h) \neq \{0\}$ (ya que existe un vector propio asociado al valor propio λ) entonces $Nuc(h) = V_1$. Así $h = 0$, y en consecuencia $f - \lambda Id = 0$.

□

Definición 1.7. La **traza** de una matriz A (la matriz debe ser necesariamente cuadrada), se define como la suma de las entradas de su diagonal. La traza de A se denota por Tr(A).

Observaciones. Resultados elementales de álgebra lineal nos garantizan que la traza del producto de dos matrices AB es igual a la traza de la matriz BA. Además, la traza de la matriz asociada a una transformación lineal no depende de la base con la que se haya expresado.

Corolario 1.4. *Sean $\rho_1 : G \to \mathrm{GL}(V_1)$ y $\rho_2 : G \to \mathrm{GL}(V_2)$ dos representaciones irreducibles de G. Sea $h : V_1 \to V_2$ una transformación lineal arbitraria, y sea*

$$h^0 := \frac{1}{|G|} \sum_{g \in G} \rho_2(g^{-1}) h \rho_1(g)$$

entonces,

(i) *Si ρ_1 y ρ_2 no son isomorfas se tiene que $h^0 = 0$.*

(ii) *Si $\rho_1 = \rho_2$, entonces $h^0 = (\frac{Tr(h)}{d_\rho}) Id$.*

Demostración. (i) Note que para cualquier $s \in G$ se tiene

$$\begin{aligned} \rho_2(s^{-1}) h^0 \rho_1(s) &= \frac{1}{|G|} \sum_{g \in G} \rho_2(s^{-1}) \rho_2(g^{-1}) h \rho_1(g) \rho_1(s) \\ &= \frac{1}{|G|} \sum_{g \in G} \rho_2(s^{-1} g^{-1}) h \rho_1(gs) \\ &= h^0. \end{aligned}$$

Así que $f = h^0$ cumple las hipótesis del lema de Schur ($h^0 \rho_1(s) = \rho_2(s) h^0$). Aplicando el inciso (i) de dicho lema se sigue el resultado.

(ii) La parte (ii) del lema de Schur nos dice que $h^0 = cId$. Aplicando la función traza obtenemos

$$\begin{aligned} cd_\rho &= Tr(cId) \\ &= Tr(h^0) \\ &= \frac{1}{|G|} Tr(\sum_{g\in G} \rho_2(g^{-1}) h \rho_1(g)) \\ &= \frac{1}{|G|}(|G| Tr(h)) \\ &= Tr(h). \end{aligned}$$

La primera igualdad se da porque cId es una matriz de tamaño $d_\rho \times d_\rho$. Las demás igualdades son cuentas sencillas, y se deben a cómo se definió h^0.

□

Corolario 1.5. *Si G es un grupo abeliano finito, entonces todas sus representaciones irreducibles son de grado $d_\rho = 1$.*

Demostración. Sea $\rho : G \to \mathrm{GL}(V)$ irreducible, G finito y abeliano. Sea $h \in G$ arbitrario, entonces para todo $g \in G$ se tiene que

$$\rho(h)\rho(g) = \rho(hg) = \rho(gh) = \rho(g)\rho(h).$$

Por tanto, el isomorfismo $\rho(h) : V \to V$ cumple las hipótesis del lema de Schur. Aplicando el inciso (ii) del lema de Schur obtenemos $\rho(h) = \lambda_h Id$, lo que implica que todo subespacio de V es invariante. De esta forma, si $v \in V$ con $v \neq 0$ tenemos que $\langle v \rangle$ es invariante, pero como ρ es irreducible entonces los únicos subespacios invariantes de V son, V y $\{0\}$. Como $\langle v \rangle \neq \{0\}$, llegamos a $\langle v \rangle = V$. □

1.2. Caracteres de grupos finitos

Definición 1.8. Sea $\rho : G \to \mathrm{GL}(V)$ una representación con $d_\rho < \infty$. El **carácter** de ρ se define como la función $\chi_\rho : G \to \mathbb{C}$ dada por

$$\chi_\rho(g) := Tr(\rho(g)).$$

Proposición 1.6. *Sean $\rho : G \to \mathrm{GL}(V)$ una representación con $d_\rho = n$ y χ su carácter, entonces*

(i) *Si ρ es compleja, entonces para todo $g \in G$, $\rho(g)$ es diagonalizable y cada valor propio de $\rho(g)$ es raíz de la unidad.*

(ii) $\chi(1) = d_\rho = n$.

(iii) $\chi(tgt^{-1}) = \chi(g)$ *para todo* $g, t \in G$.

(iv) *Si ρ es compleja, entonces* $\chi(g^{-1}) = \overline{\chi(g)}$.

(v) *Si ρ es compleja, entonces* $|\chi(g)| \leq n$.

(vi) *Si* $\rho = \rho_1 \oplus \rho_2$, *entonces* $\chi = \chi_1 + \chi_2$.

(vii) *Si ρ_1 y ρ_2 son isomorfas, entonces sus caracteres son iguales.*

Demostración. (i) Sean $g \in G$, m el orden de g, y $H = \langle g \rangle \leq G$. Considere $\eta = \rho|_H$. Entonces, por el ejemplo 1.2 se tiene que η es una representación de H en V. Utilizando el teorema 1.2, sea $\eta_{W_1} \oplus \cdots \oplus \eta_{W_k} = \eta$ una descomposición de η en subrepresentaciones irreducibles.

Como H es abeliano, obtenemos por el corolario 1.5, que cada η_{W_i} es de grado $d_{\eta_{W_i}} = 1$, lo que implica que $k = n$. Por el ejemplo 1.3, se tiene que para cada $h \in H$, $\eta_{W_i}(h)$ es una raíz m-ésima de la unidad multiplicada por la identidad, es decir, $\eta_{W_i}(h) = (e^{\frac{2\pi i}{m}})^{\alpha_i} Id$.

Finalmente, para $1 \leq i \leq n$ sean w_i, tales que $\langle w_i \rangle = W_i$ y sea $\beta = \{w_1, \ldots, w_n\}$ entonces β es base de V y

$$\rho(g)w_i = \eta(g)w_i = \eta_{W_i}(g)w_i = (e^{\frac{2\pi i}{m}})^{\alpha_i} w_i.$$

Así, $[\rho(g)]_\beta$ es una matriz diagonal.

El hecho de que ρ sea compleja es necesario en la demostración, pues hay representaciones reales que no son diagonalizables ejemplo 1.4.

(ii) Note que $\rho(1) = Id$, pues ρ es homomorfismo. Al aplicar la función traza el resultado sigue.

(iii) Se tiene que

$$\chi(tgt^{-1}) = Tr(\rho(t)\rho(g)\rho(t^{-1})) = Tr(\rho(t)\rho(t^{-1})\rho(g)) = Tr(\rho(g)) = \chi(g).$$

(iv) Dado $g \in G$ tenemos que

$$Id = \rho(e) = \rho(gg^{-1}) = \rho(g)\rho(g^{-1}),$$

por lo tanto, los valores propios de $\rho(g^{-1})$ son los inversos de los valores propios de $\rho(g)$. Es decir, λ_i es valor propio de $\rho(g)$, si y sólo si, λ_i^{-1} es valor propio de $\rho(g^{-1})$.

Por la parte (i) de esta proposición, cada valor propio de $\rho(g)$ es una raíz de la unidad y así, $\lambda_i^{-1} = \overline{\lambda_i}$. [3] En consecuencia

$$\chi(g^{-1}) = \sum_i \overline{\lambda_i} = \overline{\sum_i \lambda_i} = \overline{\chi(g)}.$$

(v) Por (i), tenemos que $\chi(g)$ es la suma de n raíces de la unidad (los valores propios). Por lo tanto $|\chi(g)| = |\sum_i \lambda_i| \leq \sum_i |\lambda_i| = n$.

(vi) Tomemos α y β bases de V_1 y V_2, entonces para todo $g \in G$

$$[\rho(g)]_{\alpha\cup\beta} = \begin{pmatrix} [\rho_1(g)]_\alpha & 0 \\ 0 & [\rho_(g)]_\beta \end{pmatrix}.$$

El resultado se sigue de la definición de carácter.

(vii) Sean χ_1 y χ_2 los caracteres de ρ_1 y ρ_2 respectivamente. Como ρ_1 y ρ_2 son isomorfas existe $f : V_1 \to V_2$ isomorfismo tal que para todo $g \in G$, $f\rho_1(g) = \rho_2(g)f$. En consecuencia $f\rho_1(g)f^{-1} = \rho_2(g)$. Aplicando la función traza, se sigue el resultado.

□

Definición 1.9. Sea G un grupo finito y sea $\mathbb{C}^G$ el conjunto de funciones de G en el conjunto de números complejos $\mathbb{C}$. Para $\phi,\ \psi \in \mathbb{C}^G$ definimos

$$\langle \phi, \psi \rangle := \frac{1}{|G|} \sum_{g\in G} \phi(g)\overline{\psi(g)}.$$

Observación. Es fácil ver que $\mathbb{C}^G$ es un espacio vectorial (de hecho si $|G| = n$, entonces $\mathbb{C}^G \simeq \mathbb{C}^n$). También es fácil ver que $\langle \cdot, \cdot \rangle$ es un producto interior, y se le conoce como producto interior estándar en $\mathbb{C}^G$. Si $\phi,\ \psi$ son caracteres, entonces la propiedad (iv) de la proposición 1.6 nos garantiza que $\langle \phi, \psi \rangle = \langle \psi, \phi \rangle$.

Teorema 1.7. *Sea G un grupo finito, entonces*

[3] Multiplicar complejos es multiplicar sus normas y sumar sus ángulos. Si λ_i es raíz de la unidad, entonces al mutliplicarlo por $\overline{\lambda_i}$ obtenemos un número complejo de norma 1 y ángulo 0.

(i) *Si χ_1 y χ_2 son caracteres de dos representaciones irreducibles no isomorfas de G, entonces $\langle \chi_1, \chi_2 \rangle = 0$, es decir, los caracteres son ortogonales.*

(ii) *Si χ es el carácter de una representación irreducible de G, entonces $\langle \chi, \chi \rangle = 1$, es decir, χ es unitario.*

Demostración. (i) Utilizaremos la notación del corolario 1.4. Así, sean $\rho_1 : G \to \mathrm{GL}(V_1)$ y $\rho_2 : G \to \mathrm{GL}(V_2)$ dos representaciones irreducibles no isomorfas de G, $h : V_1 \to V_2$ una transformación lineal arbitraria y

$$h^0 = \frac{1}{|G|} \sum_{g \in G} \rho_2(g^{-1}) h \rho_1(g).$$

Sean $[\rho_1(g)] = (a_{ij})$, $[\rho_2(g)] = (a'_{kl})$ y $[h] = (x_{st})$ las matrices correspondientes ($[h]$ se expresa con las bases con las que se expresaron ρ_1 y ρ_2).

Con esta notación tenemos que

$$[h^0] = \frac{1}{|G|} \sum_{g \in G} [\rho_2(g^{-1})][h][\rho_1(g)].$$

Ahora, como ρ_1 y ρ_2 no son isomorfas, obtenemos, por el corolario 1.4, parte (i), que $h^0 = 0$. Al evaluar en cualquier entrada ij de $[h_0]$, llegamos a

$$\begin{aligned} 0 = [h^0]_{ij} &= \frac{1}{|G|} \sum_{g \in G} \sum_{s} a'_{is}(g^{-1}) \sum_{t} x_{st} a_{tj}(g) \\ &= \frac{1}{|G|} \sum_{g \in G} \sum_{s,t} a'_{is}(g^{-1}) x_{st} a_{tj}(g). \end{aligned} \tag{1.4}$$

Lo más importante es que la igualdad (1.4) es cierta para cualquier valor que tomen de los x_{st}. En particular, si

$$x_{st} = \begin{cases} 1 & \text{si } s = i,\ t = j, \\ 0 & \text{en otro caso,} \end{cases}$$

concluimos que para cualquier i, j,

$$0 = \frac{1}{|G|} \sum_{g \in G} a'_{ii}(g^{-1}) a_{jj}(g). \tag{1.5}$$

Ahora, si χ_1 y χ_2 son los caracteres de ρ_1 y ρ_2 respectivamente, entonces

$$\begin{aligned}\langle \chi_1, \chi_2 \rangle &= \frac{1}{|G|} \sum_{g \in G} \chi_1(g)\chi_2(g^{-1}) \\ &= \frac{1}{|G|} \sum_{g \in G} \left[\left(\sum_u a_{uu}(g) \right) \left(\sum_v a'_{vv}(g^{-1}) \right) \right] \\ &= \frac{1}{|G|} \sum_{g \in G} \left[\sum_{u,v} a_{uu}(g) a'_{vv}(g^{-1}) \right]. \end{aligned} \tag{1.6}$$

De las ecuaciones (1.5) y (1.6) se sigue que $\langle \chi_1, \chi_2 \rangle = 0$.

(ii) Supongamos que $\rho_1 = \rho_2 = \rho$ y sea $d_\rho = n$. El lado derecho de la ecuación (1.4) nos dice cómo es la entrada ij de $[h^0]$, pero por otro lado el corolario 1.4, parte (ii), nos dice que

$$h^0 = (\frac{Tr(h)}{n}) Id.$$

Por consiguiente,

$$[h^0]_{ij} = (\frac{Tr(h)}{n}) \delta_{ij}.$$

Pero a su vez,

$$Tr(h) = \sum_{s,t} x_{st} \delta_{st}.$$

De modo que,

$$\frac{1}{|G|} \sum_{g \in G} \sum_{s,t} a_{is}(g^{-1}) x_{st} a_{tj}(g) = [h^0]_{ij} = \left[\frac{Tr(h)}{n} \right] \delta_{ij} = \frac{1}{n} \sum_{s,t} x_{st} \delta_{st} \delta_{ij}. \tag{1.7}$$

La ecuación (1.7) es cierta para cualquier valor que tomen los x_{st}. De esta forma, si $x_{st} = \delta_{st}$ obtenemos

$$\frac{1}{|G|} \sum_{g \in G} a_{is}(g^{-1}) a_{tj}(g) = \begin{cases} \frac{1}{n}, \text{ si } i = j \text{ y } s = t, \\ 0, \text{ en otro caso.} \end{cases} \tag{1.8}$$

Por último,

$$\begin{aligned}\langle \chi, \chi \rangle &= \frac{1}{|G|} \sum_{g \in G} \chi(g)\chi(g^{-1}) \\ &= \frac{1}{|G|} \sum_{g \in G} \left[\sum_{u} a_{uu}(g) \sum_{v} a_{vv}(g^{-1}) \right] \\ &= \frac{1}{|G|} \sum_{g \in G} \left[\sum_{u,v} a_{vv}(g^{-1}) a_{uu}(g) \right] \\ &= \frac{1}{|G|} \sum_{g \in G} \left[\sum_{v} a_{vv}(g^{-1}) a_{vv}(g) \right] = (n)\frac{1}{n} = 1. \end{aligned} \tag{1.9}$$

El lado derecho de la ecuación (1.9) es consecuencia directa de la ecuación (1.8).

□

Corolario 1.8. *Sea G un grupo finito y $\rho : G \to \mathrm{GL}(V)$ una representación con carácter χ. Sea $\rho_{W_1} \oplus \cdots \oplus \rho_{W_n} = \rho$ una descomposición de ρ en sub-representaciones irreducibles. Si $\rho_1 : G \to \mathrm{GL}(V_1)$ es una representación irreducible de G con carácter ψ, entonces $\langle \chi, \psi \rangle$ es un número entero que cuenta las ρ_{W_i} isomorfas a ρ_1. Además, el número de ρ_{W_i} isomorfas a ρ_1 no depende de la descomposición de V.*

Demostración. Sea χ_i el carácter de ρ_{W_i}, entonces

$$\langle \chi, \psi \rangle = \langle \chi_1 + \cdots + \chi_n, \psi \rangle = \sum_{i=1}^{n} \langle \chi_i, \psi \rangle .$$

Por el teorema 1.7 tenemos que

$$\langle \chi_i, \psi \rangle = \begin{cases} 1, & \text{si } \rho_{W_i} \text{ y } \rho_1 \text{ son isomorfas,} \\ 0, & \text{si no lo son.} \end{cases}$$

Para la segunda parte del corolario recuerde el hecho de que el carácter no depende de la base. Así, si cambiamos de descomposición cambiamos de base, pero no de carácter. □

El siguiente corolario es de gran utilidad y nos ayudará a simplificar la búsqueda de representaciones lineales isomorfas.

Corolario 1.9. *Dos representaciones ρ_1 y ρ_2 son isomorfas, si y sólo si, sus caracteres son iguales.*

Demostración. La proposición 1.6 parte (vii) nos dice que si ρ_1 y ρ_2 son isomorfas, entonces sus caracteres son iguales. Por otro lado si $\chi_1 = \chi_2$ concluimos, por el corolario 1.8, que cualquier representación irreducible que ocurra en ρ_1 ocurre el mismo número de veces en ρ_2 y viceversa. □

Observaciones. *1.* El nombre de carácter de una representación se debe al corolario 1.9, pues el carácter caracteriza a la representación.

2. Sea ρ una representación con carácter χ. Sean k el número de representaciones irreducibles no isomorfas de ρ y $\rho_{W_1} \oplus \cdots \oplus \rho_{W_k}$ una descomposición de ρ como suma de irreducibles.

Sea χ_i el carácter de ρ_{W_i} y sea

$$m_i = \langle \chi, \chi_i \rangle .$$

Entonces escribiremos

$$\rho = m_1\rho_{W_1} \oplus \cdots \oplus m_k\rho_{W_k}$$

para denotar que ρ contiene a ρ_{W_i} m_i veces. Al número m_i lo llamaremos la **multiplicidad** de ρ_{W_i}.

Esta notación es muy útil y nos induce una descomposición natural del carácter en combinación lineal de caracteres irreducibles de la forma

$$\chi = m_1\chi_1 + \cdots + m_k\chi_k.$$

Sabemos que si χ es irreducible entonces necesariamente $\langle \chi, \chi \rangle = 1$. El inciso (ii) del siguiente corolario nos garantiza que es suficiente que $\langle \chi, \chi \rangle = 1$ para afirmar que χ es irreducible.

Corolario 1.10. *Sea χ el carácter de ρ expuesto como en la observación 2. de arriba, entonces*

(i) $\langle \chi, \chi \rangle = \sum_i m_i^2$.

(ii) *χ es ireducible, si y sólo si,* $\langle \chi, \chi \rangle = 1$.

Demostración. (i) Tenemos que $\chi = m_1\chi_1 + \ldots + m_k\chi_k$. Por el teorema 1.7, si $i \neq j$, entonces $\langle m_i\chi_i, m_j\chi_j \rangle = 0$. Pero si $i = j$, entonces, $\langle m_i\chi_i, m_j\chi_j \rangle = m_i^2$.

(ii) Del teorema 1.7 sabemos que si χ es irreducible entonces $\langle \chi, \chi \rangle = 1$. Si $\langle \chi, \chi \rangle = 1$ concluimos por el inciso (i), que $\chi = m_i\chi_i$ para algún i, con $m_i = 1$.

□

Observación. Considere la representación regular de G (ejemplo 1.6). Así, $|G| = n$, V es espacio vectorial de dimensión n con base $\beta = \{v_g\}_{g\in G}$ y $\rho_R(g)v_h = v_{gh}$. Note que si $g \neq 1$ entonces $\rho_R(g)v_h = v_{gh} \neq v_h$. Así, si $g \neq 1$ llegamos a que la diagonal de la matriz $[\rho_R(g)]_\beta$ consta de puros ceros. Si $g = 1$ entonces $\rho_R(g)v_h = v_{1h} = v_h$ y en consecuencia, la diagonal de la matriz $[\rho_R(1)]_\beta$ consta de puros unos. Sea χ_R el carácter de la representación regular. Por las observaciones que acabamos de mencionar concluimos

$$\chi_R(g) = \begin{cases} n, \text{ si } g = 1, \\ 0, \text{ en otro caso.} \end{cases} \tag{1.10}$$

Proposición 1.11. *Toda representación irreducible ρ_i de G está contenida en la representación regular de G con multiplicidad $m_i = d_{\rho_i}$.*

Demostración. Si $g \neq 1$, entonces por la ecuación (1.10) tenemos $\chi_R(g) = 0$. Sea χ_i el carácter de ρ_i entonces

$$\begin{aligned} m_i &= \langle \chi_R, \chi_i \rangle \\ &= \frac{1}{|G|} \sum_{g\in G} \chi_R(g)\overline{\chi_i(g)} \\ &= \frac{1}{|G|} \chi_R(1)\overline{\chi_i(1)} \\ &= \frac{1}{|G|} |G| d_{\rho_i} = d_{\rho_i}. \end{aligned}$$

La primera igualdad es por definición de m_i, la segunda, por cómo definimos el producto interior. La tercera, porque $\chi_R(g) = 0$ para $g \neq 1$. La cuarta se debe a la ecuación (1.10) ($|G| = n = \chi_R(1)$). □

Corolario 1.12. (i) *Si G es un grupo finito, entonces G tiene sólo un número finito de representaciones irreducibles no isomorfas.*

(ii) *Si $\rho_1, \ldots, \rho_k$ son todas las representaciones irreducibles de G y si d_i es el grado de ρ_i, entonces $\sum_i d_i^2 = |G|$.*

(iii) *Sea $g \in G$, y $g \neq 1$. Entonces $\sum_i d_i \chi_i(g) = 0$.*

Demostración. (i) Cualquier representación irreducible de G está contenida en la representación regular y el espacio de la representación regular tiene dimensión $|G| = n < \infty$. Supongamos que existen $\rho_1, \rho_2, \ldots$ una infinidad de representaciones irreducibles no isomorfas de G, cada una de grado $d_i \geq 1$. Se sigue que

$$\infty = \sum_{i=1}^{\infty} d_i \leq n,$$

lo cual es imposible.

(ii) Para $1 \leq i \leq k$, sea χ_i el carácter de ρ_i. La proposición 1.11 nos dice que

$$\chi_R = d_i\chi_i + \cdots + d_k\chi_k. \tag{1.11}$$

En particular, al evaluar la ecuación (1.11) en 1, obtenemos

$$|G| = \chi_R(1) = d_i\chi_i(1) + \cdots + d_k\chi_k(1) = \sum_i d_i^2.$$

(iii) Evaluando la ecuación (1.11) en $g \neq 1$, tenemos

$$0 = \chi_R(g) = \sum_i d_i\chi_i(g).$$

□

Nuestro siguiente objetivo es obtener el número exacto de representaciones irreducibles no isomorfas de un grupo finito G. Para ver esto recordemos el concepto de **clase de conjugación**.

Definición 1.10. Dado un grupo G y a, $b \in G$, decimos que a y b son **conjugados** si existe $g \in G$ tal que $a = gbg^{-1}$. Esta es una relación de equivalencia y llamamos a sus clases, clases de conjugación. Una función $f : G \to \mathbb{C}$ es una **función de clases** si es constante en cada clase de conjugación, lo que equivale a decir que $f(b) = f(gbg^{-1})$ para todo b, $g \in G$.

Observación. Sea W el conjunto de todas las funciones de clases de G, es fácil ver que W es subespacio de $\mathbb{C}^G$. Podemos heredarle a W el producto interior de funciones que tenemos en $\mathbb{C}^G$ simplemente restringiéndolo a W. Si χ es un carácter, entonces $\chi(t) = \chi(gtg^{-1})$ para todo g, $t \in G$. Así, los caracteres son funciones de clases.

Proposición 1.13. *Sean, $f : G \to \mathbb{C}$ una función de clases y $\rho : G \to \mathrm{GL}(V)$ una representación irreducible de grado n y carácter χ. Sea*

$$\widehat{f}(\rho) := \sum_{g\in G} f(g)\rho(g).$$

Entonces $\widehat{f}(\rho) = \lambda Id$, con $\lambda = \frac{1}{n}\sum_{g\in G} f(g)\chi(g) = \frac{|G|}{n}\langle f, \overline{\chi}\rangle$.

Demostración. Veamos que $\widehat{f}(\rho)$ satisface las hipótesis del lema de Schur. Sea $h \in G$ arbitrario, entonces

$$\rho(h)\widehat{f}(\rho)\rho(h^{-1}) = \sum_{g\in G} \rho(h)f(g)\rho(g)\rho(h^{-1}) = \sum_{g\in G} f(g)\rho(hgh^{-1}).$$

Ahora, si g varía en G, entonces también $v = hgh^{-1}$ varía en G. Por lo tanto

$$\sum_{g\in G} f(g)\rho(hgh^{-1}) = \sum_{v\in G} f(h^{-1}vh)\rho(v) = \sum_{v\in G} f(v)\rho(v) = \widehat{f}(\rho).$$

De esta forma demostramos que $\rho(h)\widehat{f}(\rho)\rho(h^{-1}) = \widehat{f}(\rho)$, y por el inciso (ii) del lema de Schur, tenemos $\widehat{f}(\rho) = \lambda Id$. Aplicando la función traza obtenemos

$$\lambda n = Tr(\widehat{f}(\rho)) = Tr\left(\sum_{g\in G} f(g)\rho(g)\right) = \sum_{g\in G} f(g)\chi(g) = |G|\left\langle f, \overline{\chi}\right\rangle. \qquad \square$$

Teorema 1.14. *Sea G un grupo finito, entonces*

(i) *Los caracteres irreducibles $\chi_1, \ldots, \chi_k$ de G forman una base del espacio de funciones de clases W.*

(ii) *El número de representaciones irreducibles no isomorfas de G es igual al número de clases de conjugación de G.*

Demostración. (i) Por el teorema 1.7, los caracteres irreducibles $\chi_1, \ldots, \chi_k$ de G forman un sistema ortonormal, lo que implica que son linealmente independientes. De esta forma, basta probar que generan a todo el espacio. Para demostrar esto, basta mostrar que $\{\overline{\chi_1}, \ldots, \overline{\chi_k}\}^{\perp} = \{0\}$ (aquí notemos que $f \in \{\overline{\chi_1}, \ldots, \overline{\chi_k}\}^{\perp}$, si y sólo si, $\overline{f} \in \{\chi_1, \ldots, \chi_k\}^{\perp}$).

Sea $f \in \{\overline{\chi_1}, \ldots, \overline{\chi_k}\}^{\perp}$ y sea ρ una representación con carácter χ. Considere $\widehat{f}(\rho) = \sum_{g\in G} f(g)\rho(g)$ como en la proposición 1.13. Si ρ es irreducible, entonces $\chi = \chi_i$ para algún i, y por la proposición 1.13 tenemos

$$\widehat{f}(\rho) = \frac{|G|}{n}\left\langle f, \overline{\chi}\right\rangle Id = \frac{|G|}{n}\left\langle f, \overline{\chi}_i\right\rangle Id = 0.$$

Si ρ no fuera irreducible, entonces al descomponerla en suma directa de subrepresentaciones irreducibles obtendríamos el mismo resultado, es decir, $\widehat{f}(\rho) = 0$.

Sea $\rho_R : G \to \mathrm{GL}(V)$ la representación regular de G y sea $\{v_g\}_{g\in G}$ una base de V. De esta forma, tenemos que $\widehat{f}(\rho_R) = 0$. Evaluando $\widehat{f}(\rho_R)$ en v_1 (donde 1 es el elemento neutro de G y $v_1 \in \{v_g\}_{g\in G}$), obtenemos

$$\widehat{f}(\rho_R)(v_1) = \sum_{g\in G} f(g)\rho_R(g)(v_1) = \sum_{g\in G} f(g)v_g = 0.$$

De la última igualdad y del hecho de que $\{v_g\}_{g\in G}$ es base de V, concluimos que $f(g) = 0$ para todo $g \in G$. Por lo tanto, $\{\overline{\chi_1}, \ldots, \overline{\chi_k}\}^{\perp} = \{0\}$.

(ii) Sean $C_1, \ldots, C_m$ las clases de conjugación de G. Se demostrará que la dimensión del espacio de funciones de clases de G es m. La afirmación es inmediata, para $1 \leq i \leq m$, basta definir $f_i : G \to \mathbb{C}$ dadas por

$$f_i(C_j) = \delta_{ij}.$$

Es evidente que las f_i forman una base del espacio de funciones de clases de G. Por lo tanto tienen la misma cardinalidad que el conjunto de los caracteres irreducibles de G que también es base, además de ser ortonormal.

□

Teorema 1.15. *Un grupo finito G es abeliano, si y sólo si, todas las representaciones irreducibles de G son de grado $d_i = 1$.*

Demostración. Si G es abeliano, entonces, por el corolario 1.5, todas las representaciones irreducibles de G son de grado $d_i = 1$. Recíprocamente, si todas las representaciones irreducibles de G son de grado $d_i = 1$, entonces, por el corolario 1.12 tenemos que $\sum_i d_i = \sum_i d_i^2 = |G|$. Por lo tanto, hay $|G|$ representaciones irreducibles no isomorfas. Por el teorema 1.14 este número es igual al número de clases de conjugación. De esta forma existen $|G|$ clases de conjugación.

Si existen $|G|$ clases de conjugación, entonces cada clase consta únicamente de un elemento, esto debido a que las clases de conjugación forman una partición de G. Como g siempre pertenece a su clase de conjugación concluimos que la clase de conjugación de g es $\{g\}$, lo que quiere decir que $tgt^{-1} = g$ para todo t, $g \in G$, y en consecuencia $tg = gt$ para todo t, $g \in G$. □

1.3. Funciones asociadas a representaciones

Definición 1.11. Sean G un grupo finito y P y Q probabilidades en G. Definimos la **convolución** entre P y Q como la probabilidad en G dada por

$$P * Q(s) := \sum_{t\in G} P(st^{-1})Q(t).$$

Observaciones. *1.* $P*Q$ puede ser interpretada como "primero seleccione t usando Q, luego independientemente seleccione u usando P y forme el producto ut".

2. Veamos que en efecto $P * Q$ es una probabilidad en G. Como P y Q son probabilidades es claro que $P * Q \geq 0$. Note que,

$$\begin{aligned}\sum_{s\in G}(P*Q)(s) &= \sum_{s\in G}\sum_{t\in G}P(st^{-1})Q(t)\\ &= \left(\sum_{t\in G}P(t)\right)\left(\sum_{t\in G}Q(t)\right) = (1)(1) = 1.\end{aligned}$$

Esto demuestra que $P * Q$ es una probabilidad en G.

3. Si G es abeliano, es fácil ver que $P * Q = Q * P$ (todo sumando de $(P*Q)(s)$ es sumando de $(Q*P)(s)$ y viceversa). Sin embargo, si G no es abeliano puede suceder que $P * Q \neq Q * P$. Por ejemplo si $G = S_3$, $P(123) \neq P(132)$ y $Q(23) = 1$ entonces

$$(P * Q)(12) = P(123) \neq P(132) = (Q * P)(12).$$

Definición 1.12. Sea Q una probabilidad en G. Para $n = 1$, definimos

$$Q^{*1} = Q.$$

Para $n \geq 2$, definimos

$$Q^{*n} = Q * Q^{*n-1}.$$

Decimos que Q^{*n} es la **n-ésima convolución de** Q.

Definición 1.13. Si G es un grupo de orden n la **distribución uniforme** en G es la función U que a cada elemento en G le asocia el valor $\frac{1}{n}$.

Observación. Es claro que U es una probabilidad en G y además cumple $U * P = U$ para toda probabilidad P pues

$$U * P(s) = \sum_{t\in G}\frac{1}{n}P(t) = \frac{1}{n}(1) = \frac{1}{n}.$$

Definición 1.14. Sea Q una probabilidad en G y ρ una representación con dominio G. La **transformación de Fourier** de Q en la representación ρ es la matriz

$$\widehat{Q}(\rho) := \sum_{s\in G}Q(s)\rho(s).$$

La misma definición sirve para cualquier función f.

Proposición 1.16. *Sea ρ cualquier representación con dominio G y P, Q probabilidades en G. Entonces $\widehat{P * Q}(\rho) = \widehat{P}(\rho)\widehat{Q}(\rho)$.*

Demostración. Note que,

$$\widehat{P * Q}(\rho) = \sum_{v \in G} (P * Q)(v)\rho(v) = \sum_{v \in G} \sum_{t \in G} P(vt^{-1})Q(t)\rho(v). \tag{1.12}$$

Adicionalmente,

$$\widehat{P}(\rho)\widehat{Q}(\rho) = \left(\sum_{s \in G} P(s)\rho(s) \right) \left(\sum_{t \in G} Q(t)\rho(t) \right) = \sum_{s \in G} \sum_{t \in G} P(s)Q(t)\rho(st). \tag{1.13}$$

Basta probar que el lado derecho de la ecuación (1.12) es igual al lado derecho de la ecuación (1.13).

Sea $v = st$, entonces $vt^{-1} = s$. Sustituyendo el valor vt^{-1} y sumando sobre s en la ecuación (1.12) obtenemos

$$\widehat{P * Q}(\rho) = \sum_{s \in G} \sum_{t \in G} P(s)Q(t)\rho(st),$$

y el resultado sigue. □

Corolario 1.17. *La transformación de Fourier de la $n-$ésima convolución de Q en la representación ρ es igual a la transformación de Fourier de Q en la representación ρ a la n, es decir,*

$$\widehat{Q^{*n}}(\rho) = [\widehat{Q}(\rho)]^n.$$

Demostración. Lo haremos por inducción sobre n. Para $n = 1$ la prueba es trivial. Supongamos que vale para $k = n$, entonces

$$\widehat{Q^{*n+1}}(\rho) = \widehat{Q * Q^{*n}}(\rho) = \widehat{Q}(\rho)\widehat{Q^{*n}}(\rho) = \widehat{Q}(\rho)[\widehat{Q}(\rho)]^n = [\widehat{Q}(\rho)]^{n+1}.$$

En donde en la segunda igualdad utilizamos la proposición 1.16 y en la penúltima la hipótesis de inducción. □

El corolario 1.17 será de gran utilidad en la sección 4.3 de la tesis.

Proposición 1.18. *Si ρ es la representación trivial, entonces $\widehat{U}(\rho) = Id$. Si ρ es cualquier representación irreducible no trivial, entonces $\widehat{U}(\rho) = 0$.*

Demostración. Si ρ es la representación trivial, entonces

$$\widehat{U}(\rho) = \sum_{g \in G} U(g)\rho(g) = \sum_{g \in G} \frac{1}{|G|} Id = Id.$$

Sea $\rho : G \to \mathrm{GL}(V)$ una representación irreducible no trivial de G y sea $\{v_1, \ldots, v_n\}$ base de V. Considere

$$a_i = \widehat{U}(\rho)(v_i) = \frac{1}{|G|} \sum_{g \in G} \rho(g) v_i.$$

Entonces para todo h en G tenemos que $\rho(h)a_i = a_i$. Por lo tanto, $\langle a_i \rangle$ es un subespacio G-invariante. Así $\langle a_i \rangle = V$ o $\langle a_i \rangle = \{0\}$. Si $\langle a_i \rangle = V$, entonces ρ sería la representación trivial, en contradicción con nuestra hipótesis. □

Teorema 1.19. *Sea G un grupo finito y f, $h : G \to \mathbb{C}$ funciones arbitrarias. Entonces valen*

(i) *Teorema de la Inversión de Fourier*

$$f(s) = \frac{1}{|G|} \sum_i d_i Tr(\rho_i(s^{-1}) \widehat{f}(\rho_i)).$$

(ii) *Fórmula de Plancherel*

$$\sum_{s \in G} f(s^{-1}) h(s) = \frac{1}{|G|} \sum_i d_i Tr(\widehat{f}(\rho_i) \widehat{h}(\rho_i)).$$

Donde la suma en 'i' es sobre todas la representaciones irreducibles de G y d_i es el grado de ρ_i.

Demostración. (i) Tenemos

$$\begin{aligned}
\frac{1}{|G|} \sum_i d_i Tr(\rho_i(s^{-1}) \widehat{f}(\rho_i)) &= \frac{1}{|G|} \sum_i d_i Tr\left(\rho_i(s^{-1}) \sum_{g \in G} f(g) \rho_i(g) \right) \\
&= \frac{f(s)}{|G|} \sum_i d_i Tr(\rho_i(1)) \\
&+ \sum_{g \neq s} \frac{f(g)}{|G|} \sum_i d_i Tr(\rho_i(s^{-1}g)) \\
&= \frac{f(s)}{|G|} \sum_i d_i^2 + \sum_{g \neq s} \frac{f(g)}{|G|} \sum_i d_i Tr(\rho_i(s^{-1}g)) \\
&= \frac{f(s)}{|G|} |G| + 0 \\
&= f(s).
\end{aligned}$$

Donde en la cuarta igualdad hemos utilizado el corolario 1.12 que nos dice

$$\sum_i d_i Tr(\rho_i(1)) = \sum_i d_i^2 = |G|;$$

$$\sum_i d_i Tr(\rho_i(s^{-1}g)) = \sum_i d_i \chi_{\rho_i}(s^{-1}g)) = 0 \text{ para } g \neq s.$$

(ii) Note que

$$\begin{aligned}
\frac{1}{|G|}\sum_i d_i Tr(\widehat{f}(\rho_i)\widehat{h}(\rho_i)) &= \frac{1}{|G|}\sum_i d_i Tr\left[\sum_{s\in G} f(s)\rho_i(s)\widehat{h}(\rho_i)\right] \\
&= \sum_{s\in G} f(s)\frac{1}{|G|}\sum_i d_i Tr\left[\rho_i(s)\widehat{h}(\rho_i)\right] \\
&= \sum_{s\in G} f(s)h(s^{-1}) \\
&= \sum_{s\in G} f(s^{-1})h(s). \qquad (1.14)
\end{aligned}$$

En la tercera igualdad utilizamos $|G|$ veces el inciso anterior.

□

La fórmula de Plancherel será de vital importancia en la sección 4.2 de la tesis.

Capítulo 2

Cadenas de Markov

En este capítulo introducimos los conceptos, definiciones y teoremas elementales de cadenas de Markov, todo a manera de recordatorio y en forma compacta. Lo que se presenta aquí será usado para la parte de aplicaciones de teoría de representaciones a la estimación de tasa de convergencia al estado estacionario para cadenas de Markov. Este es un capítulo de referencia. La bibliografía correspondiente puede consultarse en [3], [4], [5] y [6].

2.1. Nociones elementales

Definición 2.1. Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad. Un **proceso estocástico** es una familia de variables aletorias

$$X = \{X_i : \Omega \to S, i \in I\}$$

indexadas en un conjunto I y que toman valores en un conjunto S.

Observación. Los procesos estocásticos se clasifican dependiendo de cómo sean I y S. Para nuestro estudio, bastará tomar $I = \mathbb{N}$ y S un conjunto a lo más, numerable.

Definición 2.2. Sean $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad y

$$X = \{X_n : \Omega \to S, n = 0, 1, 2, \ldots\}$$

una sucesión de variables aleatorias asumiendo valores en un conjunto S, donde S es finito o numerable. Decimos que dicha sucesión forma una **cadena de Markov** si para todo $n \in \mathbb{N}$ y para todo $x_0, \ldots, x_{n-1}, i, j \in S$ se cumple

$$P(X_{n+1} = j | X_n = i, \ldots, X_0 = x_0) = P(X_{n+1} = j | X_n = i) \qquad (2.1)$$

siempre y cuando los eventos con los que se está condicionando sean de probabilidad positiva. S es llamado el **espacio de estados** de la cadena X.

Observación. Si $P(X_{n+1} = j|X_n = i)$ no depende de n, diremos que la cadena es homogénea en el tiempo. Siempre consideraremos cadenas homogéneas.

Definición 2.3. La **probabilidad de transición** de la cadena se define para $i, j \in S$ como

$$p(i,j) := P(X_{n+1} = j|X_n = i),$$

la cual denotaremos por p_{ij}, mientras que la **matriz de transición** P de la cadena se define como

$$\mathrm{P}_{ij} = p_{ij}.$$

Definición 2.4. Definimos a la probabilidad π_0 como

$$\pi_0(i) = P(X_0 = i),\ i \in S.$$

Observaciones. *1.* Con la función de transición y la distribución inicial podemos encontrar la distribución de X_n para cualquier n. Esto lo hacemos notando que

$$P(X_n = j) = P(X_n = j, X_{n-1} \in S) = \sum_{i \in S} P(X_{n-1} = i)p_{ij}. \tag{2.2}$$

Para $n = 1$, tenemos que

$$P(X_1 = i) = \sum_{i \in S} \pi_0(i)p_{ij}.$$

Repitiendo este proceso $(n-1)$ veces más, obtenemos la distribución de X_n.

2. La distribución conjunta de $X_0, \ldots, X_n$ también se puede obtener en términos de la distribución inicial y las probabilidades de transición. Por inducción sobre n se verifica fácilmente que

$$P(X_0 = x_0, \ldots, X_n = x_n) = \pi_0(x_0)p_{x_0x_1}p_{x_1x_2}\cdots p_{x_{n-1}x_n}. \tag{2.3}$$

La siguiente propiedad es muy útil a la hora de calcular probabilidades conjuntas.

Proposición 2.1. *Para $X = \{X_n : n \geq 0\}$ una cadena de Markov con espacio de estados S y probabilidad de transición p_{ij}, $i, j \in S$, sean $A_0, \ldots, A_{n-1}$, $B_1, \ldots, B_m$ subconjuntos de S y $x \in S$ entonces*

$$P(X_{n+m} \in B_m, \ldots, X_{n+1} \in B_1|X_n = x, X_{n-1} \in A_{n-1}, \ldots, X_0 \in A_0) = \sum_{y_1 \in B_1} \cdots \sum_{y_m \in B_m} p_{xy_1}p_{y_1y_2}\cdots p_{y_{m-1}y_m}. \tag{2.4}$$

Demostración. Ver [5], p.12. □

Definición 2.5. La **probabilidad de transición en m-pasos** ($m > 1$) de la cadena X se define como

$$p_{ij}(m) := P(X_{n+m} = j | X_n = i) = \sum_{y_1 \in S} \cdots \sum_{y_{m-1} \in S} p_{iy_1} p_{y_1 y_2} \cdots p_{y_{m-1} j}.$$

La última igualdad se da en virtud de la proposición 2.1. Para $m = 0$ la función de transición de la cadena X se define como la delta de Kronecker, es decir, $p_{ii}(0) = 1$ y $p_{ij}(0) = 0$ para $j \neq i$.

Definición 2.6. Sea S el conjunto de estados de una cadena de Markov. Si S es finito, entonces para $m \geq 0$ se define la **matriz de transición en m-pasos** $\mathrm{P}^{(m)}$, cuya entrada ij es

$$\mathrm{P}_{ij}^{(m)} = p_{ij}(m).$$

Proposición 2.2 (Ecuación Chapman-Kolmogorov)**.** *Sea $X = \{X_n : n \geq 0\}$ una cadena de Markov con espacio de estados S y probabilidades de transición en m-pasos $p_{ij}(m)$, i, $j \in S$, $m \in \mathbb{N}$. Entonces, para todo i, $j \in S$ y para todo n, $m \in \mathbb{N}$ se cumple*

$$p_{ij}(n + m) = \sum_{z \in S} p_{iz}(n) p_{zj}(m). \tag{2.5}$$

Demostración. Observe que

$$\begin{aligned} p_{ij}(n+m) &= P(X_{n+m} = j | X_0 = i) \\ &= \sum_{z \in S} P(X_n = z | X_0 = i) P(X_{n+m} = j | X_n = z, X_0 = i) \\ &= \sum_{z \in S} P(X_n = z | X_0 = i) P(X_{n+m} = j | X_n = z) \\ &= \sum_{z \in S} p_{iz}(n) p_{zj}(m). \end{aligned}$$

□

Observación. De la ecuación Chapman-Kolmogorov, con $n = m = 1$ se deduce que, para S finito, $\mathrm{P}^{(2)} = \mathrm{PP}$, es decir, la matriz de transición en dos pasos es el producto de la matriz de transición en un paso consigo misma. Suponiendo el resultado válido para n y haciendo $m = 1$, la ecuación Chapman-Kolmogorov nos garantiza que para S finito,

$$\mathrm{P}^{(n+1)} = \overbrace{\mathrm{P} \cdots \mathrm{P}}^{n+1},$$

lo que equivale a decir que para cualquier m, $\mathrm{P}^{(m)}$ es el producto de P consigo misma m-veces.

Definición 2.7. Sea $S = \{a_0, \ldots, a_d\}$, entonces para $n \in \mathbb{N}$ definimos

$$\pi_n(a_i) = P(X_n = a_i),\ a_i \in S;$$

e indicamos

$$\pi_n = (P(X_n = a_0), \ldots, P(X_n = a_d)).$$

Antes de ver cómo se relacionan las matrices de transición $\mathrm{P}^{(m)}$ y los vectores π_n, veamos una proposición.

Proposición 2.3. *Las siguientes igualdades son válidas*

(i)
$$P(X_n = j) = \sum_{i \in S} \pi_0(i) p_{ij}(n). \tag{2.6}$$

(ii)
$$P(X_{n+1} = j) = \sum_{i \in S} P(X_n = i) p_{ij}. \tag{2.7}$$

Demostración. (i) Note que

$$\begin{aligned} P(X_n = j) &= P(X_0 \in S, X_n = j) \\ &= \sum_{i \in S} P(X_0 = i, X_n = j) = \sum_{i \in S} \pi_0(i) p_{ij}(n). \end{aligned}$$

(ii) La demostración sale directamente de la ecuación (2.2).

□

Observación. En términos matriciales obtenemos las siguientes relaciones

$$\pi_n = \pi_0 \mathrm{P}^{(n)}; \tag{2.8}$$

$$\pi_{n+1} = \pi_n \mathrm{P}. \tag{2.9}$$

Estas ecuaciones son consecuencia directa de (2.6) y (2.7), respectivamente.

2.2. Clasificación de estados y teoremas límite

Utilizaremos las notaciones $P_i(\cdot)$ y $E_i[\cdot]$ para indicar probabilidades y esperanzas, respectivamente, de variables aleatorias definidas en términos de una cadena de Markov que haya empezado en el estado i (es decir, $\pi_0(i) = P(X_0 = i) = 1$).

Definición 2.8. Sea $A \subseteq S$, el **tiempo de arribo a** $\boldsymbol{A}$ lo definimos como

$$T_A := \text{mín}\{n > 0 : X_n \in A\}$$

si $X_n \in A$ para algún $n > 0$, y $T_A = \infty$ si $X_n \notin A$ para todo $n > 0$. En caso de que $A = \{i\}$, escribiremos T_i en lugar de $T_{\{i\}}$.

Definición 2.9. Un estado i es llamado **recurrente** si

$$P_i(T_i < \infty) = 1.$$

En otro caso diremos que i es **transitorio**.

Observaciones. *1.* Utilizaremos las notaciones

$$f_{ij}(n) = P_i(T_j = n), \text{ y } f_{ij} = \sum_{n=1}^{\infty} f_{ij}(n).$$

Con esta notación tenemos que i es recurrente si $f_{ii} = 1$ y transitorio si $f_{ii} < 1$.

2. Sea $N(j)$ la variable aleatoria que indica el número de visitas (sin contar el tiempo $n = 0$) al estado j. Si I_j es la función indicadora del conjunto $\{j\}$, entonces

$$N(j) = \sum_{n=1}^{\infty} I_j(X_n).$$

Observe que

$$E_i[N(j)] = E_i\left[\sum_{n=1}^{\infty} I_j(X_n)\right] = \sum_{n=1}^{\infty} E_i[I_j(X_n)] = \sum_{n=1}^{\infty} p_{ij}(n).$$

Teorema 2.4. (i) *Sea j un estado transitorio, entonces*

$$P_i(N(j) < \infty) = 1.$$

Además,

$$\sum_{n=1}^{\infty} p_{ij}(n) = E_i[N(j)] = \frac{f_{ij}}{1 - f_{jj}}$$

es finito para todo $i \in S$.

(ii) *Sea j un estado recurrente, entonces*

$$P_j(N(j) = \infty) = 1, \text{ y } \sum_{n=1}^{\infty} p_{jj}(n) = E_j[N(j)] = \infty.$$

Además, para todo $i \in S$ se tiene

$$P_i(N(j) = \infty) = P_i(T_j < \infty) = f_{ij}.$$

Si $f_{ij} = 0$, entonces $E_i[N(j)] = 0$. Si $f_{ij} > 0$, entonces $E_i[N(j)] = \infty$.

Demostración. Ver [5], p.19-20. □

Definición 2.10. El **tiempo medio de recurrencia** μ_i del estado i se define por

$$\mu_i := E_i(T_i) = \sum_{n=1}^{\infty} n f_{ii}(n),$$

la última igualdad siendo cierta sólo cuando $P_i(T_i < \infty) = 1$.

Observación. Note que $P_i(T_i = \infty) > 0$, si y sólo si, el estado i es transitorio. En este caso, $\mu_i = E_i(T_i) = \infty$.

Definición 2.11. A los estados recurrentes para los cuales $\mu_i = \infty$, ($\mu_i < \infty$) los llamaremos **nulos recurrentes**, (**recurrentes positivos**).

Teorema 2.5. *Un estado recurrente i es nulo, si y sólo si, $p_{ii}(n) \to 0$ cuando $n \to \infty$. Si esto sucede, entonces $p_{ji}(n) \to 0$ para todo $j \in S$.*

Demostración. Ver [4], p.222. □

Definición 2.12. El **periodo** $d(i)$ del estado i de la cadena X se define por

$$d(i) := \text{mcd}\{n : p_{ii}(n) > 0\}.$$

Decimos que i es **periódico** si $d(i) > 1$ y **aperiódico** si $d(i) = 1$.

Definición 2.13. Un estado de la cadena X es llamado **ergódico** si es recurrente positivo y aperiódico.

Definición 2.14. Decimos que j **es accesible a partir de** i, y escribimos $i \to j$, si $p_{ij}(n) > 0$ para algún $n \geq 0$. Decimos que i y j se **comunican**, si $i \to j$ y $j \to i$, y escribimos $i \leftrightarrow j$.

Observación. Es inmediato que la relación $\leftrightarrow$ es de equivalencia.

El siguiente teorema habla de las relaciones que guardan los elementos de las diversas clases de equivalencia.

Teorema 2.6. *Si $i \leftrightarrow j$, entonces,*

(i) *i y j tienen el mismo periodo,*

(ii) *i es transitorio, si y sólo si, j lo es,*

(iii) *i es nulo recurrente, si y sólo si, j lo es,*

(iv) *i es recurrente positivo, si y sólo si, j lo es.*

Demostración. Ver [4], p.224. □

Definición 2.15. Sea $A \subseteq S$, entonces A es llamado

(i) **cerrado**, si $p_{ij} = 0$ para todo $i \in A$, $j \notin A$,

(ii) **irreducible**, si $i \leftrightarrow j$ para todo i, $j \in A$.

Observación. Una cadena de Markov es llamada irreducible si todos sus estados se comunican. De igual manera, si todos sus estados son recurrentes nulos, recurrentes positivos o ergódicos entonces ésta se llama cadena recurrente nula, recurrente positiva o ergódica, respectivamente.

En virtud del teorema 2.6, si una cadena es irreducible, entonces basta saber si un estado es recurrente o ergódico para saber si los demás lo son.

Teorema 2.7 (Teorema de descomposición)**.** *El espacio de estados S de la cadena X se puede particionar de manera única como*

$$S = T \cup C_1 \cup C_2 \cup \cdots \tag{2.10}$$

donde T es el conjunto de estados transitorios y los C_i son conjuntos cerrados e irreducibles de estados recurrentes.

Demostración. Ver [4], p.224. □

Lema 2.8. *Si el espacio de estados S es finito, entonces por lo menos uno de los estados es recurrente. Todo estado recurrente es positivo recurrente.*

Demostración. Ver [4], p.225. □

Definición 2.16. Sea p_{ij} la probabilidad de transición de la cadena de Markov X. Si $\pi(i)$, $i \in S$ son números no negativos que suman 1, y si además, para todo $j \in S$ se cumple

$$\sum_{i \in S} \pi(i) p_{ij} = \pi(j),$$

entonces diremos que π es una **distribución estacionaria** para la cadena X.

Observación. En términos matriciales, tenemos que $\pi \mathrm{P} = \pi$. Así,

$$\pi \mathrm{P}^{(2)} = (\pi \mathrm{P})\mathrm{P} = \pi \mathrm{P} = \pi. \tag{2.11}$$

En general, tenemos que para todo $n \geq 0$, $\pi \mathrm{P}^{(n)} = \pi$.

La ecuación (2.11) y la ecuación (2.8) nos dicen que si la distribución inicial π es estacionaria, entonces la distribución de X_n no depende de n.

Análogamente, si la distribución de X_n no depende de n, entonces tomando $n = 1$ llegamos a que para todo $j \in S$ se cumple

$$\sum_{i \in S} \pi_0(i) p_{ij} = P(X_1 = j) = \pi_0(j).$$

En consecuencia π_0 es estacionaria. Concluimos entonces que la distribución de X_n no depende de n, si y sólo si, la distribución inicial es estacionaria.

Teorema 2.9. *Una cadena irreducible tiene distribución estacionaria π, si y sólo si, todos los estados son recurrentes positivos. En este caso, la distribución es única y está dada para $i \in S$ por $\pi(i) = \frac{1}{\mu_i}$, donde μ_i es el tiempo medio de recurrencia.*

Demostración. Ver [4], p.227-230. □

Corolario 2.10. *Cualquier cadena irreducible que toma valores en un espacio de estados finito tiene una única distribución estacionaria dada por $\pi(i) = \frac{1}{\mu_i}$, donde μ_i es el tiempo medio de recurrencia.*

Demostración. Por el lema 2.8, existe un estado recurrente positivo. Como la cadena es irreducible, entonces todos los estados son recurrentes positivos y podemos aplicar el teorema 2.9. □

Teorema 2.11. *Para una cadena irreducible y aperiódica se tiene que*

$$\lim_{n \to \infty} p_{ij}(n) = \frac{1}{\mu_j}, \text{ para todo } i, j.$$

En particular, este resultado vale para cualquier cadena ergódica.

Demostración. Ver [4], p.232-235. □

Teorema 2.12. *Una cadena ergódica que toma valores en un espacio de estados finito converge a la distribución estacionaria sin importar cuál sea la distribución inicial X_0.*

Demostración.

$$\begin{aligned}
\lim_{n\to\infty} P(X_n = j) &= \lim_{n\to\infty} P(X_0 \in S, X_n = j) \\
&= \lim_{n\to\infty} \sum_{i\in S} P(X_0 = i, X_n = j) \\
&= \lim_{n\to\infty} \sum_{i\in S} P(X_0 = i) p_{ij}(n) \\
&= \sum_{i\in S} \Big[P(X_0 = i) \lim_{n\to\infty} p_{ij}(n) \Big] \\
&= \sum_{i\in S} P(X_0 = i) \frac{1}{\mu_j} = \frac{1}{\mu_j}.
\end{aligned}$$

En la cuarta igualdad utilizamos el hecho de que S es finito. □

Capítulo 3

Caminatas aleatorias en $\mathbb{Z}$

Aquí estudiaremos las caminatas aleatorias en $\mathbb{Z}$. En la sección 3.1 utilizaremos el principio de dualidad para obtener resultados elementales. En la sección 3.2 daremos un criterio para decir si una caminata aleatoria en $\mathbb{Z}$ es o no recurrente. La bibliografía de este capítulo es [4] y [6].

3.1. Dualidad en caminatas aleatorias

Definición 3.1. Sean $X_1, X_2, \ldots$ variables aleatorias independientes e idénticamente distribuidas con un mismo espacio de estados $S \subset \mathbb{Z}$ finito, y con $E[|X_i|] < \infty$. Sea $Y_0 = c$, y para $n \geq 1$ sean $Y_n = Y_0 + \sum_{i=1}^{n} X_i$. El proceso $Y = \{Y_n : n \geq 0\}$ es llamado **caminata aleatoria**.

Proposición 3.1. *La caminata aleatoria $Y = \{Y_n : n \geq 0\}$ es una cadena de Markov con probabilidad de transición $p_{ij} = a_{j-i}$, donde $a_i = P(X_1 = i)$.*

Demostración. Sean $x_1, \ldots, x_{n-1}, i, j$ estados arbitrarios en S entonces,

$$\begin{aligned}
&P(Y_{n+1} = j | Y_n = i, \ldots, Y_1 = x_1, Y_0 = c) \\
&= \frac{P(Y_{n+1} = j, Y_n = i, \ldots, Y_1 = x_1, Y_0 = c)}{P(Y_n = i, \ldots, Y_1 = x_1, Y_0 = c)} \\
&= \frac{P(X_{n+1} = j - i, X_n = i - x_{n-1}, \ldots, X_1 = x_1 - c)}{P(X_n = i - x_{n-1}, \ldots, X_1 = x_1 - c)} \\
&= P(X_{n+1} = j - i) \\
&= a_{j-i}.
\end{aligned}$$

En la tercera igualdad utilizamos la independencia de las X_i. Por otro lado,

$$P(Y_{n+1} = j | Y_n = i) = \frac{P(X_{n+1} = j - i, Y_n = i)}{P(Y_n = i)} = P(X_{n+1} = j - i) = a_{j-i}.$$

En la última igualdad volvimos a utilizar la independencia de las X_i. □

Observación. Si $c = Y_0$, diremos que la caminata aleatoria empieza en el estado c. De ahora en adelante (a menos que sea especificado lo contrario) sólo consideraremos caminatas aleatorias con $c = 0$, es decir, que empiecen en el origen.

Como $X_1, X_2, \ldots$ son independientes e idénticamente distribuidas tenemos que, la distribución de $(X_1, X_2, \ldots, X_n)$ es igual a la distribución de $(X_n, X_{n-1}, \ldots, X_1)$. Esta propiedad llamada **principio de dualidad** es muy útil y la emplearemos en la demostración de varios resultados referentes a la caminata aleatoria.

Sea R_n el número de distintos valores en el vector $(Y_0, \ldots, Y_n)$. Por ejemplo, si $Y_0 = 0$, $Y_1 = 1$, $Y_2 = 1$ y $Y_3 = 4$, entonces $R_3 = 3$. La siguiente proposición nos dice cómo calcular la probabilidad, en términos de R_n, de que la caminata aleatoria jamás regrese al origen.

Proposición 3.2.

$$\lim_{n\to\infty} \frac{E[R_n]}{n} = P(\textit{la caminata aleatoria nunca regrese a } 0).$$

Demostración. Sea

$$I_k = \begin{cases} 1 & \text{si } Y_k \neq Y_{k-1}, Y_k \neq Y_{k-2}, \ldots, Y_k \neq Y_0, \\ 0 & \text{en otro caso.} \end{cases}$$

Es claro que $R_n = 1 + \sum_{k=1}^{n} I_k$, por lo tanto

$$\begin{aligned} E[R_n] &= 1 + \sum_{k=1}^{n} E[I_k] \\ &= 1 + \sum_{k=1}^{n} P(Y_k \neq Y_{k-1}, Y_k \neq Y_{k-2}, \ldots, Y_k \neq Y_0) \\ &= 1 + \sum_{k=1}^{n} P(X_k \neq 0, X_k + X_{k-1} \neq 0, \ldots, X_k + X_{k-1} + \cdots + X_1 \neq 0) \\ &= 1 + \sum_{k=1}^{n} P(X_1 \neq 0, X_1 + X_2 \neq 0, \ldots, X_1 + \cdots + X_k \neq 0). \end{aligned}$$

En la última igualdad utilizamos el principio de dualidad. Sea T el tiempo del primer regreso al 0. Así,

$$E[R_n] = 1 + \sum_{k=1}^{n} P(Y_1 \neq 0, Y_2 \neq 0, \ldots, Y_k \neq 0) = \sum_{k=0}^{n} P(T > k).$$

Notemos que $[T > 1] \supset [T > 2] \supset \ldots$ lo que implica que

$$P(\text{nunca regresar al } 0) = P(T = \infty) = \lim_{k\to\infty} P(T > k).$$

Puesto que

$$\frac{E[R_n]}{n} = \frac{\sum_{k=1}^{n} kP(T = k) + (n + 1)P(T > n)}{n},$$

basta probar

$$\lim_{n\to\infty} \sum_{k=1}^{n} \frac{k}{n} P(T = k) = 0.$$

Sea $\alpha \in \mathbb{N}$ tal que $\sum_{k=\alpha}^{\infty} P(T = k) < \frac{\varepsilon}{2}$. Sea $\beta \in \mathbb{N}$ tal que $\frac{\alpha}{\beta} < \frac{\varepsilon}{2}$. Entonces, para todo $n \geq \text{máx}\{\alpha, \beta\}$ tenemos que

$$\begin{aligned}
\sum_{k=1}^{n} \frac{k}{n} P(T = k) &= \sum_{k=1}^{\alpha} \frac{k}{n} P(T = k) + \sum_{k=\alpha+1}^{n} \frac{k}{n} P(T = k) \\
&\leq \frac{\varepsilon}{2} \sum_{k=1}^{\alpha} P(T = k) + \sum_{k=\alpha+1}^{n} P(T = k) \\
&< \frac{\varepsilon}{2} + \frac{\varepsilon}{2}.
\end{aligned}$$

□

Definición 3.2. Una **caminata aleatoria simétrica** es una caminata aleatoria en $\mathbb{Z}$ tal que $p_{i,i+1} = \frac{1}{2} = 1 - p_{i,i-1}$ para todo $i \in \mathbb{Z}$.

Proposición 3.3. *En la caminata aleatoria simétrica, el número esperado de visitas al estado k antes de regresar al origen es 1 para todo $k \neq 0$.*

Demostración. Lo haremos para $k > 0$ el caso $k < 0$ es análogo. Sea

$$I_n = \begin{cases} 1 & \text{si al tiempo } n, \text{ la caminata visita el estado } k \\ & \text{sin antes haber visitado el origen,} \\ 0 & \text{en otro caso.} \end{cases}$$

Esto es equivalente a

$$I_n = \begin{cases} 1 & \text{si } Y_n > 0, Y_{n-1} > 0, \ldots, Y_1 > 0, Y_n = k, \\ 0 & \text{en otro caso.} \end{cases}$$

Sea Z el número de visitas al estado k antes de regresar al origen. Entonces $Z = \sum_{n=1}^{\infty} I_n$, así

$$\begin{aligned}
E[Z] &= \sum_{n=1}^{\infty} P(Y_n > 0, Y_{n-1} > 0, \ldots, Y_1 > 0, Y_n = k) \\
&= \sum_{n=1}^{\infty} P(X_n + \cdots + X_1 > 0, X_{n-1} + \cdots + X_1 > 0, \ldots, X_1 > 0, \\
&\qquad\qquad X_n + \cdots + X_1 = k) \\
&= \sum_{n=1}^{\infty} P(X_1 + \cdots + X_n > 0, X_2 + \cdots + X_n > 0, \ldots, X_n > 0, \\
&\qquad\qquad X_1 + \cdots + X_n = k) \\
&= \sum_{n=1}^{\infty} P(Y_n > 0, Y_n > Y_1, \ldots, Y_n > Y_{n-1}, Y_n = k) \\
&= \sum_{n=1}^{\infty} P(\text{la caminata aleatoria simétrica llegue a } k \text{ por vez primera} \\
&\qquad\qquad \text{al tiempo } n) \\
&= P(\text{la caminata aleatoria simétrica llegue a } k \text{ alguna vez}) \\
&= 1.
\end{aligned}$$

En la tercera igualdad utilizamos el principio de dualidad y en la última el hecho de que la caminata aleatoria simétrica es recurrente (teorema 3.5) cuya demostración es independiente a esta. □

Definición 3.3. Para $n \geq 1$, sea M_n el valor máximo que toma una caminata aleatoria hasta el tiempo n, es decir, $M_n = \max\{0, Y_1, \ldots, Y_n\}$.

Definición 3.4. Sea $X : \Omega \to \mathbb{R}$ una función. Definimos su **parte positiva** como $X^+ : \Omega \to \mathbb{R}$ definida para $\omega \in \Omega$ como $X^+(\omega) = \max\{0, X(\omega)\}$.

Proposición 3.4 (Identidad de Spitzer)**.** *La esperanza de M_n es igual a la suma desde $k = 1$ hasta $k = n$, de las esperanzas de las partes positivas de Y_k, multiplicadas cada una por $\frac{1}{k}$, es decir,*

$$E[M_n] = \sum_{k=1}^{n} \frac{1}{k} E[Y_k^+].$$

Demostración. La demostración se basa en descomponer M_n de forma adecuada. Sea $I_A(x)$ la función indicadora en A, es decir, $I_A(x) = 1$ si $x \in A$, y es 0 en otro caso. Entonces

$$M_n = I_{Y_n > 0} M_n + I_{Y_n \leq 0} M_n;$$

$$I_{Y_n>0}M_n = I_{Y_n>0}(X_1 + \max(0, X_2, \ldots, X_2 + \cdots + X_n)).$$

Aplicando la función esperanza a $I_{Y_n>0}M_n$, obtenemos

$$E[I_{Y_n>0}M_n] = E[I_{Y_n>0}X_1] + E[I_{Y_n>0}\max(0, X_2, \ldots, X_2 + \cdots + X_n)]. \quad (3.1)$$

Ahora $X_1, X_2, \ldots, X_n$ y $X_n, X_1, X_2 \ldots, X_{n-1}$ tienen la misma distribución conjunta. Por lo tanto,

$$E[I_{Y_n>0}\max(0, X_2, \ldots, X_2 + \cdots + X_n)] = E[I_{Y_n>0}M_{n-1}]. \quad (3.2)$$

Como las X_i tienen la misma distribución, y como la distribución de (X_i, Y_i) no depende de $i \in \{1, \ldots, n\}$ concluimos que,

$$E[Y_n I_{Y_n>0}] = E\left[\sum_{i=1}^{n} X_i I_{Y_n>0}\right] = nE[X_1 I_{Y_n>0}],$$

de forma que,

$$E[X_1 I_{Y_n>0}] = \frac{1}{n}E[Y_n^+]. \quad (3.3)$$

Sustituyendo las ecuaciones (3.2) y (3.3) en la ecuación (3.1) llegamos a

$$E[I_{Y_n>0}M_n] = E[I_{Y_n>0}M_{n-1}] + \frac{1}{n}E[Y_n^+],$$

y del hecho de que $I_{Y_n\leq 0}M_n = I_{Y_n\leq 0}M_{n-1}$, se tiene que

$$\begin{aligned} E[M_n] &= E[I_{Y_n>0}M_{n-1}] + \frac{1}{n}E[Y_n^+] + E[I_{Y_n\leq 0}M_{n-1}] \\ &= E[M_{n-1}] + \frac{1}{n}E[Y_n^+]. \end{aligned}$$

Repitiendo este proceso para $2 \leq k \leq n-1$ y teniendo en cuenta que $M_1 = Y_1^+$ obtenemos el resultado. □

3.2. Un teorema sobre recurrencia

Definición 3.5. Un proceso estocástico $Z = \{Z_n : n \geq 1\}$ es una **martingala** si para todo $n \geq 1$ se cumple que $E[|Z_n|] < \infty$ y $E[Z_{n+1}|Z_1, \ldots, Z_n] = Z_n$.

Es inmediato de la definición que $E[Z_n] = E[Z_1]$ para todo n.

Ejemplo 3.1. Sean $X_1, X_2, \ldots$ variables aleatorias independientes con media 0. Sea $Z_n = \sum_{i=1}^{n} X_i$, entonces $Z = \{Z_n : n \geq 1\}$ es martingala. Esto se debe a que

$$\begin{aligned} E[Z_{n+1}|Z_1, \ldots, Z_n] &= E[Z_n + X_{n+1}|Z_1, \ldots, Z_n] \\ &= E[Z_n|Z_1, \ldots, Z_n] + E[X_{n+1}|Z_1, \ldots, Z_n] \\ &= Z_n + E[X_{n+1}]. \end{aligned}$$

Ejemplo 3.2. Sean $X_1, X_2, \ldots$ variables aleatorias independientes con media 1. Sea $Z_n = \prod_{i=1}^{n} X_i$, entonces $Z = \{Z_n : n \geq 1\}$ es martingala. Esto se debe a que

$$\begin{aligned} E[Z_{n+1}|Z_1, \ldots, Z_n] &= E[Z_n X_{n+1}|Z_1, \ldots, Z_n] \\ &= Z_n E[X_{n+1}|Z_1, \ldots, Z_n] \\ &= Z_n E[X_{n+1}]. \end{aligned}$$

Teorema 3.5. *Sean $X_1, X_2, \ldots$ variables aleatorias independientes e idénticamente distribuidas, las cuales pueden tomar valores en $0, \pm 1, \pm 2, \ldots, \pm M < \infty$. Entonces $Y = \{Y_n : n \geq 0\}$ es una cadena de Markov recurrente, si y sólo si, $E[X] = 0$.*

Demostración. Si $E[X] > 0$, entonces, por la Ley Fuerte de los Grandes Números, $Y_n \to \infty$ cuando $n \to \infty$ con probabilidad 1 y por consiguiente la cadena es transitoria. El caso $E[X] < 0$ es análogo.

Supongamos ahora que $E[X] = 0$, lo cual implica que $Y = \{Y_n : n \geq 1\}$ es martingala. Sea

$$A = \{-M, -(M-1), \ldots, -1\}.$$

Supongamos que el proceso empieza en el estado $i \geq 0$. Para $j > i$, sea

$$A_j = \{j, j+1, \ldots, j+M\}$$

y sea τ la variable aleatoria que indica la primera vez que entramos en A o A_j. Si sabemos los valores de $Y_1, \ldots, Y_n$, es claro que podemos decir si $\tau = n$ o no. Así, τ es tiempo de paro de Y_n. Además la propiedad (i) del teorema A.2 se cumple, ya que para $n \geq 0$ se tiene

$$-M < Y_{n \wedge \tau} < j + M,$$

lo que implica que

$$E[Y_\tau] = E[Y_0] = i.$$

Ahora,

$$\begin{aligned} i &= E[Y_\tau | Y_\tau \in A]P(Y_\tau \in A) + E[Y_\tau | Y_\tau \in A_j]P(Y_\tau \in A_j) \\ &\geq -MP(Y_\tau \in A) + j(1 - P(Y_\tau \in A)). \end{aligned}$$

Aislando $P(Y_\tau \in A)$ llegamos a

$$P(Y_\tau \in A) \geq \frac{j-i}{j+M}.$$

Por tanto,

$$P(\text{el proceso entre a } A) \geq P(Y_\tau \in A) \geq \frac{j-i}{j+M}.$$

Lo importante de esta desigualdad es que vale para toda $j > i \geq 0$. Haciendo $j \to \infty$ llegamos a

$$P(\text{el proceso entre a } A | \text{empezó en } i \geq 0) = 1.$$

Sea $B = \{1, 2, \ldots, M\}$. Repitiendo el mismo argumento, llegamos a

$$P(\text{el proceso entre a } B | \text{empezó en } i \leq 0) = 1.$$

Lo que implica que

$$P(\text{el proceso entre a } A \cup B | \text{empezó en } i \in \mathbb{Z}) = 1.$$

Lo anterior implica (se demostrará) que el conjunto finito $A \cup B$ será visitado infinitas veces. Si la cadena fuera transitoria, entonces todo conjunto finito sería visitado un número finito de veces y como esto no sucede, llegamos a que el proceso es recurrente.

Sólo falta demostrar que $A \cup B$ será visitado una infinidad de veces. Una forma de hacerlo es la siguiente, sea $N(A \cup B)$ la variable aleatoria que cuenta el número de visitas a $A \cup B$, entonces para cualquier $\omega \in [N(A \cup B) < \infty]$ se cumple que existe $k \in \mathbb{N}$ tal que si $n > k$ entonces $Y_n(\omega) \notin A \cup B$, por lo tanto

$$
\begin{aligned}
P[N(A\cup B)<\infty] &\leq P[Y_0\in S, Y_1\notin A\cup B, Y_2\notin A\cup B,\ldots]\\
&\quad + P[Y_0\in S, Y_1\in S, Y_2\notin A\cup B, Y_3\notin A\cup B,\ldots]\\
&\quad\quad +\cdots\\
&= \sum_{x\in S} P[Y_0=x, Y_1\notin A\cup B,\ldots]\\
&\quad + \sum_{x\in S} P[Y_1=x, Y_2\notin A\cup B,\ldots]+\cdots\\
&= \sum_{x\in S} P[\textit{no regresar a } A\cup B|Y_0=x]P[Y_0=x]\\
&\quad + \sum_{x\in S} P[\textit{no regresar a } A\cup B|Y_1=x]P[Y_1=x]\\
&\quad\quad +\cdots\\
&= 0.
\end{aligned}
$$

□

Capítulo 4

Caminatas aleatorias en grupos

En este capítulo presentamos los resultados referentes a caminatas aleatorias en grupos. En la sección 4.1 se define qué es una caminata aleatoria en un grupo finito y vemos ciertas propiedades elementales de dicho concepto. La sección finaliza introduciendo la distancia de variación total, noción que será útil para dar criterios para la convergencia de distribuciones relacionadas con caminatas aleatorias. En la sección 4.2 probaremos el lema de la cota superior (e inferior), el cual, como su nombre lo dice, acota la distancia entre la distribución al tiempo n de la caminata aleatoria y la distribución uniforme en términos de variación total. En la sección 4.3 damos dos aplicaciones del lema de la cota superior y estimamos en cuánto tiempo se alcanza el estado estacionario. La primera aplicación es en la caminata aleatoria simple y simétrica en $\mathbb{Z}_p$, mientras que la segunda es sobre cierta caminata aleatoria en $\mathbb{Z}_2^d$. La bilbliografía de este capítulo es [1].

4.1. Fundamentos

Antes de definir formalmente qué es una caminata aleatoria en un grupo finito veamos una definición informal pero ilustrativa. Sean G un grupo finito con operación $\bullet$, y Q una probabilidad en G. Caminar aleatoriamente en G significa movernos en el grupo G de la siguiente manera; supongamos que al instante $n = 0$ nos encontramos en el neutro $e \in G$. Ahora, para el instante $n = 1$ nos moveremos de e al elemento $s_1 \in G$ con probabilidad $Q(s_1)$. Para el instante $n = 2$ nos moveremos de $s_1 \in G$ al elemento $s_2 \bullet s_1 \in G$ con probabilidad $Q(s_2)$. En general, si en el instante $n = k$ estamos en $s_k \bullet \cdots \bullet s_1 \in G$, entonces nos moveremos al elemento $s_{k+1} \bullet s_k \bullet \cdots \bullet s_1 \in G$ con probabilidad $Q(s_{k+1})$.

Visto de esta manera, es natural hablar de la caminata aleatoria generada

por el grupo G y la probabilidad Q. Ahora bien:

¿Cuál es la probabilidad de que la caminata aleatoria se encuentre en el estado $s \in G$ al tiempo $n = k$?

¿Se puede encontrar un n tal que para todo $k > n$ la distribución de la caminata aleatoria generada por Q al tiempo k "diste en menos de ε"[1] de la distribución uniforme?

La primera pregunta es fácil de responder. Veremos que la respuesta afirmativa es la k-ésima convolución de Q, es decir, $Q^{*k}(s)$. La segunda pregunta es más difícil, dependerá eventualmente del grupo G y de cómo sea Q.

Definición 4.1. Sean G un grupo finito con operación $\bullet$ y Q una probabilidad en G. Para $i \geq 1$, sean $X_i : \Omega \to G$ variables aleatorias independientes e idénticamente distribuidas con distribución común Q. En particular, para todo $i \geq 1$ y para cualquier $s \in G$ se tiene que

$$P(X_i = s) = Q(s). \tag{4.1}$$

Para $n \geq 1$, sea

$$Y_n = X_n \bullet \cdots \bullet X_1. \tag{4.2}$$

Decimos que el proceso $Y = \{Y_n : n \geq 1\}$ es **la caminata aleatoria en** G, **generada por** Q.

Proposición 4.1. *La caminata aleatoria en G generada por Q es una cadena de Markov con probabilidad de transición $p_{ij} = Q(ji^{-1})$.*

Demostración. Sean $x_1, \ldots, x_{n-1}, i, j$ elementos arbitrarios en G entonces,

$$\begin{aligned}
&P(Y_{n+1} = j | Y_n = i, Y_{n-1} = x_{n-1} \ldots, Y_1 = x_1) \\
&= \frac{P(Y_{n+1} = j, Y_n = i, Y_{n-1} = x_{n-1} \ldots, Y_1 = x_1)}{P(Y_n = i, \ldots, Y_1 = x_1)} \\
&= \frac{P(X_{n+1} = ji^{-1}, X_n = ix_{n-1}^{-1}, \ldots, X_2 = x_2x_1^{-1}, X_1 = x_1)}{P(X_n = ix_{n-1}^{-1}, \ldots, X_2 = x_2x_1^{-1}, X_1 = x_1)} \\
&= P(X_{n+1} = ji^{-1}) \\
&= Q(ji^{-1}).
\end{aligned}$$

En la segunda igualdad, para $1 \leq i \leq n+1$, escribimos Y_n en función de las X_i de la ecuación (4.2). En la tercera igualdad utilizamos la independencia

[1] Posteriormente definiremos formalmente una métrica en el conjunto de las medidas de probabilidad en G.

de las X_i. Por otro lado,

$$\begin{aligned} P(Y_{n+1} = j | Y_n = i) &= \frac{P(Y_{n+1} = j, Y_n = i)}{P(Y_n = i)} \\ &= \frac{P(X_{n+1} = ji^{-1}, Y_n = i)}{P(Y_n = i)} \\ &= P(X_{n+1} = ji^{-1}) \\ &= Q(ji^{-1}). \end{aligned}$$

En la segunda igualdad, utilizamos el hecho de que $Y_{n+1} = X_{n+1} \bullet Y_n$ y en la tercera volvimos a utilizar la independencia de las X_i. □

Observación. Hemos considerado caminatas aleatorias que empiezan en el neutro $e \in G$. Podemos considerar también caminatas aleatorias que empiecen en cualquier $s \in G$, modificando Y_n por

$$Y'_n = Y_n \bullet s = X_n \bullet \cdots \bullet X_1 \bullet s.$$

La prueba de que $Y' = \{Y'_n : n \geq 1\}$ es cadena de Markov es análoga. En general consideraremos caminatas aleatorias que empiezan en el neutro $e \in G$ a menos que sea especificado lo contrario.

A continuación resolvemos la primera pregunta que planteamos al principio de la sección.

Proposición 4.2. *La probabilidad de que la caminata aleatoria en el grupo G generada por Q al instante n, se encuentre en $s \in G$, es $Q^{*n}(s)$, es decir,*

$$P(Y_n = s) = Q^{*n}(s).$$

Demostración. Lo haremos por inducción sobre n. El caso $n = 1$ es trivial. Por definición tenemos que $Q^{*1} = Q$, así, la distribución de Q es igual a la distribución de X_1 pero $X_1 = Y_1$ y por tanto, el resultado es cierto para $n = 1$. Supongamos ahora que el resultado es cierto para n. Entonces tenemos que

$$\begin{aligned} Q^{*n+1}(s) &= Q * Q^{*n}(s) \\ &= \sum_{t \in G} Q(st^{-1}) Q^{*n}(t) \\ &= \sum_{t \in G} Q(st^{-1}) P(Y_n = t) \\ &= \sum_{t \in G} P(X_{n+1} = st^{-1}) P(Y_n = t) \\ &= \sum_{t \in G} P(X_{n+1} = st^{-1}, Y_n = t) = P(Y_{n+1} = s). \end{aligned}$$

La primera y segunda igualdad son por definición, mientras que en la tercera utilizamos la hipótesis de inducción. La cuarta igualdad se da por cómo definimos las X_i ecuación (4.1). Para la quinta igualdad utilizamos el hecho de que las X_i son independientes. □

Definición 4.2. Sean G un grupo finito y P, Q probabilidades en G. La **distancia de variación total** entre P y Q se define como

$$||P - Q||_{TV} := \max_{A \subset G} |P(A) - Q(A)|.$$

Esta definición será muy usada, por lo que veremos algunas de sus propiedades.

Proposición 4.3. *Sean $f : G \to [-1, 1]$ y $P(f) = \sum_{s \in G} f(s)P(s)$ el valor esperado de f bajo P, entonces*

$$||P - Q||_{TV} = \frac{1}{2} \sum_{s \in G} |P(s) - Q(s)| = \frac{1}{2} \max_{|f| \leq 1} |P(f) - Q(f)|.$$

Demostración. Como $P(A^c) = 1 - P(A)$, tenemos que

$$|P(A) - Q(A)| = |P(A^c) - Q(A^c)|.$$

Sea

$$A = \{s \in G : P(s) \geq Q(s)\}.$$

Sea $B \subseteq G$, entonces

$$\begin{aligned} |P(B) - Q(B)| &= |(P(A \cap B) - Q(A \cap B)) + (P(A^c \cap B) - Q(A^c \cap B))| \\ &\leq \max\{|P(A \cap B) - Q(A \cap B)|, |P(A^c \cap B) - Q(A^c \cap B)|\} \\ &\leq \max\{|P(A) - Q(A)|, |P(A^c) - Q(A^c)|\} \\ &= |P(A) - Q(A)|. \end{aligned}$$

La primera desigualdad se da porque

$$(P(A^c \cap B) - Q(A^c \cap B)) \leq 0 \leq (P(A \cap B) - Q(A \cap B)).$$

La segunda desigualdad se debe a las siguientes dos desigualdades

$$\begin{aligned} |P(A \cap B) - Q(A \cap B)| &= P(A \cap B) - Q(A \cap B) \\ &\leq P(A \cap B) - Q(A \cap B) \\ &\quad + P(A - A \cap B) - Q(A - A \cap B) \\ &= P(A) - Q(A). \end{aligned}$$

Análogamente

$$|P(A^c \cap B) - Q(A^c \cap B)| \leq Q(A^c) - P(A^c).$$

Dado que $A = \{s \in G : P(s) \geq Q(s)\}$ es el conjunto en el que se alcanza el máximo valor de $|P(\cdot) - Q(\cdot)|$, es inmediato que

$$\begin{aligned} 2||P - Q||_{TV} &= |P(A) - Q(A)| + |P(A^c) - Q(A^c)| \\ &= P(A) - Q(A) + Q(A^c) - P(A^c) \\ &= \sum_{s \in G} |P(s) - Q(s)|, \end{aligned}$$

lo que prueba la primera igualdad.

Sean $f : G \to [-1, 1]$ arbitraria y $h : G \to [-1, 1]$ dada por

$$h(s) = \begin{cases} 1 & \text{si } s \in A, \\ -1 & \text{si } s \in A^c. \end{cases}$$

Note que

$$\begin{aligned} |P(h) - Q(h)| &= \left| \sum_{s \in G} h(s)(P(s) - Q(s)) \right| \\ &= \left| \sum_{s \in A} h(s)(P(s) - Q(s)) + \sum_{s \in A^c} h(s)(P(s) - Q(s)) \right| \\ &= \sum_{s \in G} |P(s) - Q(s)|. \end{aligned}$$

Tenemos entonces,

$$\begin{aligned} |P(f) - Q(f)| &= \left| \sum_{s \in G} f(s)(P(s) - Q(s)) \right| \\ &\leq \sum_{s \in G} |f(s)||P(s) - Q(s)| \\ &\leq \sum_{s \in G} |P(s) - Q(s)| \\ &= |P(h) - Q(h)|. \end{aligned}$$

Por tanto,

$$|P(h) - Q(h)| = \max_{||f|| \leq 1} |P(f) - Q(f)|. \qquad \square$$

Corolario 4.4. *La distancia de variación total $||\cdot,\cdot||_{TV}$ es una métrica en el conjunto de las medidas de probabilidad en G, es decir, para cualesquiera P, Q y H probabilidades en G, tenemos que*

(i) $||P-Q||_{TV} = ||Q-P||_{TV}$,

(ii) $||P-Q||_{TV} = 0$ *si, y sólo si,* $P = Q$,

(iii) $||P-Q||_{TV} \leq ||P-H||_{TV} + ||H-Q||_{TV}$. *Adicionalmente,*

(iv) $0 \leq ||P-Q||_{TV} \leq 1$.

Demostración. En todos los casos utilizaremos la proposición 4.3.

(i)

$$||P-Q||_{TV} = \frac{1}{2}\sum_{s\in G}|P(s)-Q(s)| = \frac{1}{2}\sum_{s\in G}|Q(s)-P(s)| = ||Q-P||_{TV}.$$

(ii) $||P-Q||_{TV} = 0$, si y sólo si, $\frac{1}{2}\sum_{s\in G}|P(s)-Q(s)| = 0$, si y sólo si, $|P(s)-Q(s)| = 0$ para todo $s \in G$.

(iii) Note que

$$\begin{aligned} ||P-Q||_{TV} &= \frac{1}{2}\sum_{s\in G}|P(s)-Q(s)| \\ &= \frac{1}{2}\sum_{s\in G}|P(s)-H(s)+H(s)-Q(s)| \\ &\leq \frac{1}{2}\sum_{s\in G}(|P(s)-H(s)|+|H(s)-Q(s)|) \\ &= ||P-H||_{TV} + ||H-Q||_{TV}. \end{aligned}$$

(iv) Por definición tenemos que $||P-Q||_{TV} \geq 0$. Ahora

$$||P-Q||_{TV} = \frac{1}{2}\sum_{s\in G}|P(s)-Q(s)| \leq \frac{1}{2}\sum_{s\in G}(|P(s)|+|Q(s)|) = \frac{1}{2}(2) = 1.$$

□

Observación. Ahora podemos hablar formalmente acerca de la distancia entre dos distribuciones definidas en G. Recuerde de la proposición 4.2, que $Q^{*n}(s)$ es la probabilidad de que la caminata aleatoria en G generada por Q

al instante n, se encuentre en $s \in G$. Teniendo esto en cuenta, note que el número

$$\varepsilon = ||Q^{*n} - U||_{TV} \tag{4.3}$$

se interpreta como "la distribución de la caminata aleatoria en G generada por Q al tiempo n, dista en ε de la distribución uniforme U." Gran parte de lo que resta del trabajo nos dedicaremos a ver en qué casos se puede hacer tender ε a 0 conforme n tiende a ∞.

El siguiente es un ejemplo muy útil para ver cómo funciona la distancia de variación total $||Q - U||_{TV}$.

Ejemplo 4.1. Encontrar la variación total $||Q-U||_{TV}$ para los siguientes dos casos. Sean $G = S_n$, U la distribución uniforme en S_n, es decir, $U(\sigma) = \frac{1}{n!}$, $\sigma \in S_n$ y $A \subseteq \{1, \ldots, n\}$, con $|A| = k \leq n$. Definimos las medidas de probabilidad Q en S_n dadas por

(i) "La primera carta está hasta el frente, y todas las demás son aleatorias."

$$Q(\sigma) = \begin{cases} \frac{1}{(n-1)!} & \text{si } \sigma(1) = 1, \\ 0 & \text{en otro caso.} \end{cases}$$

Para encontrar $||Q - U||_{TV}$ sólo hay que considerar dos casos. Cuando $\sigma(1) = 1$ y cuando $\sigma(1) \neq 1$. Note que

$$|\{\sigma \in S_n : \sigma(1) = 1\}| = (n-1)!$$

y que

$$|\{\sigma \in S_n : \sigma(1) \neq 1\}| = n! - (n-1)! = (n-1)!(n-1).$$

Utilizando la proposición 4.3, tenemos que

$$\begin{aligned} ||Q - U||_{TV} &= \frac{1}{2} \sum_{\sigma \in S_n} |Q(\sigma) - U(\sigma)| \\ &= \frac{1}{2} \sum_{\sigma:\sigma(1)=1} |Q(\sigma) - U(\sigma)| + \frac{1}{2} \sum_{\sigma:\sigma(1)\neq 1} |Q(\sigma) - U(\sigma)| \\ &= \frac{1}{2}(n-1)! \left| \frac{1}{(n-1)!} - \frac{1}{n!} \right| + \frac{1}{2}(n-1)!(n-1) \left| 0 - \frac{1}{n!} \right| \\ &= \frac{1}{2}(n-1)! \frac{(n-1)}{n!} + \frac{1}{2} \frac{(n-1)}{n} \\ &= \frac{1}{2} \frac{(n-1)}{n} + \frac{1}{2} \frac{(n-1)}{n} = 1 - \frac{1}{n}. \end{aligned}$$

Note que la distancia de variación total entre Q y U, que es $1 - \frac{1}{n}$, concuerda con la intuición, pues Q se distribuye uniformemente sobre $\sigma(1) = 1$.

(ii) "La primera carta es distribuida aleatoriamente en A, y todas las demás son aleatorias."

$$Q(\sigma) = \begin{cases} \frac{1}{k(n-1)!} & \text{si } \sigma(1) \in A, \\ 0 & \text{en otro caso.} \end{cases}$$

Evidentemente este caso es, una generalización de (i). Para encontrar $||Q - U||_{TV}$ hay que considerar dos casos. Cuando $\sigma(1) \in A$ y cuando $\sigma(1) \notin A$. Note que

$$|\{\sigma \in S_n : \sigma(1) \in A\}| = k(n-1)!$$

y que

$$|\{\sigma \in S_n : \sigma(1) \notin A\}| = n! - k(n-1)! = (n-1)!(n-k).$$

Utilizando la proposición 4.3, tenemos que

$$\begin{aligned} ||Q - U||_{TV} &= \frac{1}{2} \sum_{\sigma \in S_n} |Q(\sigma) - U(\sigma)| \\ &= \frac{1}{2} \sum_{\sigma:\sigma(1)\in A} |Q(\sigma) - U(\sigma)| + \frac{1}{2} \sum_{\sigma:\sigma(1)\notin A} |Q(\sigma) - U(\sigma)| \\ &= \frac{1}{2} k(n-1)! \left| \frac{1}{k(n-1)!} - \frac{1}{n!} \right| + \frac{1}{2}(n-1)!(n-k) \left| 0 - \frac{1}{n!} \right| \\ &= \frac{1}{2} k(n-1)! \left| \frac{n}{kn!} - \frac{k}{kn!} \right| + \frac{1}{2}(n-1)!(n-k)\frac{1}{n!} \\ &= \frac{1}{2n}(n-k) + \frac{n-k}{2n} = 1 - \frac{k}{n}. \end{aligned}$$

Observe que la distancia de variación total entre Q y U, que es $1 - \frac{k}{n}$, concuerda con la intuición, pues Q se distribuye uniformemente sobre $\sigma(1) \in A$ (además de que $|A| = k$). Finalmente, si $A = \{1, \ldots, n\}$ entonces

$$||Q - U||_{TV} = 1 - \frac{n}{n} = 0,$$

lo cual es lógico ya que $Q = U$.

4.2. Lema de la cota superior

En esta sección acotaremos $||Q - U||_{TV}$ (más precisamente, acotaremos $||Q - U||_{TV}^2$) para cualquier probabilidad Q en G, es decir, encontraremos

$$0 \leq \varepsilon_0 \leq \varepsilon_1 \leq 1,$$

tales que

$$\varepsilon_0 \leq ||Q - U||_{TV}^2 \leq \varepsilon_1.$$

Los números ε_0 y ε_1 dependerán de quién sea G y de cómo sea Q. La siguiente proposición nos será de gran ayuda para encontrar dichas cotas.

Proposición 4.5. *Sean G un grupo finito, y Q cualquier probablidad en G, entonces*

$$\sum_{\rho} d_\rho Tr(\widehat{Q}(\rho)\widehat{Q}(\rho)^*) = |G| \sum_{s \in G} Q^2(s) - 1.$$

Donde la suma en 'ρ' es sobre todas las representaciones irreducibles de G, salvo la trivial.

Demostración. Debido a la observación *4.* del teorema 1.1, podemos asumir que nuestras representaciones son unitarias, es decir, para cualquier ρ representación irreducible de G y para cualquier $s \in G$ tenemos que la matriz $\rho(s)$ es unitaria, es decir,

$$\overline{\rho(s)^t} = \rho(s)^* = \rho(s)^{-1} = \rho(s^{-1}). \tag{4.4}$$

Definimos la probabilidad H en G dada por

$$H(s) = Q(s^{-1}) \text{ para todo } s \in G.$$

Note que

$$\widehat{Q}(\rho)^* = \overline{\left[\sum_{s \in G} Q(s)\rho(s)\right]}^t = \sum_{s \in G} Q(s)\overline{\rho(s)^t} = \sum_{s \in G} Q(s)\rho(s^{-1}) = \widehat{H}(\rho). \tag{4.5}$$

La primera igualdad es por definición, mientras que la segunda se da por la linealidad, tanto de la transpuesta como de la conjugada. La tercera igualdad se da en virtud de que las representaciones son unitarias, ecuación (4.4). La cuarta igualdad se da por cómo definimos H. Si $\rho = 1$ es la representación

tivial, entonces

$$\begin{aligned} d_\rho Tr(\widehat{Q}(\rho)\widehat{Q}(\rho)^*) &= d_\rho Tr\left[\left(\sum_{s\in G} Q(s)\rho(s)\right)\left(\sum_{s\in G} Q(s)\overline{\rho(s)^t}\right)\right] \\ &= (1)Tr\left[\left(\sum_{s\in G} Q(s)(1)\right)\left(\sum_{s\in G} Q(s)\overline{1^t}\right)\right] \\ &= (1)Tr\left[\left(\sum_{s\in G} Q(s)\right)\left(\sum_{s\in G} Q(s)\right)\right] \\ &= 1. \end{aligned} \tag{4.6}$$

Por último, recuerde que la representación trivial es irreducible. Si ρ' recorre todas las representaciones irreducibles de G (incluyendo la trivial) obtenemos

$$\begin{aligned} \sum_\rho d_\rho Tr(\widehat{Q}(\rho)\widehat{Q}(\rho)^*) &= \sum_{\rho'} d_{\rho'} Tr(\widehat{Q}(\rho')\widehat{Q}(\rho')^*) \ - d_1 Tr(\widehat{Q}(1)\widehat{Q}(1)^*) \\ &= |G|\frac{1}{|G|}\sum_{\rho'} d_{\rho'} Tr(\widehat{Q}(\rho')\widehat{H}(\rho')) - 1 \\ &= |G|\sum_{s\in G} Q(s^{-1})H(s) - 1 \\ &= |G|\sum_{s\in G} Q(s^{-1})Q(s^{-1}) - 1 \\ &= |G|\sum_{s\in G} Q^2(s^{-1}) - 1 = |G|\sum_{s\in G} Q^2(s) - 1. \end{aligned}$$

La primera igualdad se debe a que, es lo mismo sumar sobre todas las irreducibles salvo la trivial (parte izquierda), a sumar sobre todas las irreducibles y restarle el término correspondiente de la trivial (parte derecha). La segunda igualdad se da por la ecuación (4.6). La tercera ecuación consiste simplemente en aplicar la fórmula de Plancherel, teorema 1.19 inciso (ii). □

Lema 4.6 (Lema de la cota superior-Diaconis y Shahshahani)**.** *Sean G un grupo finito, U la probabilidad uniforme en G y Q cualquier probablidad en G, entonces*

$$||Q - U||_{TV}^2 \leq \frac{1}{4}\sum_\rho d_\rho Tr(\widehat{Q}(\rho)\widehat{Q}(\rho)^*).$$

Donde la suma en 'ρ' es sobre todas las representaciones irreducibles de G, salvo la trivial.

Demostración. Definimos las funciones $f : G \to \mathbb{R}$ y $g : G \to \mathbb{R}$ dadas por

$$f(s) = |Q(s) - U(s)| \tag{4.7}$$

$$g(s) = 1 \tag{4.8}$$

para todo $s \in G$. Ahora, con el producto interior usual de funciones de G a $\mathbb{R}$,

$$\langle g, f \rangle = \sum_{s \in G} g(s) f(s),$$

tenemos la desigualdad de Cauchy-Schwarz que nos afirma

$$| \langle g, f \rangle | \leq ||g|| \cdot ||f||.$$

Donde $||f|| = \sqrt{\langle f, f \rangle}$ es la norma de f. Aplicando la desigualdad de Cauchy-Schwarz para f y g definidas como en (4.7) y (4.8), llegamos a,

$$\sum_{s \in G} |Q(s) - U(s)| \leq \sqrt{|G|} \sqrt{\sum_{s \in G} |Q(s) - U(s)|^2}.$$

Elevando al cuadrado obtenemos

$$\begin{aligned}
\left[\sum_{s \in G} |Q(s) - U(s)| \right]^2 &\leq |G| \left[\sum_{s \in G} |Q(s) - U(s)|^2 \right] \\
&= |G| \left[\sum_{s \in G} (Q^2(s) - 2Q(s)U(s) + U^2(s)) \right] \\
&= |G| \left[\sum_{s \in G} (Q^2(s) - \frac{2Q(s)}{|G|} + \frac{1}{|G|^2}) \right] \\
&= |G| \left[\sum_{s \in G} Q^2(s) + \sum_{s \in G} (\frac{1}{|G|^2} - \frac{2Q(s)}{|G|}) \right] \\
&= |G| \left[\sum_{s \in G} Q^2(s) - \frac{1}{|G|} \right] \\
&= |G| \sum_{s \in G} Q^2(s) - 1 = \sum_{\rho} d_\rho Tr(\widehat{Q}(\rho) \widehat{Q}(\rho)^*). \qquad (4.9)
\end{aligned}$$

La última de las igualdades se da por la proposición 4.5. Todas las demás igualdades son cuentas sencillas. Finalmente, por la proposición 4.3, tenemos que

$$||Q - U||_{TV} = \frac{1}{2} \sum_{s \in G} |Q(s) - U(s)|.$$

Por lo tanto,

$$4||Q-U||_{TV}^2 = \left[\sum_{s\in G}|Q(s)-U(s)|\right]^2. \tag{4.10}$$

Sustitúyase el lado izquierdo de la ecuación (4.10) en el lado izquierdo de la desigualdad (4.9). □

Observaciones. *1.* Sea G cualquier grupo finito y Q cualquier probabilidad en G. Supongamos que tenemos información que nos hace pensar que la caminata aleatoria en G generada por Q converge a la distribución uniforme. Una forma para verificar lo anterior es utilizar el lema de la cota superior. Para eso, hay que encontrar todas las representaciones irreducibles de G y acotar el término

$$||Q^{*n}-U||_{TV}^2 \leq \frac{1}{4}\sum_{\rho} d_\rho Tr(\widehat{Q^{*n}}(\rho)\widehat{Q^{*n}}(\rho)^*), \tag{4.11}$$

donde la suma en 'ρ' es sobre todas las representaciones irreducibles de G, salvo la trivial. Más importante aún, el término a la derecha en (4.11) está en función de n, por lo que si converge, sabremos qué tan rápido converge.

2. El término $\widehat{Q^{*n}}(\rho)$ no es tan complicado como parece, recuerde que el corolario 1.17 nos dice que $\widehat{Q^{*n}}(\rho) = [\widehat{Q}(\rho)]^n$.

Lema 4.7 (Lema de la cota inferior)**.** *Sean G un grupo finito, U la probabilidad uniforme en G y Q cualquier probablidad en G, entonces*

(i)
$$||Q-U||_{TV} \geq \frac{1}{2|G|}\sum_{\rho} d_\rho Tr(\widehat{Q}(\rho)\widehat{Q}(\rho)^*).$$

(ii)
$$||Q-U||_{TV}^2 \geq \frac{1}{4|G|}\sum_{\rho} d_\rho Tr(\widehat{Q}(\rho)\widehat{Q}(\rho)^*).$$

Donde la suma en 'ρ' es sobre todas las representaciones irreducibles de G, salvo la trivial.

Demostración. (i) Tenemos que

$$\begin{aligned}\sum_{s\in G}|Q(s)-U(s)|^2 &= \sum_{s\in G}(Q^2(s)-2Q(s)U(s)+U^2(s))\\ &= \sum_{s\in G}(Q^2(s)-\frac{2Q(s)}{|G|}+\frac{1}{|G|^2})\\ &= \sum_{s\in G}Q^2(s)+\sum_{s\in G}(\frac{1}{|G|^2}-\frac{2Q(s)}{|G|})\\ &= \sum_{s\in G}Q^2(s)-\frac{1}{|G|}. \qquad (4.12)\end{aligned}$$

Como $|Q(s)-U(s)|^2 \leq |Q(s)-U(s)|$ para todo $s \in G$, obtenemos por la ecuación (4.12) que

$$\sum_{s\in G}Q^2(s)-\frac{1}{|G|}=\sum_{s\in G}|Q(s)-U(s)|^2 \leq \sum_{s\in G}|Q(s)-U(s)|.$$

Por lo tanto,

$$\frac{1}{|G|}\left[|G|\sum_{s\in G}Q^2(s)-1\right] \leq \sum_{s\in G}|Q(s)-U(s)| = 2||Q-U||_{TV}.$$

Utilizando la proposición 4.5 llegamos a

$$\frac{1}{|G|}\sum_{\rho}d_\rho Tr(\widehat{Q}(\rho)\widehat{Q}(\rho)^*)=\frac{1}{|G|}\left[|G|\sum_{s\in G}Q^2(s)-1\right]\leq 2||Q-U||_{TV}.$$

(ii) Note que

$$\begin{aligned}4||Q-U||_{TV}^2 &= \left[\sum_{s\in G}|Q(s)-U(s)|\right]^2\\ &\geq \sum_{s\in G}|Q(s)-U(s)|^2\\ &= \sum_{s\in G}Q(s)^2-\frac{1}{|G|}\\ &= \frac{1}{|G|}\left[|G|\sum_{s\in G}Q^2(s)-1\right]\\ &= \frac{1}{|G|}\sum_{\rho}d_\rho Tr(\widehat{Q}(\rho)\widehat{Q}(\rho)^*).\end{aligned}$$

La primera igualdad se da en virtud de la proposición 4.3. La tercera se debe a la ecuación (4.12), y la última se da por la proposición 4.5. □

4.3. Dos aplicaciones del lema de la cota superior

El siguiente teorema nos habla de la caminata aleatoria simple y simétrica en $\mathbb{Z}_p$ con p impar (el caso cuando p es par lo tratamos en la observación posterior al teorema 4.8). El inciso (i) nos dice que se necesitan $n = p^2$ pasos para que dicha caminata aleatoria tenga una distribución cercana a la uniforme en términos de variación total. El inciso (ii) nos dice que aún cuando el límite sea la distribución uniforme, la distribución de la caminata aleatoria al tiempo n, es distinta de la distribución uniforme para todo $n \in \mathbb{N}$.

Teorema 4.8. *Sean $\mathbb{Z}_p$ el grupo de enteros módulo p, y Q la probabilidad en $\mathbb{Z}_p$ definida por*

$$Q(1) = \frac{1}{2} = Q(-1), \text{ y cero en otro caso.}$$

(i) *Para $n \geq p^2$, con p impar tenemos*

$$||Q^{*n} - U||_{TV} \leq \exp(-\frac{\pi^2 n}{2p^2}).$$

(ii) *Si $p \geq 7$, entonces para cualquier n,*

$$||Q^{*n} - U||_{TV} \geq \frac{1}{2}\exp(-\frac{\pi^2 n}{2p^2} - \frac{\pi^4 n}{11p^4}).$$

Demostración. (i) Como $\mathbb{Z}_p$ es abeliano, tiene $|\mathbb{Z}_p| = p$ clases de conjugación. Por el teorema 1.14 inciso (ii), tenemos que existen p representaciones irreducibles no isomorfas de $\mathbb{Z}_p$. Además, por el corolario 1.5, todas sus representaciones son de grado $d_\rho = 1$. Como $\mathbb{Z}_p$ es cíclico, basta definir cualquier representación en el generador $1 \in \mathbb{Z}_p$. Para $0 \leq j \leq p-1$ definimos

$$\rho_j : \mathbb{Z}_p \to \mathrm{GL}(1, \mathbb{C}), \text{ dadas por, } \rho_j(1) = e^{(\frac{2\pi i j}{p})}.$$

Es claro que cada ρ_j es una representación irreducible de $\mathbb{Z}_p$, y si k, j $\in \mathbb{Z}_p$ con $k \neq j$, entonces $\rho_k \neq \rho_j$, puesto que $\rho_k(1) \neq \rho_j(1)$. Por tanto,

hemos encontrado todas las representaciones irreducibles de $\mathbb{Z}_p$. Note que

$$\widehat{Q}(\rho_j) = \sum_{h\in\mathbb{Z}_p} Q(h)\rho_j(h) = \frac{1}{2}(e^{(\frac{2\pi ij}{p})} + e^{(\frac{2\pi ij(p-1)}{p})}) = \cos(\frac{2\pi j}{p}). \quad (4.13)$$

Ahora

$$\begin{aligned} ||Q^{*n} - U||_{TV}^2 &\leq \frac{1}{4}\sum_{j=1}^{p-1} d_{\rho_j} Tr(\widehat{Q^{*n}}(\rho_j)\widehat{Q^{*n}}(\rho_j)^*) \\ &= \frac{1}{4}\sum_{j=1}^{p-1} Tr([\widehat{Q}(\rho_j)]^n([\widehat{Q}(\rho_j)]^n)^*) \\ &= \frac{1}{4}\sum_{j=1}^{p-1} Tr(\cos(\frac{2\pi j}{p})^n(\cos(\frac{2\pi j}{p})^n)^*) \\ &= \frac{1}{4}\sum_{j=1}^{p-1} \cos(\frac{2\pi j}{p})^{2n} = \frac{1}{2}\sum_{j=1}^{\frac{p-1}{2}} \cos(\frac{\pi j}{p})^{2n}. \end{aligned} \quad (4.14)$$

La primera desigualdad es aplicar el lema de la cota superior (lema 4.6). La segunda igualdad se da porque $d_{\rho_j} = 1$ para toda j, y por el corolario 1.17. En tanto que la tercera, se da en virtud de la ecuación (4.13). La última igualdad se da por las propiedades del coseno y porque p es impar. De esta forma, solo basta acotar el término

$$\frac{1}{2}\sum_{j=1}^{\frac{p-1}{2}} \cos(\frac{\pi j}{p})^{2n}$$

de la desigualdad (4.14). Definimos $h : [0, \frac{\pi}{2}) \to \mathbb{R}$ dada por

$$h(x) = \ln(e^{\frac{x^2}{2}}\cos(x)).$$

Entonces tenemos que

$$h'(x) = \frac{xe^{\frac{x^2}{2}}\cos(x) - e^{\frac{x^2}{2}}\sin(x)}{e^{\frac{x^2}{2}}\cos(x)} = x - \tan(x);$$

$$h''(x) = 1 - \sec^2(x) < 0 \text{ para } x \in (0, \frac{\pi}{2}).$$

Lo que implica que h es decreciente en $(0, \frac{\pi}{2})$, y en consecuencia

$$\ln(e^{\frac{x^2}{2}}\cos(x)) = h(x) < h(0) = 0 \text{ para } x \in (0, \frac{\pi}{2}).$$

Por tanto,

$$\begin{aligned} \ln(e^{\frac{x^2}{2}}\cos(x)) &< 0 \\ e^{\frac{x^2}{2}}\cos(x) &< 1 \\ \cos(x) &< e^{\frac{-x^2}{2}}, \end{aligned} \tag{4.15}$$

lo cual es válido para $x \in (0, \frac{\pi}{2})$. Observe que

$$\begin{aligned} \frac{1}{2}\sum_{j=1}^{\frac{p-1}{2}}\cos(\frac{\pi j}{p})^{2n} &\leq \frac{1}{2}\sum_{j=1}^{\frac{p-1}{2}}\exp(\frac{-(\frac{\pi j}{p})^2}{2})^{2n} \\ &= \frac{1}{2}\sum_{j=1}^{\frac{p-1}{2}}\exp(\frac{-\pi^2 j^2 n}{p^2}) \\ &< \frac{1}{2}\sum_{j=1}^{\infty}\exp(\frac{-\pi^2 j^2 n}{p^2}) \\ &= \frac{\exp(\frac{-\pi^2 n}{p^2})}{2}\sum_{j=1}^{\infty}\exp(\frac{-\pi^2(j^2-1)n}{p^2}) \\ &< \frac{\exp(\frac{-\pi^2 n}{p^2})}{2}\sum_{j=0}^{\infty}\exp(\frac{-3\pi^2 jn}{p^2}) \\ &= \frac{\exp(\frac{-\pi^2 n}{p^2})}{2(1-\exp(\frac{-3\pi^2 n}{p^2}))}. \end{aligned} \tag{4.16}$$

La primera desigualdad ocurre debido a la desigualdad (4.15). Lo demás son cuentas sencillas. Note que para $n \geq p^2$ obtenemos

$$0 < \frac{1}{2(1-\exp(\frac{-3\pi^2 n}{p^2}))} < 1,$$

y por tanto,

$$\frac{1}{2}\sum_{j=1}^{\frac{p-1}{2}}\cos(\frac{\pi j}{p})^{2n} < \exp(\frac{-\pi^2 n}{p^2}).$$

Finalmente, de la desigualdad (4.14) obtenemos

$$||Q^{*n} - U||_{TV}^2 \leq \frac{1}{2}\sum_{j=1}^{\frac{p-1}{2}} \cos(\frac{\pi j}{p})^{2n} < \exp(\frac{-\pi^2 n}{p^2}),$$

el resultado se sigue calculando raíz cuadrada.

(ii) Sea $k = \frac{p-1}{2}$. Definimos $f : \mathbb{Z}_p \to [-1, 1]$ dada por

$$f(j) = \cos(\frac{2\pi jk}{p})$$

para toda $j \in \mathbb{Z}_p$. El valor esperado de la n-ésima convolución de Q bajo f es

$$Q^{*n}(f) = \sum_{j=0}^{p-1} Q^{*n}(j)\cos(\frac{2\pi jk}{p}). \tag{4.17}$$

Como la caminata aleatoria es simétrica, tenemos para $0 \leq j \leq p-1$, que

$$Q^{*n}(j) = Q^{*n}(-j) = Q^{*n}(p-j). \tag{4.18}$$

Para $k = \frac{p-1}{2}$, tenemos

$$\begin{aligned}
\widehat{Q^{*n}}(\rho_k) &= \sum_{j=0}^{p-1} Q^{*n}(j)\rho_k(j) \\
&= \sum_{j=0}^{p-1} Q^{*n}(j)e^{\frac{2\pi ijk}{p}} \\
&= \sum_{j=0}^{p-1} Q^{*n}(j)\cos(\frac{2\pi jk}{p}) + i\sum_{j=0}^{p-1} Q^{*n}(j)\sin(\frac{2\pi jk}{p}) \\
&= \sum_{j=0}^{p-1} Q^{*n}(j)\cos(\frac{2\pi jk}{p}) \\
&\qquad + i\sum_{j=0}^{\frac{p-1}{2}} \left[Q^{*n}(j)(\sin(\frac{2\pi jk}{p}) + \sin(\frac{2\pi(p-j)k}{p}))\right] \\
&= \sum_{j=0}^{p-1} Q^{*n}(j)\cos(\frac{2\pi jk}{p}).
\end{aligned} \tag{4.19}$$

La cuarta igualdad se debe a la ecuación (4.18). En la quinta igualdad utilizamos el hecho de que la función seno es impar.

De las ecuaciones (4.17) y (4.19) obtenemos

$$Q^{*n}(f) = \widehat{Q^{*n}}(\rho_k).$$

Pero a su vez

$$\begin{aligned}\widehat{Q^{*n}}(\rho_k) &= \left[\widehat{Q}(\rho_k)\right]^n \\ &= \cos(\frac{2\pi k}{p})^n \\ &= \cos(\frac{2\pi(\frac{p-1}{2})}{p})^n \\ &= \cos(\pi - \frac{\pi}{p})^n = (-1)^n[\cos(\frac{\pi}{p})]^n.\end{aligned}$$

La primera igualdad se da por el corolario 1.17. La segunda se debe a la ecuación (4.13). La tercera, a que $k = \frac{p-1}{2}$. Por tanto,

$$Q^{*n}(f) = \widehat{Q^{*n}}(\rho_k) = (-1)^n[\cos(\frac{\pi}{p})]^n.$$

Ahora, el valor esperado de f bajo la distribución uniforme es

$$U(f) = \sum_{j=0}^{p-1} U(j)f(j) = \frac{1}{p}\sum_{j=0}^{p-1}\cos(\frac{2\pi jk}{p}) = \frac{1}{p}\sum_{j=0}^{p-1}\cos(\frac{2\pi j}{p}) = 0.$$

Por la proposición 4.3 obtenemos

$$2||Q^{*n} - U|| \geq |Q^{*n}(f) - U(f)| = |(-1)^n[\cos(\frac{\pi}{p})]^n - 0| = |\cos(\frac{\pi}{p})^n|.$$

Si $x \leq \frac{1}{2}$, entonces

$$\cos(x) \geq \exp(-\frac{x^2}{2} - \frac{x^4}{11}).$$

Si $p \geq 7$, entonces $\frac{\pi}{p} < \frac{1}{2}$ y obtenemos

$$2||Q^{*n} - U|| \geq |\cos(\frac{\pi}{p})^n| \geq \left[\exp(-\frac{(\frac{\pi}{p})^2}{2} - \frac{(\frac{\pi}{p})^4}{11})\right]^n.$$

□

Observación. El hecho de que p sea impar es fundamental en la demostración del teorema 4.8. Si p fuera par, tendríamos, por el lema de la cota inferior, que

$$\begin{aligned} ||Q^{*n} - U||_{TV} &\geq \frac{1}{2|G|} \sum_{j=1}^{p-1} d_{\rho_j} Tr(\widehat{Q^{*n}}(\rho_j)\widehat{Q^{*n}}(\rho_j)^*) \\ &= \frac{1}{2p} \sum_{j=1}^{p-1} \cos(\frac{2\pi j}{p})^{2n} \\ &> \frac{1}{2p} \cos(\frac{2\pi \frac{p}{2}}{p})^{2n} = \frac{1}{2p}. \end{aligned}$$

De modo que, sin importar qué tan grande sea n, siempre sucederá

$$||Q^{*n} - U||_{TV} > \frac{1}{2p}.$$

Definición 4.3. El grupo $\mathbb{Z}_2^d$, es el producto cartesiano de $\mathbb{Z}_2$ consigo mismo d veces. La suma se define entrada a entrada módulo 2. Los elementos de $\mathbb{Z}_2^d$, son vectores binarios. Para $x \in \mathbb{Z}_2^d$, definimos el peso de x, como el número de unos que tiene el vector x, y lo denotamos por $\omega(x)$. Llamamos a $\mathbb{Z}_2^d$, el **cubo d dimensional**.

El siguiente teorema acota la distancia de variación total entre "la caminata aleatoria del vecino mas cercano" y la distribución uniforme en el grupo $\mathbb{Z}_2^d$. La caminata aleatoria del vecino mas cercano en $\mathbb{Z}_2^d$ consta en moverse a uno de los vecinos mas cercanos o no moverse, donde los vecinos mas cercanos de $x \in \mathbb{Z}_2^d$ son los $y \in \mathbb{Z}_2^d$ tales que $||x - y|| = 1$.

Teorema 4.9. *Sea $G = \mathbb{Z}_2^d$, definimos la probabilidad Q en $\mathbb{Z}_2^d$, dada por,*

$$Q(0) = Q(1, 0, \ldots, 0) = \cdots = Q(0, \ldots, 0, 1) = \frac{1}{d+1}$$

y cero en otro caso. Sean $c \in \mathbb{R}_+$ y $k = \frac{1}{4}(d+1)(\ln(d) + c)$. Entonces para $n \in \mathbb{N}$, con $n \geq k$, tenemos

$$||Q^{*n} - U||_{TV}^2 \leq \frac{1}{2}(e^{e^{-c}} - 1).$$

Demostración. Como $\mathbb{Z}_2^d$ es abeliano, tiene $|\mathbb{Z}_2^d| = 2^d$ clases de conjugación. Por el teorema 1.14 inciso (ii), tenemos que existen 2^d representaciones

irreducibles no isomorfas de $\mathbb{Z}_2^d$. Además, por el corolario 1.5, todas sus representaciones son de grado $d_\rho = 1$. Para $x \in \mathbb{Z}_2^d$, definimos $\rho_x : \mathbb{Z}_2^d \to \mathrm{GL}(1, \mathbb{C})$ dadas por

$$\rho_x(y) = (-1)^{\langle x,y \rangle},$$

donde $\langle x, y \rangle$ es el producto interior usual. Sean y, $z \in \mathbb{Z}_2^d$, entonces

$$\rho_x(y+z) = (-1)^{\langle x,y+z \rangle} = (-1)^{\langle x,y \rangle + \langle x,z \rangle} = (-1)^{\langle x,y \rangle}(-1)^{\langle x,z \rangle} = \rho_x(y)\rho_x(z).$$

Lo que implica que cada ρ_x es una representación irreducible de $\mathbb{Z}_2^d$. Sean x y $x_0 \in \mathbb{Z}_2^d$ con $x \neq x_0$. Sin pérdida de generalidad, supongamos que difieren en la j-ésima entrada, lo que implica,

$$\rho_x(e_j) \neq \rho_{x_0}(e_j).$$

Por lo tanto las ρ_x son todas las representaciones irreducibles de $\mathbb{Z}_2^d$. La transformación de Fourier de la probabilidad Q en la representación ρ_x es

$$\begin{aligned} \widehat{Q}(\rho_x) &= \sum_{y \in \mathbb{Z}_2^d} Q(y)\rho_x(y) \\ &= \frac{1}{d+1} + \frac{d-\omega(x)}{d+1} - \frac{\omega(x)}{d+1} \\ &= 1 - \frac{2\omega(x)}{d+1}. \end{aligned} \tag{4.20}$$

Donde el primer término en la segunda igualdad corresponde a $y = 0$. El segundo término corresponde a los $y = e_j$ que difieren de x en la j-ésima entrada, y el tercer término corresponde a los $y = e_j$ que son iguales a x en la j-ésima entrada.

Antes de utilizar el lema de la cota superior veamos dos desigualdades que nos serán de utilidad. Supondremos que d es par, así $\frac{d}{2}$ es entero (el caso impar es análogo y utilizariamos el número $\lfloor \frac{d}{2} \rfloor + 1$, que es entero). Tenemos para $1 \leq j \leq \frac{d}{2}$, que

$$\begin{aligned} \binom{d}{d-(j-1)}(1 - \frac{2(d-(j-1))}{d+1})^{2n} &= \binom{d}{j-1}(1 - \frac{2(d-(j-1))}{d+1})^{2n} \\ &< \binom{d}{j}(\frac{d+1-2d+2j-2)}{d+1})^{2n} \\ &= \binom{d}{j}(\frac{2j-(d+1)}{d+1})^{2n} \\ &= \binom{d}{j}(1 - \frac{2j}{d+1})^{2n}. \end{aligned} \tag{4.21}$$

La desigualdad se da en virtud de que $\binom{d}{j}$ es creciente para $0 \leq j \leq \frac{d}{2}$. La última igualdad se da porque $2n$ es par.

Ahora acotaremos el término $\binom{d}{j}(1-\frac{2j}{d+1})^{2n}$. Sea $1 \leq j \leq \frac{d}{2}$, entonces para $n \geq k = \frac{1}{4}(d+1)(\ln(d)+c)$, tenemos que

$$\begin{aligned}
\binom{d}{j}(1-\frac{2j}{d+1})^{2n} &\leq \frac{d!}{(d-j)!j!}(1-\frac{2j}{d+1})^{(2)\frac{1}{4}(d+1)(\ln(d)+c)} \\
&\leq \frac{d^j}{j!}(1-\frac{2j}{d+1})^{\frac{1}{2}(d+1)(\ln(d)+c)} \\
&= \frac{d^j}{j!}\left[(1-\frac{2j}{d+1})^{d+1}\right]^{\frac{1}{2}(\ln(d)+c)} \\
&\leq \frac{d^j}{j!}\left[e^{-2j}\right]^{\frac{1}{2}(\ln(d)+c)} \\
&= \frac{d^j}{j!}e^{-j(\ln(d)+c)} \\
&= \frac{d^j}{j!}e^{\ln(d^{-j})}e^{-jc} = \frac{e^{-jc}}{j!}. \qquad (4.22)
\end{aligned}$$

En este caso, todas las igualdades y desigualdades son sencillas y utilizan propiedades elementales del logaritmo y la exponencial.

Utilizando el lema de la cota superior, obtenemos

$$\begin{aligned}
||Q^{*n}-U||_{TV}^2 &\leq \frac{1}{4}\sum_{x\neq 0} d_{\rho_x} Tr(\widehat{Q^{*n}}(\rho_x)\widehat{Q^{*n}}(\rho_x)^*) \\
&= \frac{1}{4}\sum_{x\neq 0} Tr\left[[\widehat{Q}(\rho_x)]^n([\widehat{Q}(\rho_x)]^n)^*\right] \\
&= \frac{1}{4}\sum_{x\neq 0} Tr\left[(1-\frac{2\omega(x)}{d+1})^n\overline{[(1-\frac{2\omega(x)}{d+1})^n]^t}\right] \\
&= \frac{1}{4}\sum_{x\neq 0} Tr\left[(1-\frac{2\omega(x)}{d+1})^{2n}\right] \\
&= \frac{1}{4}\sum_{j=1}^{d}\binom{d}{j}(1-\frac{2j}{d+1})^{2n}.
\end{aligned}$$

La segunda igualdad se da porque $d_{\rho_x} = 1$ para todo x, y por el corolario 1.17. La tercera igualdad se da en virtud de la ecuación (4.20). Concluimos

que

$$\frac{1}{4}\sum_{j=1}^{d}\binom{d}{j}(1-\frac{2j}{d+1})^{2n} = \frac{1}{4}\sum_{j=1}^{\frac{d}{2}}\binom{d}{j}(1-\frac{2j}{d+1})^{2n} + \frac{1}{4}\sum_{j=\frac{d}{2}+1}^{d}\binom{d}{j}(1-\frac{2j}{d+1})^{2n}$$

$$< (2)\frac{1}{4}\sum_{j=1}^{\frac{d}{2}}\binom{d}{j}(1-\frac{2j}{d+1})^{2n}$$

$$\leq \frac{1}{2}\sum_{j=1}^{\frac{d}{2}}\frac{e^{-jc}}{j!} \leq \frac{1}{2}\sum_{j=1}^{\infty}\frac{e^{-jc}}{j!} = \frac{1}{2}(e^{e^{-c}}-1).$$

La primera desigualdad es consecuencia de la ecuación (4.21) mientras que la segunda desigualdad se da en virtud de la ecuación (4.22). □

Apéndice A

Tiempos de Paro

Definición A.1. Una variable aleatoria τ que toma valores en $\mathbb{Z}_+ \cup \{\infty\}$ es un **tiempo de paro** del proceso $Z = \{Z_n : n \geq 1\}$ si el evento $[\tau = n]$ queda determinado por $Z_1, \ldots, Z_n$, es decir, sabiendo $Z_1, \ldots, Z_n$, podemos decir si $\tau = n$ o no. Si $P(\tau < \infty) = 1$, entonces diremos que τ es un **tiempo de paro finito**.

Definición A.2. Sea τ un tiempo de paro del proceso $\{Z_n : n \geq 1\}$, y sea

$$Z_n^\tau = Z_{n\wedge\tau} = \begin{cases} Z_n & \text{si } n \leq \tau, \\ Z_\tau & \text{si } n > \tau. \end{cases}$$

Llamamos a $Z^\tau = \{Z_n^\tau : n \geq 1\}$ el **proceso detenido** en τ.

Proposición A.1. *Si τ es un tiempo de paro de la martingala $\{Z_n : n \geq 1\}$, entonces Z^τ también es una martingala.*

Demostración. Puesto que $Z_1, \ldots, Z_{n-1}$ determinan a $Z_{1\wedge\tau}, \ldots, Z_{(n-1)\wedge\tau}$, basta probar que $E[Z_{n\wedge\tau}|Z_1, \ldots, Z_{n-1}] = Z_{(n-1)\wedge\tau}$. Sea

$$I_n = \begin{cases} 1 & \text{si } n \leq \tau, \\ 0 & \text{si } n > \tau \end{cases}$$

entonces

$$Z_{n\wedge\tau} = Z_{(n-1)\wedge\tau} + I_n(Z_n - Z_{n-1}).$$

El resultado se verifica dividiendo en los casos $n \leq \tau$ y $n > \tau$ y evaluando por separado. Ahora,

$$\begin{aligned} E[Z_{n\wedge\tau}|Z_1, \ldots, Z_{n-1}] &= E[Z_{(n-1)\wedge\tau} + I_n(Z_n - Z_{n-1})|Z_1, \ldots, Z_{n-1}] \\ &= Z_{(n-1)\wedge\tau} + I_n E[Z_n - Z_{n-1}|Z_1, \ldots, Z_{n-1}] \\ &= Z_{(n-1)\wedge\tau}. \end{aligned}$$

La segunda igualdad se da debido a que $Z_{(n-1)\wedge\tau}$ e I_n quedan determinados por $Z_1, \ldots, Z_{n-1}$, y la tercera, porque $\{Z_n : n \geq 1\}$ es martingala. □

Observación. Puesto que $Z_{1\wedge\tau} = Z_1$, tenemos que para todo n, $E[Z_{n\wedge\tau}] = E[Z_1]$. Si τ es un tiempo de paro finito, es decir, si $P(\tau < \infty) = 1$, entonces $Z_{n\wedge\tau} \to Z_\tau$ cuando $n \to \infty$ con probabilidad 1. En caso de que $E[Z_{n\wedge\tau}] \to E[Z_\tau]$, cuando $n \to \infty$ tendríamos que $E[Z_\tau] = E[Z_1]$.

Teorema A.2 (Teorema de muestreo opcional)**.** *Si se cumple cualquiera de estas propiedades*

(i) $\{Z_{n\wedge\tau}\}$ *está uniformemente acotado o*

(ii) τ *es acotada o*

(iii) $E[\tau] < \infty$ *y existe* $M < \infty$ *tal que* $E[|Z_{n+1} - Z_n||Z_1, \ldots, Z_n] < M$.

Entonces $E[Z_{n\wedge\tau}] \to E[Z_\tau]$ *cuando* $n \to \infty$*, y por tanto,* $E[Z_\tau] = E[Z_1]$.

Demostración. Ver [6], p.300. □

Corolario A.3 (Ecuación de Wald)**.** *Sean* $X_1, X_2, \ldots$ *variables aleatorias independientes e idénticamente distribuidas con* $E[|X|] < \infty$*, y sea* τ *un tiempo de paro para* $X_1, X_2, \ldots$ *con* $\mu = E[X_i]$ *y* $E[\tau] < \infty$*, entonces*

$$E\left[\sum_{i=1}^{\tau} X_i\right] = E[\tau]E[X].$$

Demostración. Sea $Z_n = \sum_{i=1}^{n}(X_i - \mu)$. Es claro que $\{Z_n\}$ es martingala. Si conocemos los valores $Z_1, \ldots, Z_n$, entonces quedan determinados los valores de $X_1, \ldots, X_n$. Así, τ es tiempo de paro para $\{Z_n\}$. La propiedad (iii) del teorema A.2 se verifica dado que

$$\begin{aligned} E[|Z_{n+1} - Z_n||Z_1, \ldots, Z_n] &= E[|X_{n+1} - \mu||Z_1, \ldots, Z_n] \\ &= E[|X_{n+1} - \mu|] \\ &\leq E[|X|] + |\mu|, \end{aligned}$$

lo que implica que

$$E[Z_\tau] = E[Z_1] = 0.$$

Ahora note que

$$\begin{aligned} E[Z_\tau] &= E\left[\sum_{i=1}^{\tau}(X_i - \mu)\right] \\ &= E\left[\sum_{i=1}^{\tau} X_i - \tau\mu\right] = E\left[\sum_{i=1}^{\tau} X_i\right] - E[\tau]\mu. \end{aligned}$$

La segunda igualdad se da porque $P(\tau < \infty) = 1$. □

Apéndice B

Teoría de Grupos y álgebra lineal

Definición B.1. Un grupo es un conjunto no vacío G dotado de una operación binaria $\bullet$ ($\bullet : G \times G \to G$) que satisface las siguientes propiedades

(i) La operación $\bullet$ es asociativa. Así, para cualesquiera $a, b, c \in G$ se tiene

$$a \bullet (b \bullet c) = (a \bullet b) \bullet c.$$

(ii) Existe un elemento neutro $e \in G$. Así, para cualquier $a \in G$ se tiene

$$a \bullet e = a = e \bullet a.$$

(iii) Para cada elemento $a \in G$ existe un elemento $a' \in G$ tal que

$$a \bullet a' = e = a' \bullet a.$$

Observaciones. *1.* Si la operación $\bullet$ conmuta, es decir, si $a \bullet b = b \bullet a$ para cualesquiera a, $b \in G$, diremos que G es abeliano.

2. Algunas veces denotaremos al neutro e de G por 1 ó 0.

3. Por la importancia de la operación algunas veces escribiremos $G = (G, \bullet)$.

Ejemplo B.1. El conjunto de los números reales con la suma $\mathbb{R} = (\mathbb{R}, +)$ es un grupo abeliano. La suma, por definición es asociativa y conmutativa. El neutro es $e = 0$ y el inverso de a es $-a$.

Ejemplo B.2. El conjunto de números complejos sin el número 0 y con la multiplicación usual, $\mathbb{C}^* = (\mathbb{C} - \{0\}, \cdot)$, es un grupo abeliano. La multiplicación asocia y conmuta en $\mathbb{C}$. El neutro es $e = 1 + i0$. El inverso de $x = a + ib$ es $y = \frac{a}{a^2+b^2} + i\frac{-b}{a^2+b^2}$.

Ejemplo B.3. Sean $n, m \in \mathbb{Z}$, decimos que n divide a m si existe $k \in \mathbb{Z}$ tal que $nk = m$ y lo denotamos por $n|m$. Decimos que a es congruente con b módulo n si $a - b$ es divisible por n, es decir, si $n|a - b$ y lo denotamos por $a \equiv b \pmod{n}$.

La definición anterior es una relación de equivalencia en $\mathbb{Z}$ y sus clases de equivalencia son llamadas clases residuales módulo n. Veamos por ejemplo, que la definición de congruencia es transitiva (la reflexividad y simetría son triviales). Supongamos que $a \equiv b \pmod{n}$ y que $b \equiv c \pmod{n}$ entonces existen k_1, $k_2 \in \mathbb{Z}$ tales que $nk_1 = a - b$ y $nk_2 = b - c$. De modo que $n(k_1 + k_2) = a - c$ y por tanto, $a \equiv c \pmod{n}$.

Definimos

$$\mathbb{Z}_n := \mathbb{Z}/n\mathbb{Z} = \{[a] : a \in \mathbb{Z}\},$$

donde

$$[a] = \{b \in \mathbb{Z} : a \equiv b \pmod{n}\}$$

es la clase de equivalencia de a. Sea $A = \{[0], \ldots, [n-1]\}$. Entonces $A = \mathbb{Z}_n$. Para ver esto, basta probar que $\mathbb{Z}_n \subset A$. Si $[a] \in \mathbb{Z}_n$ entonces, por el algoritmo de la división, existen únicos k, $r \in \mathbb{Z}$ tales que $nk + r = a$, con $0 \leq r < n$. De modo que $a \equiv r \pmod{n}$ y por consiguiente $[a] = [r]$. Así, $[a] \in A$. Es fácil ver que ningún elemento en A se repite de modo que hay n clases residuales. En $\mathbb{Z}_n$, definimos la siguiente operación $[a] + [b] := [a + b]$. También es fácil ver que la operación no depende del representante. Tenemos entonces que $\mathbb{Z}_n$ es un grupo abeliano con esta operación. La asociatividad y conmutatividad se heredan de $\mathbb{Z}$, $[0]$ sirve de neutro y el inverso de $[a]$ es $[-a]$.

Definición B.2. Sea G un grupo, definimos el orden de G como el cardinal asociado a $|G|$. Un grupo es finito si $|G| \in \mathbb{N}$, de otro modo, es infinito.

Definición B.3. Sea $B \subseteq A$ y $f : A \to C$ una función. Definimos la restricción de f a B denotada $f|_B$ como la función $f|_B : B \to C$, con $f|_B(x) = f(x)$.

Definición B.4. Sea V un espacio vectorial sobre F (donde $F = \mathbb{R}$ o $\mathbb{C}$). Un producto interior en V es una función $\langle , \rangle : V \times V \to F$ tal que para todo $x, y, z \in V$ y para todo $c \in F$,

(i) $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$.

(ii) $\langle cx, y\rangle = c\,\langle x, z\rangle$.

(iii) $\langle x, y\rangle = \overline{\langle y, x\rangle}$ (conjugación compleja).

(iv) $\langle x, x\rangle > 0$ si $x \neq 0$.

Glosario de Simbología

$\overline{A}$: Matriz cuya entrada ij es $\overline{A_{ij}}$.

A^t: Matriz cuya entrada ij es A_{ji}.

$A^* = \overline{A^t}$: Matriz cuya entrada ij es $\overline{A_{ji}}$.

$\langle\beta\rangle$: Espacio vectorial generado a partir del conjunto β.

$\mathbb{C}$: Números complejos.

d_ρ: Grado de la representación ρ.

$e_i = (0, \ldots, 1, \ldots, 0)$: El i-ésimo vector canónico.

$|G|$: Cardinalidad del conjunto G.

$\mathrm{GL}(n, F)$: Grupo de matrices invertibles de $n \times n$, sobre F.

$\mathrm{GL}(V)$: Grupo de transformaciones lineales invertibles de V en V.

$H \leq G$, $(H \lneq G)$: H es subgrupo de G (subgrupo propio).

$\mathbb{N}$: Números naturales.

$p_{ij} = P(X_{n+1} = j | X_n = i)$: Probabilidad de transición.

$p_{ij}(m) = P(X_{n+m} = j | X_n = i)$: Probabilidad de transición en m-pasos.

$||P - Q||_{TV}$: Distancia de variacion total ente P y Q.

$P * Q$: Convolución entre P y Q $((P * Q)(s) = \sum_{t \in G} P(st^{-1})Q(t))$.

$Q^{*n} = Q * Q^{*n-1}$: La n-ésima convolución de Q (para $n = 1$, $Q^{*1} = Q$).

$\widehat{Q}(\rho) = \sum_{s \in G} Q(s)\rho(s)$: La transformación de Fourier de Q en ρ.

$\rho : G \to GL(V)$: Representación lineal de G en V.

$\mathbb{R}$: Números reales.

S_n: Grupo de todas las permutaciones en $\{1, 2 \ldots, n\}$.

$\langle u, v \rangle$: Producto interior entre u y v.

$Tr(A)$: Traza de la matriz A.

$||v|| = \sqrt{\langle v, v \rangle}$: Norma de v.

$\lfloor x \rfloor$: Parte entera de x.

$W \leq V$, $(W \lneq V)$: W es subespacio de V (subespacio propio).

$\chi_\rho = \chi$: Carácter de la representación ρ.

$\mathbb{Z}$: Números enteros.

$\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z} = \{[0], \ldots, [n-1]\}$: Grupo de enteros módulo n.

Bibliografía

[1] P. Diaconis: *Group Representations in Probability and Statistics.* Institute of Mathematical Statistics. Lecture Notes-Monograph Series Vol. 11, 1988.

[2] P. Diaconis, L. Saloff-Coste: *Comparison Techniques for Random Walk on Finite Groups.* The Annals of Probability, Vol. 21, No.4, 1993.

[3] W. Feller: *Introducción a la Teoría de Probabilidades y sus Aplicaciones.* Limusa, Tercera Edición, 1973.

[4] G. Grimmett, D. Stirzaker: *Probability and Random Processes.* Oxford University Press, Third Edition, 2001.

[5] P. Hoel, S. Port, C. Stone: *Introduction to Stochastic Processes.* Houghton Mifflin Company, 1972.

[6] S. Ross: *Stochastic Processes.* Wiley Series in Probability and Mathematical Statistics, Second Edition, 1996.

[7] J. Serre: *Linear Representations of Finite Groups.* Springer, G.T.M., Vol.42, Second Edition, 1977.

[8] F. Zaldívar: *Introducción a la Teoria de Grupos.* S.M.M.-Reverté, Vol. 32, 2006.