Teoremas Ergódicos en Espacios Hiperbólicos

Autor: Pablo Lessa
Orientador: Martin Sambarino

5 de noviembre de 2009

Para Marcela.

Agradecimientos

A la barra que cursó Dinámica compleja allá por el 2006, y que ahora están desperdigados por el mundo... ¡Qué nos volvamos a juntar! Al Leva, por dejarme interrumpir con preguntas cuando estaba por salir el ejemplo. Al Mati, por el semestre complejo que nos mandamos. Al Gordo, por las tertulias, en el centro, y en los bares. Al Rata, por el mes y medio de procesos. Al Rata y a Juliana por Chile. Al Dieguito, por siempre tener algo divertido para contar de matemática... ¡Ojalá pinte colaborar en algo, por más que la última vez quedamos anulados! A Marianita, por enseñarme un poco de Álgebra casi sin lágrimas de por medio y por organizar tantas cosas. Al nuevo Alfonso... ¡Defendé esa tesis de una vez loco! A Marcela, esta vez por las charlas sobre física. ¡Ojalá fueran más! Al Chermi, por la buena onda y buena comida. Al Guarino por las fiestas, y las clases de filosofía práctica. Al Frodo, por pensar en matemática en voz alta. Al Sambita, por tratar de sacar adelante el coloquiooleis. A Sambita y Juliana, por hospedar a este viajero enfermo. A Ricardo, por introducirme a la probabilidad. A Mario, por hacerse tiempo para escuchar. A Roberto y Matilde por emprender la tarea de leer y tratar de entender esta novela que escribí. A Martín, por tener la paciencia y la flexibilidad que me permitió estar tan feliz mientras la escribía. A Marcela, por toda la felicidad de estos tres años. A mis padres María Paz y Enrique, y a mi hermana Carolina, por aceptarme y quererme, aunque les salí matemático.

A todos, por la calidez humana que tuve el privilegio de sentir en estos años, y sin la cual toda actividad pierde su sentido.

Introducción

Esta tesis trata de teoremas ergódicos. Entendemos por teorema ergódico aquel que hace una afirmación sobre lo que le pasará casi seguramente a cierto tipo de sucesiones aleatorias.

Históricamente el primer teorema de este tipo es la Ley de Grandes Números, que afirma que para sucesiones de números aleatorios que son independientes e idénticamente distribuidos y además cumplen una condición de acotación (que en este caso es que tengan esperanza finita) casi seguramente la sucesión de promedios parciales convergerá a un número fijo (no aleatorio) que es la esperanza de cualquiera de las variables. Las primeras versiones de este teorema fueron demostradas a principios del siglo XIX, pero las observaciones empíricas que motivan el teorema son mucho más antiguas.

El Teorema Ergódico de Birkhoff, que fué publicado por primera vez en [Bir31], es análogo a la Ley de Grandes Números, con la diferencia de que el tipo de sucesiones aleatorias a las cuales se aplica es más amplio (sucesiones estacionarias de esperanza finita, en lugar de independientes) pero a su vez la conclusión es más débil (casi seguramente los promedios parciales tendrán limite, pero este límite será aleatorio).

Este último teorema tiene importancia aún en el estudio de sucesiones deterministas (y de hecho el teorema tiene sus orígenes en la Mecánica Estadística). El vínculo viene dado de la siguiente manera: Si $f : M \to M$ es una transformación medible de un espacio de medida M que además preserva una medida de probabilidad en M, el teorema de Birkhoff implica que dada una función integrable $\phi : M \to \mathbb{R}$ se cumple que para casi todo punto $x \in M$ los promedios

$$\frac{1}{n}\sum_{k=0}^{n-1}\phi(f^k(x))$$

tienen límite. Esto muestra que el teorema da información sobre casi todas las trayectorias de un sistema determinista.

Los conceptos que hoy se llaman “Exponente de Lyapunov” y “Regularidad Lyapunov” fueron introducidos por Lyapunov en el siglo XIX para el estudio de la persistencia de la estabilidad asintótica de cierta trayectoria de una ecuación diferencial, al alterar la ecuación ligeramente (ver [BP01] y las referencias allí contenidas para un resumen histórico más detallado).

En la década de 1960, empezó a verse que la regularidad Lyapunov, y la existencia de exponentes de Lyapunov no nulos está también vinculada con el comportamiento caótico de los iterados de un difeomorfismo de una variedad compacta.

En este contexto fué que, a mediados de la década del ’60, Oseledets demostró un teorema que garantiza la regularidad Lyapunov para casi todas las órbitas de un difeomorfismo f con respecto a cualquier medida invariante (ver [Ose08]).

El teorema de Oseledets puede verse como una generalización del Teorema de Birkhoff. Es una afirmación sobre lo que le pasará casi seguramente a la sucesión de productos parciales $B_n = A_n \cdots A_1$ de una sucesión aleatoria estacionaria de matrices invertibles $\{A_n\} \subset \mathrm{GL}(N)$ que cumple cierta condición de acotación (análoga a la condición de integrabilidad necesaria en el teorema de Birkhoff).

Notemos que el teorema de Descomposición Polar nos permite escribir $B_n = O_n P_n$ donde O_n es ortogonal y P_n es simétrica y positiva. Una vez hecho esto, puede verse que la regularidad Lyapunov (como se define en el capítulo 6) es una afirmación únicamente sobre la sucesión de matrices positivas $\{P_n\}$.

A fines de los ’80 (e.g. en [Kaĭ89]), Kaimanovich observó que, por lo menos en el caso de matrices de determinante 1, la conclusión del teorema de Oseledets puede obtenerse de la afirmación de que casi seguramente la sucesión $\{P_n\}$ se mantiene cerca de una geodésica en el espacio de de las matrices positivas $\mathrm{Pos}(N)$.

En [KM99] Karlsson y Margulis demostraron un teorema muy general que puede resumirse en lo siguiente: Sea X un espacio métrico completo que cumple ciertas hipótesis de hiperbolicidad. Si $\{\rho_n\}$ es una sucesión aleatoria estacionaria de isometrías de X, que cumple una condición de acotación, entonces dado $x \in X$ casi seguramente la sucesión $(\rho_1 \circ \cdots \circ \rho_n)(x)$ se mantendrá cerca de una geodésica en X.

El objeto de esta tesis es mostrar la conexión entre las distintas versiones del teorema de Oseledets y los teoremas ergódicos análogos al de Karlsson y Margulis (i.e. teoremas que muestran la existencia casi segura de una geodésica cercana a cierta sucesión aleatoria). Surge naturalmente la siguiente pregunta ¿Puede generalizarse el teorema de Karlsson y Margulis a sucesiones aleatorias que no provienen de aplicar isometrías al azar?

Una potencial aplicación de este tipo de generalización sería responder a la siguiente pregunta: Sea $f : M \to M$ un homeomorfismo isotópico a la identidad en una variedad de curvatura seccional no-positiva M y asociemos a cada $x \in M$ la sucesión de puntos en el cubrimiento universal $\tilde{M}$ de M dada por $\{F^n\tilde{x}\}$ donde $\tilde{x}$ es un levantado de x y F es un levantado de f. ¿Existe una geodésica cercana a $\{F^n\tilde{x}\}$ para casi todo $x \in M$?

Una respuesta afirmativa a la anterior pregunta, generalizaría la noción de vector de rotación que actualmente se utiliza para homeomorfismos isotópicos a la identidad en el toro. Notemos que en este contexto no puede aplicarse el teorema de Karlsson y Margulis ya que (salvo en casos excepcionales) no existe una isometría que lleve $\tilde{x}$ a $F\tilde{x}$. Aún en el caso de que el grupo de isometrias actúe transitivamente en la variedad, no es claro como se podría hacer para aplicar dicho teorema.

El teorema 5.14 de esta tesis generaliza el teorema de Karlsson y Margulis a sucesiones aleatorias que no necesariamente provienen de aplicar isometrias al azar. Además, en el mismo capítulo, se da una respuesta afirmativa a la pregunta sobre vectores de rotación, en el caso de superficies hiperbólicas compactas. Si bien ese caso ya era conocido, los métodos del capítulo (i.e. el uso del teorema 5.14 sumado a un lema geométrico) pueden generalizarse a dimensiones mayores.

En el capítulo 4 se le da la estructura hiperbólica al espacio de matrices positivas y simetricas $\mathrm{Pos}(N)$ y en los capítulos 6 y 7 se demuestra el teorema de Oseledets a partir del teorema 5.14.

En los primeros tres capítulos se desarrollan las herramientas necesarias de la teoría ergódica clásica (i.e. la teoría ergódica de sucesiones de números aleatorios) en forma escencialmente autocontenida. De estas herramientas, el Teorema Ergódico de Kingman (ver [Kin68]) y el Lema de Karlsson-Margulis (lema 4.1 en [KM99]) son especialmente relevantes para la demostración del teorema 5.14.

Índice general

Capítulo 1

Ley de Grandes Números

En este capítulo nos adentramos en el tema de los teoremas ergódicos empezando por el más simple, la Ley de Grandes Números.

Empezamos introduciendo la notación y la terminología probablística para los conceptos de teoría de la medida (ver por ejemplo [Ash00]).

Luego damos una demostración de las cotas de Chernoff para variables Bernoulli (una excelente exposición del tema de las desigualdades exponenciales está dada en [Tal96]) y con esto demostramos el teorema para este tipo de variables.

Damos luego una pequeña discusión de equidistribución de una sucesión y utilizamos el concepto para mostrar como se deduce el teorema para variables acotadas a partir del teorema para variables Bernoulli.

El caso general, lo obtenemos a partir de la clásica desigualdad maximal (debida a Garsia, ver [Shi96] o [Mañ83]) que se interpreta como una afirmación sobre la continuidad de los promedios $\frac{1}{n}\sum_{k\leq n} f(x_k)$ cuando f varía poco en distancia L^1.

Hemos ido armando este camino que mezcla las demostraciones clásicas del Teorema Ergódico de Birkhoff con las cotas de Chernoff, más que nada para entender el teorema maximal en el contexto más sencillo posible.

1.1. Variables Aleatorias y Distribuciones

Denotaremos por $(\Omega, \mathcal{F}, \mathbb{P})$ a un espacio de probabilidad, donde $\mathcal{F}$ es la σ-álgebra de conjuntos medibles (que también llamaremos eventos), y $\mathbb{P}$ es la medida de probabilidad.

Una variable aletoria X es una función $X : \Omega \to \mathbb{R}$ que es Borel medible. Si $A \subset \mathbb{R}$ es un conjunto boreliano, tiene asociado vía X un evento que denotaremos de las siguientes formas:

$$\{\omega \in \Omega : X(\omega) \in A\} = \{X \in A\} = X^{-1}(A)$$

Escribiremos la probabilidad de este evento como sigue:

$$\mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}) = \mathbb{P}(X \in A)$$

En general probabilidades escritas en la forma:

$$\mathbb{P}(X \text{ cumple la propiedad } P)$$

deben ser interpretadas como se explicó anteriormente tomando en este caso $A = \{x \in \mathbb{R} : x \text{ cumple la propiedad } P\}$ que debe ser un conjunto boreliano.

Si $A \subset \Omega$ es un evento (equivalente a $A \in \mathcal{F}$ o A medible) tiene asociada la variable aleatoria 1_A que vale 1 en A y 0 en $\Omega \setminus A$. A esta variable la llamaremos la indicatriz de A.

Se dice que una variable aleatoria X tiene esperanza finita si se cumple que:

$\mathbb{E}(|X|) = \int_\Omega |X| \mathrm{d}\mathbb{P} < +\infty$

En este caso se dice que la integral de Lebesgue

$$\mathbb{E}(X) = \int_\Omega X \mathrm{d}\mathbb{P}$$

es la esperanza de X (equivalentemente su valor medio o valor esperado).

La distribución de una variable aleatoria X es la medida μ en los borelianos de $\mathbb{R}$ definida mediante:

$$\mu(A) = \mathbb{P}(X \in A)$$

Puede demostrarse por inducción (en L^1) que para toda función medible $f : \mathbb{R} \to \mathbb{R}$ que es μ-integrable se tiene:

$$\mathbb{E}(f(X)) = \int_\mathbb{R} f(x) \mathrm{d}\mu(x)$$

en particular tomando $f(x) = x$ para todo x se tiene una fórmula para la esperanza en función de la distribución.

Si $X_1, \ldots, X_n$ son variables aleatorias, su distribución conjunta es la medida μ en los borelianos de $\mathbb{R}^n$ que cumple:

$$\mu(A_1 \times \cdots \times A_n) = \mathbb{P}(X_1 \in A_1 \cap \cdots \cap X_n \in A_n) = \mathbb{P}(X_1 \in A_1, \cdots, X_n \in A_n)$$

para toda n-upla de borelianos $A_1, \ldots, A_n \subset \mathbb{R}$.

En este caso si $f : \mathbb{R}^n \to \mathbb{R}$ es μ-integrable se tiene:

$$\mathbb{E}(f(X_1, \ldots, X_n)) = \int_{\mathbb{R}^n} f(x_1, \ldots, x_n) \mathrm{d}\mu(x_1, \ldots, x_n)$$

En el caso de una sucesión de variables aleatorias $X_1, \ldots, X_n, \ldots$ si denotamos por $\mathbb{R}^{\mathbb{N}}$ al conjunto de sucesiónes de reales dotado con la topología de la convergencia puntual se tiene que la función $F : \Omega \to \mathbb{R}^{\mathbb{N}}$ definida mediante:

$$F(\omega)(k) = X_k(\omega) \ \forall k \in \mathbb{N}$$

es borel medible. Se define entonces la distribución de la sucesión $\{X_n\}$, como la medida μ en los borelianos de $\mathbb{R}^{\mathbb{N}}$ que cumple:

$$\mu(A) = \mathbb{P}(F \in A)$$

El conjunto $A \subset \mathbb{R}^{\mathbb{N}}$ formado por las sucesiones que tienen límite es un conjunto boreliano. Por lo tanto dado una sucesión de variables aleatorias $X_1, \ldots, X_n, \ldots$ queda definida la probabilidad:

$$\mathbb{P}(X_n \text{ tiene límite cuando } n \to +\infty) = \mathbb{P}(F \in A)$$

Definimos a partir de $\{X_n\}$ una nueva sucesión de variables aleatorias dadas por las sumas parciales $S_n = X_1 + \cdots + X_n$. Tendrá particular interés para nosotros la siguiente probabilidad:

$$\mathbb{P}(\frac{1}{n} S_n \text{ tiene límite cuando } n \to +\infty)$$

En particular nos interesa la pregunta siguiente: ¿Que condiciones se pueden imponer sobre la distribución conjunta que garanticen que la probabilidad anterior sea 1?

Una primer condición intuitiva (una vez que se entendió lo que trata de modelar este formalismo) es la hipótesis "Todas las X_n tienen la misma distribución". Con esto se intenta modelar lo que sucede si tiramos muchas veces un mismo dado o una misma moneda (el resultado de la tirada n-ésima sería $X_n(\omega)$ para cierto ω desconocido, y la hipótesis es que las probabilidades de cada resultado no dependen de n). Sin embargo esto resulta insuficiente como mostraremos a continuación.

Tomemos $\{x_n\}$ una sucesión de ceros y unos tal que $\frac{1}{n}s_n$ (la definición de s_n es obvia) no tiene límite cuando $n \to +\infty$. Luego tomemos X_1 una variable aleatoria tal que $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = 0) = \frac{1}{2}$ y definamos:

$$X_n = \begin{cases} x_n & \text{si } X_1 = 0 \\ 1 - x_n & \text{si } X_1 = 1 \end{cases}$$

Todas las variables de esta sucesión tienen la misma distribución pero se tiene:

$$\mathbb{P}(\frac{1}{n}S_n \text{ tiene límite cuando } n \to +\infty) = 0$$

Vemos entonces que para tener éxito en demostrar que los promedios convergen con probabilidad 1 se debe tener alguna hipótesis sobre las distribuciónes conjuntas de las variables X_n.

1.2. Independencia

Sean $X_1, \dots, X_n$ variables aleatorias con distribuciónes $\mu_1, \dots, \mu_n$ respectivamente. Sea μ su distribución conjunta. Se dice que las variables son independientes si:

$$\mu = \mu_1 \times \cdots \times \mu_n$$

Es decir si su distribución conjunta es la medida producto de las distribuciones individuales.

Se dice que un conjunto numerable de variables aleatorias $\{X_n\}$ son independientes si cualquier subconjunto finito lo es.

El teorema central de este capítulo el siguiente (la demostración se da en la sección 1.5):

Teorema 1.1 (Ley de Grandes Números)**.** *Si $X_1, \dots, X_n, \dots$ son variables aleatorias de esperanza finita, independientes, e idénticamente distribuidas entonces:*

$$\mathbb{P}(\frac{1}{n}S_n \to E(X_1)) = 1$$

Notemos que dado que todas las variables tienen la misma distribución $E(X_1) = E(X_n)$ para todo n. Además definiendo $Y_n = X_n - E(X_n)$ se tiene $E(Y_n) = 0$ y es claro que alcanza demostrar el teorema solo para variables de esperanza nula.

1.3. Borel-Cantelli

El siguiente clásico lema (que no demostraremos) es una consecuencia casi trivial de la desigualdad $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$

Lema 1.2 (Borel-Cantelli). *Si $A_1, \ldots, A_n, \ldots \in \mathcal{F}$ y $\sum_n \mathbb{P}(A_n) < +\infty$ entonces*

$$\mathbb{P}(\bigcap_m \bigcup_{n \geq m} A_n) = \mathbb{P}(\{\omega \in \Omega : \omega \in A_n \textit{ para infinitos valores de } n\}) = 0$$

El lema muestra que si para una sucesión de variables aleatorias:

$$\sum_n \mathbb{P}(\frac{1}{n}|S_n| > \epsilon) < +\infty$$

entonces

$$\mathbb{P}(\frac{1}{n}|S_n| > \epsilon \text{ para infinitos valores de } n) = 0$$

Si la desigualdad anterior se cumple para numerables $\epsilon > 0$ arbitrariamente pequeños entonces obtenemos:

$$\mathbb{P}(\frac{1}{n}S_n \to 0) = 1$$

Esto muestra que el lema reduce el problema de demostrar el teorema 1.1 a encontrar buenas cotas para $\mathbb{P}(|S_n| > n\epsilon)$.

Esto puede lograrse en algunos casos importantes mediante la siguiente observación:

$$\mathbb{P}(S_n > n\epsilon) = \mathbb{E}(1_{\{S_n - n\epsilon > 0\}}) \leq \mathbb{E}(\exp(S_n t - n\epsilon t)) \text{ para todo } t > 0$$

El método consiste en elegir $t > 0$ para minimizar el lado derecho de la desigualdad. Bajo ciertas hipótesis esta cota es finita y decrece exponencialmente en n.

En particular la desigualdad:

$$\frac{\exp(t)+\exp(-t)}{2} \leq \exp(\frac{t^2}{2})$$

que se verifica tomando series de potencias de ambos lados nos permite demostrar el siguiente teorema:

Teorema 1.3. *Si $X_1, \ldots, X_n$ son variables aleatorias independientes tales que para todo n se tiene $\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = \frac{1}{2}$ entonces para todo $\epsilon > 0$ se cumple que:*

$$\mathbb{P}(\frac{1}{n}|S_n| > \epsilon) \leq 2\exp(-n\frac{\epsilon^2}{2})$$

Demostración: Por las observaciones anteriores tenemos la desigualdad

$$P(\frac{1}{n}S_n > \epsilon) \leq \exp(-n\epsilon t)\mathbb{E}(\exp(S_n t)) \text{ para todo } t > 0$$

Para variables independientes la esperanza del producto es el producto de las esperanzas (esto es un caso particular del teorema de Fubini). Como en nuestro caso además tienen todas la misma distribución se obtiene:

$$\mathbb{E}(\exp(S_n t)) = \mathbb{E}(\exp(X_1 t))^n = (\frac{\exp(t)+\exp(-t)}{2})^n \leq \exp(n\frac{t^2}{2})$$

Substituyendo en lo anterior y tomando $t = \epsilon$ se obtiene:

$$\mathbb{P}(\frac{1}{n}S_n > \epsilon) \leq \exp(-n\frac{\epsilon^2}{2})$$

La desigualdad para $\mathbb{P}(\frac{1}{n}S_n < -\epsilon)$ se obiene por el mismo método, y con esto el resultado deseado. □

Con esto estamos en condiciones de demostrar el siguiente teorema:

Teorema 1.4. *Si $X_1, \ldots, X_n, \ldots$ son variables aleatorias independientes tales que $\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = 0) = \frac{1}{2}$ para todo n, entonces:*

$$\mathbb{P}(\frac{1}{n}S_n \to \frac{1}{2}) = 1$$

Demostración: Definamos $Y_n = 2(X_n - \frac{1}{2})$ el teorema se reduce a demostrar que $\frac{1}{n}(Y_1 + \cdots + Y_n) = \frac{1}{n}\tilde{S}_n \to 0$ cuando $n \to +\infty$ en un conjunto de probabilidad 1.

Como la sucesión $\{Y_n\}$ está en hipótesis del teorema anterior se tiene que:

$$\sum_n \mathbb{P}(\frac{1}{n}|\tilde{S}_n| > \epsilon) < +\infty$$

para todo $\epsilon > 0$. Por lo tanto por el lema de Borel-Cantelli se obtiene el resultado. $\square$

Con los mismos métodos obtenemos el siguiente teorema:

Teorema 1.5 (Ley de Grandes Números para variables Bernoulli)**.** *Si* $X_1, \ldots, X_n, \ldots$ *son variables aleatorias indpendientes tales que* $P(X_n = 1) = p, P(X_n = 0) = 1 - p$ *para todo* n*, entonces:*

$$\mathbb{P}(\frac{1}{n}S_n \to p) = 1$$

Demostración: Si se cumple el teorema para la sucesión de variables $\{1 - X_n\}$ entonces se cumple para $\{X_n\}$ por lo tanto podemos asumir $p > \frac{1}{2}$ o lo que es lo mismo $p > 1 - p$. Otra simplificación sencilla se obtiene definiendo $Y_n = \frac{1}{1-p}(X_n - p)$ para lograr $\mathbb{E}(Y_n) = 0$. Alcanza demostrar

$$\mathbb{P}(\frac{1}{n}\tilde{S}_n \to 0) = 1 \text{ donde } \tilde{S}_n = Y_1 + \cdots + Y_n$$

La única dificultad técnica para repetir los argumentos ya utilizados está en acotar:

$$\mathbb{E}(\exp(-Y_1 t)) \leq \mathbb{E}(\exp(Y_1 t)) = p\exp(t) + (1-p)\exp(\frac{-p}{1-p}t)$$

para $t > 0$.

Tomando series de potencias y descartando los términos negativos se obtiene:

$$\mathbb{E}(\exp(Y_1 t)) \leq 1 + \sum_{n \geq 1}(p + (1-p)\frac{p^{2n}}{(1-p)^{2n}})\frac{t^{2n}}{(2n)!}$$

Por último las desigualdades groseras $p(1-p)^{2n} + (1-p)p^{2n} \leq 2$ y $\frac{1}{(2n)!} \leq \frac{1}{2 \cdot n!}$ nos dan:

$$\mathbb{E}(\exp(Y_1 t)) \leq 1 + \sum_{n \geq 1} \frac{(\frac{t}{1-p})^{2n}}{n!} = \exp(\frac{t^2}{(1-p)^2})$$

Con esta cota y el método ya utilizado (esta vez con $t = \frac{(1-p)^2\epsilon}{2}$) se obtiene:

$$\mathbb{P}(\frac{1}{n}|\tilde{S}_n| > \epsilon) \leq 2\exp(-n\frac{(1-p)^2\epsilon^2}{4})$$

Con lo cual por el lema de Borel-Cantelli se obtiene el resultado. □

1.4. Equidistribución

Definición 1.6 (Sucesión Equidistribuida)**.** Dada una medida de probabilidad μ definida en los borelianos de $\mathbb{R}$ y una familia numerable de conjuntos borelianos $A_n \subset \mathbb{R}$, decimos que una sucesión $\{x_n\} \subset \mathbb{R}$ es equidistribuida si se cumple que:

$$\lim_n \frac{1}{n}\sum_{k\leq n} 1_{A_i}(x_k) \to \mu(A_i) \text{ para todo } i \in \mathbb{N}$$

Cuando la medida μ es la de Lebesgue en el intervalo $[0,1]$ y la familia A_i consiste en los intervalos de extremos racionales la definición anterior coincide con la noción pertinente en el Teorema de Equidistribución de Weil[1].

En nuestro caso cobra interés gracias al siguiente lema:

Lema 1.7 (Equidistribución de las Sucesiones Aleatorias)**.** *Si $X_1, \ldots, X_n, \ldots$ es una sucesión de variables independientes de distribución μ entonces para toda familia numerable de borelianos $\{A_i\}_{i\in I}$ se tiene que:*

$$\mathbb{P}(X_n \textit{ equidistribuida respecto a } \mu \textit{ y } \{A_i\}) = 1$$

Demostración: Para cada A_i las variables $Y_n = 1_{A_i} \circ X_n$ son independientes, de igual distribución, y toman solo dos valores, por lo tanto por el teorema 1.5 se tiene que:

$$\mathbb{P}(\frac{1}{n}\sum_{k\leq n} Y_k \to \mathbb{E}(Y_1)) = 1$$

Intersectando los numerables conjuntos de probabilidad 1 que se obtienen para cada i se deduce el resultado. □

Esto ya es suficiente para demostrar la ley de grandes números en el caso de variables acotadas. La clave es el siguiente lema:

[1]En una versión simple este teorema muestra que $\{n\alpha \mod 1\}_{n\in\mathbb{N}}$ es equidistribuida si α es irracional. Una discusión detallada puede encontrarse en [Fur81]

Lema 1.8. *Sea μ una medida de probabilidad boreliana en $\mathbb{R}$ y $f : \mathbb{R} \to \mathbb{R}$ una función medible acotada. Si $\{I_i\}_{i\in\mathbb{R}}$ es la familia de todos los intervalos de extremos racionales y $\{x_n\} \subset \mathbb{R}$ es una sucesión equidistribuida respecto a μ y $\{A_i = f^{-1}(I_i)\}$ entonces:*

$$\frac{1}{n}\sum_{k\leq n} f(x_n) \to \int_{\mathbb{R}} f(x)\mathrm{d}\mu(x)$$

Demostración: Definamos para cada $N \in \mathbb{N}$ las siguientes versiones "redondeadas" de f:

$$e = \sum_{k\in\mathbb{Z}} \frac{k}{N} 1_{\{f^{-1}([\frac{k}{N},\frac{k+1}{N}))\}}$$

y

$$g = \sum_{k\in\mathbb{Z}} \frac{k+1}{N} 1_{\{f^{-1}([\frac{k}{N},\frac{k+1}{N}))\}}$$

Se tiene que $e \leq f \leq g$ y por lo tanto:

$$\frac{1}{n}\sum_{k\leq n} e(x_k) \leq \frac{1}{n}\sum_{k\leq n} f(x_k) \leq \frac{1}{n}\sum_{k\leq n} g(x_k)$$

Dado que f es acotada se tiene que e y g son simples (combinaciones lineales finitas de funciones indicatrices). La equidistribución de x_n implica que el lado izquierdo tiende a $\int_{\mathbb{R}} e(x)\mathrm{d}\mu(x)$ y el lado derecho a $\int_{\mathbb{R}} g(x)\mathrm{d}\mu(x)$ cuando $n \to +\infty$. Como esto se cumple para todo N se obtiene el lema, dado que $\int_{\mathbb{R}} g(x) - e(x)\mathrm{d}\mu(x) \to 0$ cuando $N \to +\infty$. □

Es clave en la demostración que f sea acotada. Pongamos por ejemplo $f(x) = x$ y μ la distribución normal estandard, μ cumple:

$$\mu([a,b]) = \frac{1}{\sqrt{2\pi}}\int_a^b \exp(-\frac{1}{2}x^2)\mathrm{d}x \ \forall a < b$$

Si x_n es una sucesión equidistribuida respecto a los intervalos de extremos racionales es posible cambiar los valores de $x_{n!}$ (cosa que no afecta la equidistribución, ya que $n!$ tiene densidad cero en $\mathbb{N}$) poniendo valores positivos enormes y lograr que:

$$\lim_n \frac{1}{n}\sum_{k\leq n} x_k = +\infty \neq \int_{\mathbb{R}} x\mathrm{d}\mu(x) = 0$$

Concluimos esta sección con una versión de la Ley de Grandes Números para variables acotadas.

Teorema 1.9 (Ley de Grandes Números para variables acotadas). *Si $X_1, \ldots, X_n, \ldots$ es una sucesión de variables aleatorias acotadas, independientes, e idénticamente distribuidas, de esperanza finita entonces:*

$$\mathbb{P}(\frac{1}{n}S_n \to E(X_1) \text{ cuando } n \to +\infty) = 1$$

donde $S_n = X_1 + \cdots + X_n$.

Demostración: Sea μ la distribución de X_1 y tomemos $f(x) = x$ en un intervalo acotado que contiene a $X_1(\Omega)$ y $f(x) = 0$ en el complemento de este intervalo. Existe un conjunto de probabilidad 1 en el que $f(X_n) = X_n$ para todo n. Intersectando este conjunto con el conjunto en el cual la sucesión es equidistribuida para la familia $\{A_i\}$ de intervalos de extremos racionales en $\mathbb{R}$ (obtenido en el lema 1.7) y aplicando el lema 1.8 se obtiene el resultado. $\square$

1.5. Desigualdad Máximal

Hasta ahora hemos demostrado el siguiente resultado:

Teorema 1.10. *Si $X_1, \ldots, X_n, \ldots$ es una sucesión de variables aleatorias, independientes e idénticamente distribuidas, y $f : \mathbb{R} \to \mathbb{R}$ es una función medible y acotada se tiene que:*

$$\mathbb{P}(\frac{1}{n}\tilde{S}_n \to E(f(X_1)) = 1 \text{ donde } \tilde{S}_n = f(X_1) + \cdots + f(X_n)$$

Para demostrar el teorema 1.1 la idea natural de tomar una sucesión cualquiera de variables de esperanza finita y truncarla en cotas cada vez mayores funciona. Pero requiere el siguiente teorema que muestra la continuidad de los promedios en función de la norma L^1.

Teorema 1.11 (Desigualdad Maximal). *Sean $X_1, \ldots, X_n, \ldots$ variables aleatorias independientes de distribución μ. Y sean $f, g : \mathbb{R} \to \mathbb{R}$ dos funciones medibles y μ-integrables. Para todo $\epsilon > 0$ se cumple:*

$$\mathbb{P}(\sup_n\{\frac{1}{n}\sum_{k\leq n}|f(X_k) - g(X_k)|\} > \epsilon) \leq \frac{1}{\epsilon}\mathbb{E}(|f(X_1) - g(X_1)|)$$

Demostración: Sea $h(x) = |f - g|(x)$ para todo x. Y, fijado $\epsilon > 0$, definamos las variables:

$$M_{m,n} = \text{máx}\{\sum_{k=m}^{l}(h(X_k) - \epsilon) : m \leq l \leq n\}$$

.

Dado que los eventos $\{M_{1,n} > 0\}$ son crecientes en n se tiene que:

$$\mathbb{P}(\sup_n\{\frac{1}{n}\sum_{k\leq n}|f(X_k) - g(X_k)|\} > \epsilon) = \lim_n \mathbb{P}(M_{1,n} > 0)$$

Por otro lado:

$$M_{1,n} = h(X_1) - \epsilon + M_{2,n} \leq h(X_1) - \epsilon + M_{2,n+1}$$

multiplicando por la indicatriz de $A = \{M_{1,n} > 0\}$ y tomando esperanzas se obtiene:

$$\begin{aligned} 0 \leq \mathbb{E}(M_{1,n}1_A - M_{2,n+1}1_A) &\leq \mathbb{E}(h(X_1)1_A) - \epsilon\mathbb{P}(M_{1,n} > 0) \\ &\leq \mathbb{E}(h(X_1)) - \epsilon\mathbb{P}(M_{1,n} > 0) \end{aligned}$$

donde la primer desigualdad es cierta porque las variables $M_{1,n}$ y $M_{2,n+1}$ tienen la misma distribución y estamos tomando valor esperado en el conjunto donde la primera de ellas es positiva.

Por lo tanto para todo n se tiene:

$$\epsilon\mathbb{P}(M_{1,n} > 0) \leq \mathbb{E}(h(X_1))$$

de lo cual tomando límite se deduce el enunciado. □

Con esto estamos en condiciones de completar nuestra demostración de la Ley de Grandes Números.

Demostración de la ley de Grandes Números: Sean $X_1, \ldots, X_n, \ldots$ variables aleatorias de esperanza finita, independientes e idénticamente distribuidas con distribución μ. Consideremos en $L^1(\mathbb{R}, \mathcal{B}, \mu)$ (i.e. el conjunto de funciones $f : \mathbb{R} \to \mathbb{R}$, medibles y μ-integrables) el siguiente subconjunto:

$$\mathcal{F} = \{f \in L^1(\mathbb{R}, \mathcal{B}, \mu) : \mathbb{P}(\frac{1}{n}\sum_{k\leq n} f(X_k) \to \mathbb{E}(f(X_1))) = 1\}$$

Nuestro objetivo es demostrar que $\mathcal{F} = L^1(\mathbb{R}, \mathcal{B}, \mu)$.

Evidentemente $\mathcal{F}$ es un subespacio vectorial. Además las funciones constantes están incluidas en $\mathcal{F}$ trivialmente, mientras que las funciones indicatrices lo están gracias a la demostración de la Ley de Grandes Números para variables Bernoulli que dimos anteriormente.

Por último, consideremos una sucesión $\{f_M\}_{M\in\mathbb{N}} \subset \mathcal{F}$ de funciones positivas que crece a un límite $f \in L^1(\mathbb{R}, \mathcal{B}, \mu)$ cuando $M \to +\infty$. Alcanza con mostrar que $f \in \mathcal{F}$ para terminar la demostración.

Para cada $\epsilon > 0$, a partir de cierto M se cumple $|\mathbb{E}(f(X_1) - f_M(X_1))| < \frac{\epsilon}{2}$ y por lo tanto, por la desigualdad maximal, se obtiene lo siguiente:

$$\mathbb{P}(\limsup_n |\frac{1}{n}\sum_{k\leq n} f(X_k) - \mathbb{E}(f(X_1))| > \epsilon) \leq$$

$$\mathbb{P}(\limsup_n |\frac{1}{n}\sum_{k\leq n} f(X_k) - \mathbb{E}(f_M(X_1))| > \frac{\epsilon}{2}) \leq$$

$$\mathbb{P}(\limsup_n |\frac{1}{n}\sum_{k\leq n} f(X_k) - \frac{1}{n}\sum_{k\leq n} f_M(X_k)| > \frac{\epsilon}{2}) \leq$$

$$\mathbb{P}(\sup_n\{\frac{1}{n}\sum_{k\leq n} |f(X_k) - f_M(X_k)|\} > \frac{\epsilon}{2}) \leq \frac{2}{\epsilon}\mathbb{E}(|f(X_1) - f_M(X_1)|)$$

Como el lado derecho tiende a cero cuando $M \to +\infty$ obtuvimos para todo $\epsilon > 0$

$$\mathbb{P}(\limsup_n |\frac{1}{n}\sum_{k\leq n} f(X_k) - \mathbb{E}(f(X_1))| > \epsilon) = 0$$

lo cual muestra que f pertenece a $\mathcal{F}$, como se buscaba. $\square$

Capítulo 2

Teorema de Birkhoff

En este capítulo demostramos el Teorema Ergódico de Birkhoff (la demostración clásica está bien explicada en [Shi96] y [Mañ83]) y armamos un pequeño diccionario entre los lenguajes probabilísticos y dinámicos.

Para discutir el concepto de ergodicidad, damos la demostración clásica de la Ley Cero-Uno de Kolmogorov (ver [Ash00]).

Luego para mostrar el caso ergódico del teorema de Birkhoff, utilizamos una versión aditiva del lema de A.Karlsson y G.Margulis (ver [KM99]).

El lema es equivalente en esta versión al clásico teorema ergódico maximal de Garsia, pero la demostración es distinta, y tiene la virtud de que puede adaptarse para dar cotas inferiores a procesos subaditivos (la demostración del teorema maximal se extiende a procesos subaditivos pero sólo da cotas superiores).

A.Karlsson y G.Margulis utilizaron el lema para dar una demostración del caso ergódico del teorema subaditivo de Kingman, nosotros adaptamos esta demostración al Teorema de Birkhoff y queda especialmente sencilla.

Luego introducimos la esperanza condicional, y la utilizamos para extender la misma demostración al caso no ergódico.

2.1. Procesos Estacionarios

Mientras que la motivación para demostrar la Ley de Grandes Números viene de consideraciones puramente probabilísticas, la preocupación principal de este capítulo será demostrar un teorema que tiene sus orígenes en consideraciones dinámicas.

Sin embargo, la matemática detrás de ambos resultados es básicamente la misma, y existe un diccionario completo entre el lenguaje dinámico y el lenguaje probabilístico que intentaremos presentar en parte aquí.

Empecemos con las definiciones:

Definición 2.1 (Transformación que preserva medida, sistema dinámico que preserva probabilidad)**.** Sea $(X, \mathcal{X}, \mu)$ un espacio de probabilidad. Se dice que $T : X \to X$ preserva μ, si es medible y se cumple $\mu(T^{-1}A) = \mu(A)$ para todo $A \in \mathcal{X}$.

Un sistema dinámico que preserva probabilidad es una cuaterna $(X, \mathcal{X}, \mu, T)$ donde $(X, \mathcal{X}, \mu)$ es un espacio de probabilidad y $T : X \to X$ preserva μ.

Este es el punto de partida de la teoría ergódica. Puede demostrarse que si X es un espacio topológico compacto, $\mathcal{X}$ su σ-álgebra de Borel, y $T : X \to X$ una función continua, entonces existe alguna medida μ de probabilidad que es preservada por T.

Nuestro interés en este capítulo es demostrar el siguiente teorema:

Teorema 2.2 (Teorema Ergódico de Birkhoff)**.** *Si $(X, \mathcal{X}, \mu, T)$ es un sistema dinámico que preserva probabilidad, y $f \in L^1(X, \mathcal{X}, \mu)$ entonces:*

$$\mu(\{x \in X : \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k x \textit{ tiene límite cuando } n \to +\infty\}) = 1$$

La similitud de este teorema con la Ley de Grandes Números (Teorema 1.1) es evidente. Se hará aún más explíta con las siguiente definiciones.

Definición 2.3 (Proceso Estacionario)**.** Un proceso estacionario (o sucesión estacionaria de variables aleatorias) es una sucesión $X_1, \ldots, X_n, \ldots$ de variables aleatorias que cumplen que para todo $k, n \in \mathbb{N}$ la distribución conjunta de $X_{k+1}, \ldots, X_{k+n}$ es igual a la distribución conjunta de $X_1, \ldots, X_n$.

Los dos ejemplos que nos interesan son los siguientes:

- Si $X_1, \ldots, X_n, \ldots$ son variables aleatorias independientes e idénticamente distribuidas, entonces forman un proceso estacionario.
- Si $(X, \mathcal{X}, \mu, T)$ es un sistema dinámico que preserva probabilidad y $f : X \to \mathbb{R}$ es medible, la sucesión de variables aleatorias definida mediante $X_n = f \circ T^{n-1}$ es estacionaria.

Hemos mostrado entonces que el siguiente enunciado es, en principio, más general que el teorema ergódico de Birkhoff como lo enunciamos antes:

Teorema 2.4 (Teorema Ergódico de Birkhoff)**.** *Si $X_1, \ldots, X_n, \ldots$ es una sucesión estacionaria de variables aleatorias de esperanza finita entonces:*

$$\mathbb{P}(\frac{1}{n}S_n \textit{ tiene límite cuando } n \to +\infty) = 1$$

Mostraremos ahora que ambos enunciados son equiventes. Para esto fijemos de aquí en adelante:

$$\mathbb{R}^{\mathbb{N}} = \{x : \mathbb{N} \to \mathbb{R}\}$$

el conjunto de sucesiones de números reales (convengamos que $0 \notin \mathbb{N}$ de aquí en más).

En $\mathbb{R}^{\mathbb{N}}$ hay una topología natural que es la topología producto (en la cual $x_n \to x$ si y sólo si $x_n(k) \to x(k)$ para todo $k \in \mathbb{N}$). Sea $\mathcal{B}$ la σ-álgebra de Borel asociada a esta topología.

Por último consideremos la transformación "Shift", $T : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ dada por:

$$(Tx)(k) = x(k+1) \; \forall k \in \mathbb{N}$$

Existe una gran variedad de medidas de probabilidad μ tales que $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}, \mu, T)$ es un sistema dinámico que preserva probabilidad, por ejemplo:

- Una medida concentrada en una única sucesión constante (i.e. $x(k) = x(1) \; \forall k \in \mathbb{N}$).
- Una medida concentrada en la órbita bajo el shift de una sucesión periódica. Es decir, dados $x \in \mathbb{R}^{\mathbb{N}}$ y $p \in \mathbb{N}$ tales que $x(k) = x(k+p) \; \forall k \in \mathbb{N}$, consideramos una medida que asigna igual probabilidad a cada uno de los puntos $x, Tx, T^2x, \ldots, T^{p-1}x$.
- Una medida producto de un único factor (i.e. μ es el producto de numerables copias de cierta medida Boreliana ν en los reales)
- Combinaciones lineales convexas de medidas invariantes, como las anteriores.

De hecho mostraremos que cualquiera de las dos versiones enunciadas del Teorema de Birkhoff, es equivalente a demostrar que el teorema se cumple para caso particular $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}, \mu, T)$ con $f(x) = x(1)$ y μ cualquier probabilidad invariante (la dificultad está en gran diversidad de posibilidades para μ).

Esto se hace de la siguiente manera, sea $X_1, \ldots, X_n, \ldots$ un proceso estacionario proveniente de cierto espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$. La función $F : \Omega \to \mathbb{R}^{\mathbb{N}}$ dada por:

$$(F(\omega))(k) = X_k(\omega)\ \forall k \in \mathbb{N}$$

es Borel medible (porque cada X_k es medible) y por lo tanto se puede definir una probabilidad μ (llamada la distribución del proceso) dada por:

$$\mu(A) = \mathbb{P}(\{\omega \in \Omega : F(\omega) \in A\})$$

La definición de proceso estacionario es equivalente a decir que $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}, \mu, T)$ es un sistema dinámico que preserva probabilidad (i.e. $X_1, \ldots, X_n, \ldots$ es estacionaria si y sólo si su distribución es invariante bajo el shift) . Además se tiene por definición:

$$\mathbb{P}(\frac{1}{n}S_n \text{ tiene límite cuando } n \to +\infty) = \mu(A)$$

donde

$$A = \{x \in \mathbb{R}^{\mathbb{N}} : \exists \lim_n \frac{1}{n}\sum_{k=1}^{n} x(k)\} = \{x \in \mathbb{R}^{\mathbb{N}} : \exists \lim_n \frac{1}{n}\sum_{k=0}^{n-1} f(T^k x)\}$$

Hacemos notar que la construcción anterior, muestra que la Ley de Grandes Números es un caso particular del Teorema Ergódico de Birkhoff, de hecho se deduce del caso $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}, \mu, T)$ donde μ es el producto de numerables copias de cierta probabilidad ν definida en los borelianos de $\mathbb{R}$.

2.2. La Ley Cero-Uno de Kolmogorov

Dedicaremos ahora una sección a exponer una propiedad de las variables independientes que ayuda a entender el Teorema Ergódico de Birkhoff y las dificultades adicionales que ofrece con respecto a la Ley de Grandes Números.

Un hecho poco intuitivo en principio es la existencia de funciones de infinitas variables que no dependen de ninguna cantidad finita de sus variables.

Por ejemplo, la función que le asocia el límite a cada sucesión que tiene límite, o simplemente la función indicatriz de las sucesiones que poseen límite.

Análogamente existen subconjuntos borelianos de $\mathbb{R}^{\mathbb{N}}$ tales que pertenecer a ellos no depende de ninguna cantidad finita de coordenadas. Son los llamados "eventos de cola":

Definición 2.5 (Evento de cola)**.** Sean $\mathbb{R}^{\mathbb{N}}, \mathcal{B}$ y T las sucesiónes de reales, su σ-álgebra de Borel, y la transformación shift. Y definamos

$$\mathcal{B}_n = T^{-n}\mathcal{B}$$

Un evento $A \in \mathcal{B}$ es un evento de cola si

$$A \in \mathcal{B}_n \ \forall n \in \mathbb{N}$$

Los eventos de cola forman una σ-álgebra que llamaremos la σ-álgebra de Cola y denotamos por

$$\mathcal{B}_\infty = \cap_n \mathcal{B}_n$$

El típico ejemplo de evento de cola es, el ya mencionado, conjunto de sucesiones que tienen límite.

El siguiente teorema debido a Kolmogorov muestra que la σ-álgebra de cola es casi trivial desde el punto de vista de las sucesiones de variables aleatorias independientes.

Teorema 2.6 (Ley Cero-Uno de Kolmogorov)**.** *Sean $X_1, \ldots, X_n, \ldots$ variables aleatorias independientes, y sea $A \in \mathcal{B}_\infty$ entonces:*

$$\mathbb{P}(\{X_n\}_{n\in\mathbb{N}} \in A\}) = 0 \ \textit{o} \ \mathbb{P}(\{X_n\}_{n\in\mathbb{N}} \in A\}) = 1$$

Demostración: Sea μ la distribución del proceso $\{X_n\}$ (en particular $\mathbb{P}(\{X_n\}_{n\in\mathbb{N}} \in A) = \mu(A)$ y μ es un producto numerable de ciertas medidas de probabilidad borelianes reales) y definamos la familia de conjuntos independientes de A con esta medida:

$$\mathcal{B}_A = \{B \in \mathcal{B} : \mu(A \cap B) = \mu(A)\mu(B)\}$$

mostraremos que $A \in \mathcal{B}_A$ de modo que $\mu(A) = \mu(A)^2$ y se obtiene el resultado.

Para esto verificamos que:

- Para cualquier conjunto $B \in \mathcal{B}$ la familia de conjuntos independientes de B es cerrada bajo uniones crecientes numerables e intersecciones decrecientes numerables (i.e. $\mathcal{B}_B$ es una clase monótona y por lo tanto si mostramos que contiene un álgebra de conjuntos contiene la σ-álgebra generada).

- Para cualquier $B \in \mathcal{B}$ que es de la forma $B = \{x \in \mathbb{R}^{\mathbb{N}} : x(1) \in B_1, \ldots, x(n) \in B_n\}$ para cierto $n \in \mathbb{N}$ y ciertos borelianos $B_n \subset \mathbb{R}$ se tiene que $\mathcal{B}_{n+1} \subset \mathcal{B}_B$. Esto se cumple porque en este caso la clase monótona $\mathcal{B}_B$ contiene un álgebra generadora de la σ-álgebra $\mathcal{B}_{n+1}$ porque μ es una medida producto.

- Como $A \in \mathcal{B}_n$ para todo $n \in \mathbb{N}$ se deduce de lo anterior que $\mathcal{B}_A$ contiene un generador de $\mathcal{B}$ y por lo tanto $\mathcal{B}_A = \mathcal{B}$. En particular $A \in \mathcal{B}_A$.

□

El resultado anterior permite demostrar que en el caso de variables independientes se tiene:

$$\mathbb{P}(\frac{1}{n}S_n \text{ tiene límite cuando } n \to +\infty) = 0 \text{ o } 1$$

Otro resultado interesante es que en caso de existir el límite debería ser constante (notemos que no asumimos que las variables tenían la misma distribución y por lo tanto este resultado no se deduce de la Ley de Grandes Números). Formalizamos esto con la siguiente definición y corolario:

Definición 2.7 (Función de Cola)**.** $f : \mathbb{R}^{\mathbb{N}} \to \{-\infty\} \cup \mathbb{R} \cup \{+\infty\}$ es una función de cola si es medible respecto a la σ-álgebra $\mathcal{B}_\infty$ es decir si cumple:

$$f^{-1}(B) \in \mathcal{B}_\infty$$

para todo boreliano $B \subset \{-\infty\} \cup \mathbb{R} \cup \{+\infty\}$.

Corolario 2.8. *Si $X_1, \ldots, X_n, \ldots$ es una sucesión de variables aleatorias independientes y $f : \mathbb{R}^{\mathbb{N}} \to \{-\infty\} \cup \mathbb{R} \cup \{+\infty\}$ es una función de cola entonces existe $a \in \{-\infty\} \cup \mathbb{R} \cup \{+\infty\}$ tal que:*

$$\mathbb{P}(f(X_1, X_2, \ldots) = a) = 1$$

En particular el corolario muestra que en el caso independiente las variables $\limsup_n \frac{1}{n}S_n$ y $\liminf_n \frac{1}{n}S_n$ son constantes en un conjunto de probabilidad 1.

2.3. Ergodicidad

Una dificultad del Teorema Ergódico de Birkhoff es que existen procesos estacionarios tales que $\lim_n \frac{1}{n} S_n$ no es constante.

Para ver un ejemplo de esto consideremos una variable X_1 tal que $\mathbb{P}(X_1 = 0) = \mathbb{P}(X_1 = 1) = \frac{1}{2}$ y tomemos:

$$X_n = X_1 \ \forall n \in \mathbb{N}$$

El proceso así definido es estacionario pero:

$$\mathbb{P}(\lim_n \frac{1}{n} S_n = 0) = \mathbb{P}(\lim_n \frac{1}{n} S_n = 1) = \frac{1}{2}$$

Desde el punto de vista dinámico el fenómeno se entiende de la siguiente manera. Supongamos que $(X, \mathcal{X}, \mu, T)$ es un sistema dinámico que preserva probabilidad y que existe un conjunto $A \in \mathcal{X}$ con $0 < \mu(A) < 1$ y tal que $T^{-1}A = A$. En este caso es claro que la órbita de los puntos de A y los de $X \setminus A$ no tienen relación y por lo tanto los promedios de Birkhoff no necesariamente serán iguales. De hecho:

$$0 < \mu(\{x \in X : \frac{1}{n} \sum_{k=0}^{n-1} 1_A(T^k x) = 1\}) = \mu(A) < 1$$

Queremos distinguir ahora los procesos, o sistemas dinámicos que evitan este fenómeno. Lo hacemos con las siguiente definiciones.

Definición 2.9 (Conjunto invariante)**.** Sea $(X, \mathcal{X}, \mu, T)$ un sistema dinámico que preserva probabilidad. Se dice que un conjunto $A \in \mathcal{X}$ es invariante si cumple $T^{-1}A = A$. Los conjuntos invariantes forman una σ-álgebra que denotaremos $\mathcal{X}_T$.

Definición 2.10 (Sistema Dinámico Ergódico)**.** Sea $(X, \mathcal{X}, \mu, T)$ un sistema dinámico que preserva probabilidad. Se dice que el sistema es ergódico (a veces se dice que T es ergódica o que μ es una medida ergódica) si todo conjunto invariante tiene probabilidad cero o uno.

La traducción de este concepto a procesos estacionarios se hace mediante la transformación shift.

Definición 2.11 (Proceso Ergódico)**.** Sea $X_1, \ldots, X_n, \ldots$ un proceso estacionario de distribución μ. Se tiene que $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}, \mu, T)$ es un sistema dinámico que preserva probabilidad, donde T es la transformación shift. Se dice que el proceso es ergódico si para todo $A \in \mathcal{B}_T$ se tiene $\mu(A) = 0$ o $\mu(A) = 1$.

Una primera observación es que aunque un sistema dinámico que preserva probabilidad $(X, \mathcal{X}, \mu, T)$ no sea ergódico, existen funciones $f : X \to \mathbb{R}$ medibles tales que el proceso definido mediante $X_n = f \circ T^{n-1}$ si lo es. El ejemplo obvio son las funciones constantes.

Un sistema dinámico $(X, \mathcal{X}, \mu, T)$ es ergódico si y sólo si para toda función $f : X \to \mathbb{R}$ medible el proceso $X_n = f \circ T^{n-1}$ lo es.

En el caso de la transformación shift se tiene que todo conjunto invariante es también un evento de cola (aunque el recíproco no es cierto como muestra el conjunto de sucesiones que valen cero en infinitas de sus coordenadas pares) esto nos da el siguiente corolario de la Ley Cero-Uno de Kolmogorov.

Corolario 2.12. *Si $X_1, \ldots, X_n, \ldots$ son variables aleatorias independientes e idénticamente distribuidas entonces forman un proceso estacionario y ergódico.*

El análogo a lo que eran las funciones de cola para procesos independientes son las llamadas funciones invariantes:

Definición 2.13 (Función Invariante)**.** Dada una transformación $T : X \to X$ de un conjunto cualquiera X en si mismo se dice que una función $f : X \to \{-\infty\} \cup \mathbb{R} \cup \{+\infty\}$ es invariante si se cumple $f = f \circ T$.

El siguiente lema es análogo al que enunciamos sobre funciones de cola para procesos independientes:

Lema 2.14. *Si $(X, \mathcal{X}, \mu, T)$ es un sistema dinámico ergódico y $f : X \to \mathbb{R}$ es medible e invariante entonces existe $c \in \{-\infty\} \cup \mathbb{R} \cup \{+\infty\}$ tal que:*

$$\mu(\{x \in X : f(x) = c\}) = 1$$

es decir f es constante c.t.p. (casi todo punto).

Las funciones:

$$\limsup_n \frac{1}{n} S_n \text{ y } \liminf_n \frac{1}{n} S_n$$

son invariantes bajo el shift cuando pasamos el proceso a $\mathbb{R}^{\mathbb{N}}$ (mediante el viejo truco de la distribución) y por lo tanto obtenemos lo siguiente.

Lema 2.15. *Si $X_1, \dots, X_n, \dots$ es un proceso estacionario ergódico entonces:*

$$\operatorname*{lím\,sup}_n \frac{1}{n} S_n \text{ y } \operatorname*{lím\,inf}_n \frac{1}{n} S_n$$

son constantes casi seguramente (i.e. en un conjunto de probabilidad 1*)*

Vemos entonces que para los procesos ergódicos es razonable que haya una demostración más sencilla del Teorema Ergódico de Birkhoff, porque el límite de los promedios, en caso de existir, debería ser constante. Como la Ley de Grandes Números es un caso particular, es razonable que el limite sea $\mathbb{E}(X_1)$. En la siguiente sección damos una demostración de este hecho.

2.4. Lema de los Líderes

Consideremos la sucesión finita $2, 1, -2, 10, -8, 0, 4, -6$. Llamamos líderes a aquellos términos que encabezan alguna suma de términos consecutivos cuyo resultado es menor o igual a cero. En este caso los líderes son:

$$L = \{1, -2, 10, -8, 0, 4, -6\}$$

La inclusión de $-2, -8, 0$ y -6 es trivial porque los términos mismos son menores o iguales a cero. El término 1 es lider por la suma $1 - 2$. 10 es lider porque tenemos la suma $10 - 8 + 0 + 4 - 6 \leq 0$.

La observación crucial para esta sección es que la suma de los lideres es menor o igual a cero.

Formalizamos la definición como sigue:

Definición 2.16 (Líderes de una sucesión finita)**.** Si $x_1, \dots, x_n \in \mathbb{R}$ decimos que x_k es lider de la sucesión si existe $l \geq k$ tal que

$$x_k + \cdots + x_l \leq 0$$

La observación sobre la suma de los líderes se convierte en el siguiente lema:

Lema 2.17 (Lema de los líderes (F.Riesz))**.** *Si $x_1, \dots, x_n \in \mathbb{R}$ y se define $L = \{k \in \{1, \dots, n\} : x_k \text{ es lider en } x_1, \dots, x_n\}$ entonces:*

$$\sum_{k \in L} x_k \leq 0$$

Demostración: Si $L = \emptyset$ entonces el enunciado es cierto (con la convención clásica de que sumas vacias valen cero).

Si $L \neq \emptyset$ tomemos $k = \min(L)$ y $l \geq k$ tal que $x_k + \cdots + x_l \leq 0$. Dado que los términos no lideres son positivos se tiene:

$$\sum_{i \in L \cap \{1,\ldots,l\}} x_i = \sum_{i \in L \cap \{k,\ldots,l\}} x_i \leq x_k + \cdots + x_l \leq 0$$

El resultado se deduce por inducción dado que $L \cap \{l+1,\ldots,n\}$ es simplemente el conjunto índices de los líderes de la sucesión $x_{l+1},\ldots,x_n$ (la propiedad de ser lider depende sólamente de los términos siguientes al término dado). □

La consecuencia de interés para nosotros es la siguiente:

Lema 2.18 (Karlsson-Margulis)**.** *Si $X_1,\ldots,X_n,\ldots$ es una sucesión estacionaria de variables aleatorias de esperanza positiva entonces:*

$$\mathbb{P}(S_n > 0 \text{ para todo } n) > 0$$

Demostración: Por el lema de los líderes se tiene que para todo n se cumple:

$$\sum_{1\leq k\leq n} X_k 1_{\{X_k \text{ es lider en } X_k,\ldots,X_n\}} \leq 0$$

Tomando esperanza:

$$\sum_{1\leq k\leq n} \mathbb{E}(X_k 1_{\{X_k \text{ es lider en } X_k,\ldots,X_n\}}) \leq 0$$

Utilizando la estacionariedad:

$$\sum_{1\leq k\leq n} \mathbb{E}(X_1 1_{\{X_1 \text{ es lider en } X_1,\ldots,X_{n+1-k}\}}) \leq 0$$

Reordenando los índices:

$$\sum_{1\leq k\leq n} \mathbb{E}(X_1 1_{\{X_1 \text{ es lider en } X_1,\ldots,X_k\}}) \leq 0$$

Usando la definición de lider se tiene que si definimos

$$A_k = \{S_i \leq 0 \text{ para algún } i \leq k\}$$

obtuvimos la desigualdad:

$$\sum_{1\leq k\leq n} \mathbb{E}(X_1 1_{A_k}) \leq 0$$

Observemos que $A_k \subset A_{k+1}$ para todo k. Si se tuviera $\lim_k \mathbb{P}(A_k) = 1$ entonces existe $K > 0$ tal que $\mathbb{E}(X_1 1_{A_k}) > \frac{1}{2}\mathbb{E}(X_1) > 0$ para todo $k > K$. Esto está en contradicción con la validéz de la desigualdad obtenida para todo n. □

Lo anterior es suficiente para demostrar el Teorema Ergódico de Birkhoff en el caso ergódico:

Teorema 2.19 (Teorema Ergódico de Birkhoff: Caso Ergódico)**.** *Si* $X_1, \ldots, X_n, \ldots$ *es una sucesión estacionaria y ergódica de variables aleatorias de esperanza finita entonces:*

$$\mathbb{P}(\frac{1}{n}S_n \to \mathbb{E}(X_1)) = 1$$

Demostración: Fijemos $a < \mathbb{E}(X_1)$. Aplicando el lema anterior a $X_n - a$ se obtiene que:

$$\mathbb{P}(\frac{1}{n}S_n > a \text{ para todo } n) > 0$$

De esto se deduce que

$$\mathbb{P}(\liminf\{\frac{1}{n}S_n\} \geq a) > 0$$

y por la ergodicidad

$$\mathbb{P}(\liminf\{\frac{1}{n}S_n\} \geq a) = 1$$

Como esto vale para todo $a < \mathbb{E}(X_1)$ se obtiene:

$$\mathbb{P}(\liminf\{\frac{1}{n}S_n\} \geq \mathbb{E}(X_1)) = 1$$

Aplicando el mismo razonamiento a la sucesión $2\mathbb{E}(X_1) - X_n$ se obtiene:

$$\mathbb{P}(\limsup\{\frac{1}{n}S_n\} \leq \mathbb{E}(X_1)) = 1$$

Por lo tanto se concluye el enunciado. □

Notemos que en el caso de variables independientes, lo anterior es una demostración de la Ley de Grandes Números que utiliza sólamente la Ley Cero-Uno de Kolmogorov.

2.5. Radón-Nikodym, Esperanza Condicional

Demostraremos ahora el caso no ergódico del Teorema Ergódico de Birkhoff. En esta sección utilizaremos el lenguaje dinámico, principalmente porque facilita hablar del concepto de conjunto invariante.

Sea entonces $(X, \mathcal{X}, \mu, T)$ un sistema dinámico que preserva probabilidad. Y tomememos una función $f \in L^1(X, \mathcal{X}, \mu)$. La identidad:

$$\int_A f \mathrm{d}\mu = \int_A \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k \mathrm{d}\mu \ \forall n \in \mathbb{N}$$

válida para todo conjunto invariante $A \in \mathcal{X}_T$, hace razonable pensar que el límite de los promedios en caso de existir debe integrar lo mismo que f en cualquier conjunto invariante.

El primer paso para demostrar esto entonces es mostrar que existe una tal función.

Teorema 2.20. *Sea $(X, \mathcal{X}, \mu, T)$ un sistema dinámico que preserva probabilidad y sea $f \in L^1(X, \mathcal{X}, \mu)$. Existe una función $g \in L^1(X, \mathcal{X}_T, \mu)$ (i.e. g es medible, invariante, y μ integrable) tal que:*

$$\int_A f \mathrm{d}\mu = \int_A g \mathrm{d}\mu \ \forall A \in \mathcal{X}_T$$

Demostración: La ecuación $\nu(A) = \int_A f \mathrm{d}\mu$ define una medida signada en la σ-álgebra de conjuntos invariantes $\mathcal{X}_T$. Además esta medida signada es absolutamente continua respecto a μ. Por lo tanto el teorema de Radon-Nikodym garantiza la existencia de $g = \frac{\mathrm{d}\nu}{\mathrm{d}\mu_{|\mathcal{X}_T}}$ y su unicidad a menos de modificaciones en conjuntos invariantes de medida nula. $\square$

La función g del teorema anterior es llamada la esperanza condicional de f respecto a la σ-álgebra $\mathcal{X}_T$ y se denota:

$$g = \mathbb{E}(f|\mathcal{X}_T)$$

El teorema anterior nos permite restringirnos a probar el Teorema Ergódico de Birkhoff en el caso de una función $f \in L^1(X, \mathcal{X}, \mu)$ que integra cero en cada conjunto invariante (i.e. $\mathbb{E}(f|\mathcal{X}_T) = 0$).

Como es razonable, la demostración del teorema en el caso general se basa en una modificación del lema de Karlsson-Margulis para conjuntos invariantes.

Lema 2.21 (Karlsson-Margulis modificado)**.** *Sea* $(X, \mathcal{X}, \mu, T)$ *un sistema dinámico que preserva probabilidad y* $f \in L^1(X, \mathcal{X}, \mu)$*. Para todo* $A \in \mathcal{X}_T$ *de medida positiva y todo* $a < \frac{1}{\mu(A)} \int_A f \mathrm{d}\mu$ *existe un subconjunto* $B \subset A$ *medible, invariante, y con* $\mu(B) > 0$ *tal que:*

$$\liminf_n \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k x \geq a \ \forall x \in B$$

Demostración: Aplicamos el lema de Karlsson-Margulis de la sección anterior a $(A, \mathcal{X} \cap A, \frac{\mu}{\mu(A)}, T_{/A})$ y la función $f - a$. Con esto se obtiene que:

$$\mu(\{x \in A : \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k x > a \ \forall n \in \mathbb{N}\}) > 0$$

Tomemos ahora

$$B = \bigcup_{n \geq 0} T^{-n}\{x \in A : \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k x > a \ \forall n \in \mathbb{N}\}$$

se tiene que $\mu(B_0) > 0$ y B_0 cumple el enunciado. excepto que no es invariante.

Sin embargo, para todo $n \geq 0$ se cumple $T^{-n} B_0 \subset B_0$ y ambos conjuntos miden lo mismo. Por lo tanto definiendo:

$$B = \bigcap_{n \geq 0} T^{-n} B_0$$

obtenemos un conjunto invariante de probabilidad positiva que cumple el enunciado. □

Damos ahora la demostración del Teorema Ergódico de Birkhoff.

Demostración del Teorema Ergódico de Birkhoff: Supongamos $(X, \mathcal{X}, \mu, T)$ un sistema dinámico que preserva probabilidad, y $f \in L^1(X, \mathcal{X}, \mu)$. Por el Teorema 2.20 existe g medible e invariante que integra lo mismo que f en todo conjunto invariante. Alcanza probar el teorema para $f - g$ y por lo tanto asumimos sin pérdida de generalidad $\mathbb{E}(f|\mathcal{X}_T) = 0$.

Definamos (incluyendo explícitamente el caso $+\infty - (-\infty) = +\infty$)

$$A_\epsilon = \{x \in X : \limsup_n \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k x - \liminf_n \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k x > \epsilon\})$$

y supongamos por absurdo que para cierto $\epsilon > 0$ se tiene

$$\mu(A_\epsilon) > 0$$

Por el lema de Karlsson-Margulis se tiene un conjunto medible invariante $A \subset A_\epsilon$ con $\mu(A) > 0$ tal que

$$\liminf_n \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k x \geq -\frac{\epsilon}{3} \ \forall x \in A$$

Además el mismo lemma aplicado a $-f$ nos muestra que existe un conjunto medible invariante $B \subset A$ con $\mu(B) > 0$ tal que:

$$\limsup_n \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k x \leq \frac{\epsilon}{3} \ \forall x \in B$$

Pero esto contradice que $B \subset A_\epsilon$. $\square$

Capítulo 3

Teorema de Kingman

Este capítulo tiene dos objetivos principales.

El primero de ellos es demostrar el Teorema Ergódico Subaditivo de Kingman [Kin68]. El teorema es interesante por la cantidad de aplicaciones que tiene, por ejemplo da como corolario trivial el clásico teorema de Furstenberg y Kesten [FK60] sobre la norma de un producto aleatorio de matrices.

Otro objetivo es demostrar el lema de Karlsson-Margulis en su versión subaditiva (ver [KM99]).

La demostración que damos del teorema de Kingman en el caso ergódico es básicamente la de [KM99] con una notación un poco modificada. Nos tomamos el trabajo de extenderla al caso no ergódico, cosa que los autores no hicieron.

Introducimos el concepto de distancia dirigida porque captura lo que para nosotros es un proceso subaditivo. En sí, esta definición no cumple otra función que hacer más intuitivos algunos de los pasos que se utilizan para acotar este tipo de procesos.

La notación se adapta bien a la familia de procesos subaditivos que estudiaremos más adelante en la tesis, que son todos de la forma $d(x_n, x_m)$ donde $\{x_n\}$ es una cierta sucesión aleatoria en un espacio métrico (por ejemplo, la parte positiva de los productos parciales de una sucesión de matrices aleatorias).

3.1. El lema de Fekete

Definición 3.1 (Sucesión subaditiva). Decimos que una sucesión $a \in \mathbb{R}^{\mathbb{N}}$ es subaditiva si se cumple:

$$a(m+n) \leq a(m) + a(n) \ \forall m, n \in \mathbb{N}$$

Los ejempos básicos vienen de funciones concavas como $\log(n)$ y $\sqrt{n}$.

La observación básica respecto a este tipo de sucesiones es que $\frac{1}{n}a(n)$ tiene límite (posiblemente igual a $-\infty$).

Lema 3.2 (M.Fekete). *Si $a \in \mathbb{R}^{\mathbb{N}}$ es subaditiva entonces:*

$$\lim_n \frac{1}{n}a(n) = \inf_n \frac{1}{n}a(n) \in \{-\infty\} \cup \mathbb{R}$$

Demostración: Dado $t > \inf_n \frac{1}{n}a(n)$ existe $m \in \mathbb{N}$ tal que:

$$\frac{1}{m}a(m) < t$$

Para cada $n \in \mathbb{N}$ escribimos $n = qm + r$ con $q, r \geq 0$ y $r < m$. Con esto se obtiene:

$$\frac{1}{n}a(n) \leq \frac{1}{n}(qa(m) + ra(1)) \leq \frac{q}{n}a(m) + \frac{m}{n}|a(1)|$$

y por lo tanto tomando límite superior

$$\limsup_n \frac{1}{n}a(n) \leq \frac{1}{m}a(m) < t$$

□

3.2. Distancias Dirigidas

Convengamos la siguiente notación para los enteros no negativos $\mathbb{Z}_+ = \{0\} \cup \mathbb{N}$.

Definición 3.3 (Distancia Dirigida). Decimos que $a \in \mathbb{R}^{\mathbb{Z}_+^2}$ es una distancia dirigida si:

$$a(n,n) = 0 \ \forall n \in \mathbb{Z}_+$$

$$a(m,n) = -a(n,m) \ \forall m, n \in \mathbb{Z}_+$$

$$a(l,n) \leq a(l,m) + a(m,n) \ \forall l < m < n$$

El primer ejemplo básico es definir a mediante:

$$a(m,n) = n - m$$

En este caso se cumple $a(l,n) = a(l,m) + a(m,n)$ para todo $l,m,n \in \mathbb{Z}_+$. Todas las distancias dirigidas que cumplen esto son de la forma:

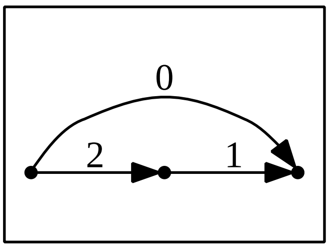
$$a(m,n) = \begin{cases} \sum_{k=m}^{n-1} a_k & \text{si } m \leq n \\ -\sum_{k=n}^{m-1} a_k & \text{si } m > n \end{cases}$$

para cierta sucesión a_k.

Otra manera de fábricar ejemplos es tomar una distancia d en $\mathbb{Z}_+$ y definir:

$$a(m,n) = \begin{cases} d(m,n) & \text{si } m \leq n \\ -d(m,n) & \text{si } m > n \end{cases}$$

El ejemplo siguiente muestra que no todas las distancias dirigidas vienen de distancias (ni siquiera las que toman valores positivos cuando $m < n$):



En este ejemplo $a(0,1) = 2, a(1,2) = 1, a(0,2) = 0$ y se tiene $|a(0,1)| = 2 > 1 = |a(0,3)| + |a(3,2)|$.

Toda función definida para $m < n$ que cumple:

$$a(l,n) \leq a(l,m) + a(m,l) \ \forall l < m < n$$

puede extenderse de manera única a una distancia dirigida definiendo $a(n,n) = 0$ para todo n y $a(n,m) = -a(m,n)$ para todo $m < n$.

Lo anterior muestra que si consideramos $\mathbb{Z}_+$ como los vértices de un grafo dirigido cuyas aristas son los pares (m,n) con $m < n$ dar una distancia dirigida es equivalente a asignar un peso a cada arista. Con esto cada camino en el grafo (dado por una sucesión finita $(n_1, \ldots, n_k) \in \mathbb{Z}_+^k$)) tiene una longitud dada por:

$$\sum_{j=1}^{k-1} a(n_j, n_{j+1})$$

La condición de subaditividad garantiza que si el camino es creciente (i.e. $n_i \leq n_{i+1}$ $\forall i$) entonces:

$$a(n_1, n_k) \leq \sum_{j=1}^{k-1} a(n_j, n_{j+1})$$

Para caminos no crecientes la longitud puede ser menor a la de la arista que une los dos extremos. De hecho una propiedad importante es:

$$a(m, n) + a(n, m+1) \leq a(m, m+1) \ \forall m, n \in \mathbb{Z}_+$$

Definamos para todo $m, n \in \mathbb{Z}_+$ la función $b(m, n) = a(m, n) + a(n, m+1)$. Se tiene entonces:

$$b(m, n) \leq a(m, m+1) \ \forall n \in \mathbb{Z}_+$$

Necesitaremos también las siguientes identidades:

$$a(0, n) = b(0, n) + b(1, n) + \cdots + b(n-1, n)$$

$$a(0, n) + a(n, k) = b(0, n) + b(1, n) + \cdots + b(k-1, n)$$

3.3. Procesos Subaditivos

Definiremos ahora lo que significa elegir una distancia dirigida al azar.

Definición 3.4 (Elemento Aleatorio)**.** Un elemento aleatorio de un espacio topológico X es una función Borel medible $a : \Omega \to X$ donde $(\Omega, \mathcal{F}, \mathbb{P})$ es un espacio de probabilidad. La distribución de a es la medida boreliana μ en X definida mediante:

$$\mu(A) = \mathbb{P}(a \in A)$$

Consideramos ahora el conjunto de las distancias dirigidas $D \subset \mathbb{R}^{\mathbb{Z}_+^2}$ con la topología de la convergencia puntual.

Definición 3.5 (Proceso Subaditivo)**.** Un proceso subaditivo es un elemento aleatorio a del espacio de distancias dirigidas.

Una definición equivalente es que un proceso subaditivo es un conjunto $\{x(n,m)\}_{n,m\in\mathbb{Z}_+}$ de variables aleatorias que cumplen:

$$x(m,n) = -x(n,m) \ \forall m,n \in \mathbb{Z}_+$$

$$x(l,n) \leq x(l,m) + x(m,n) \ \forall l,m,n \in \mathbb{Z}_+ \text{ con } l < m < n$$

El ejemplo obvio consiste en tomar una sucesión $X_1,\ldots,X_n,\ldots$ de variables aleatorias y definir:

$$a(n,m) = X_{n+1} + \cdots + X_m \text{ para todo } n < m$$

extendiendolo luego a una proceso subaditivo de la manera obvia (la mayoría de los ejemplos vendrán definidos sólamente para $m < n$ y se dará por entendido que se toma la única extensión posible).

Un ejempo un poco menos obvio a partir de las mismas variables aleatorias es:

$$a(n,m) = |X_{n+1} + \cdots + X_m| \text{ para todo } n < m$$

Si en los ejemplos anteriores la sucesión $X_1,\ldots,X_n,\ldots$ es estacionaria entonces el proceso subaditivo a es estacionario en el sentido que definiremos a continuación.

Definición 3.6. Sea el shift $T : D \to D$ dado por

$$(Ta)(n,m) = a(n+1,m+1)$$

Un proceso subaditivo a es estacionario si su distribución es invariante bajo T.

Esta definición es equivalente a decir que para todo $n \in \mathbb{N}$ y todo conjunto de borelianos $\{A_{ij}\}_{0\leq i,j\leq n}$ se tiene:

$$\mathbb{P}(a(i,j) \in A_{ij} \text{ para todo } 0 \leq i,j \leq n) = \mathbb{P}(a(i{+}1,j{+}1) \in A_{ij} \text{ para todo } 0 \leq i,j \leq n)$$

En particular, para un proceso subaditivo estacionario, la distribución de $a(m,n)$ depende únicamente de $n-m$. Más en particular aún, si se tuviera que $a(m,n)$ tiene esperanza finita entonces $\mathbb{E}(a(m,n)) = \mathbb{E}(a(0,n-m))$.

Notemos además que si a es un proceso subaditivo y $\mathbb{E}(a(m,n))$ existe para todo m,n entonces se tiene:

$$\mathbb{E}(a(0,n{+}m)) \leq \mathbb{E}(a(0,n)) + \mathbb{E}(a(n,n{+}m)) = \mathbb{E}(a(0,n)) + \mathbb{E}(a(0,m)) \ \forall n,m \in \mathbb{N}$$

y por lo tanto por el lema de Fekete

$$\lim_n \frac{1}{n}\mathbb{E}(a(0,n)) = \inf_n \frac{1}{n}\mathbb{E}(a(0,n)) \in \{-\infty\} \cup \mathbb{R}$$

Enunciamos ahora el teorema de Kingman ([Kin68]), cuya demostración es uno de los objetivos de este capítulo:

Teorema 3.7 (Teorema Ergódico de Kingman). *Si a es un proceso subaditivo y estacionario que cumple*

$$\mathbb{E}(|a(m,n)|) < +\infty \textit{ para todo } m,n \in \mathbb{Z}_+$$

$$\lim_n \frac{1}{n}\mathbb{E}(a(0,n)) > -\infty$$

entonces

$$\mathbb{P}(\exists \lim_n \frac{1}{n} a(0,n)) = 1$$

Los ejemplos muestran que este teorema generaliza el Teorema Ergódico de Birkhoff.

Dado un sistema dinámico que preserva probabilidad $(X, \mathcal{X}, \mu, T)$ y una sucesión de funciones medibles $f_n : X \to \mathbb{R}$ que cumplen:

$$f_{n+m} \leq f_n + f_m \circ T^n \ \forall n, m \in \mathbb{N}$$

se obtiene un proceso subaditivo estacionario mediante la ecuación:

$$a(m,n) = f_{n-m} \circ T^m \ \forall m < n$$

Si las funciones f_n cumplen la igualdad

$$f_{n+m} = f_n + f_m \circ T^n$$

entonces se tiene:

$$f_n = f_1 + \cdots + f_1 \circ T^{n-1}$$

Esto muestra como enunciar el Teorema Ergódico de Kingman en el contexto dinámico:

Teorema 3.8 (Teorema Ergódico de Kingman Versión Dinámica)**.** *Sea* $(X, \mathcal{X}, \mu, T)$ *un sistema dinámico que preserva probabilidad. Si la sucesión* $\{f_n\}_{n\in\mathbb{N}} \subset L^1(X, \mathcal{X}, \mu)$ *cumple:*

$$f_{n+m} \leq f_n + f_m \circ T^n \ \forall n, m \in \mathbb{N}$$

$$\lim_n \frac{1}{n} \int f_n \mathrm{d}\mu > -\infty$$

entonces:

$$\mu(\{x \in X : \exists \lim_n \frac{1}{n} f_n(x)\}) = 1$$

Las dos versiones del Teorema Ergódico de Kingman son equivalentes. La demostración de este hecho, es análoga al caso del Teorema Ergódico de Birkhoff que hicimos anteriormente.

3.4. Acotación Superior

En esta sección utilizaremos que $a(0, n)$ puede acotarse superiormente por un proceso aditivo más un resto pequeño. Esto se hace mediante la desigualdad básica

$$a(0, n) \leq a(0, m) + a(m, 2m) + \cdots + a((q-1)m, qm) + a(qm, n)$$

donde $n = qm + r$ y $0 \leq r < m$.

Aplicaremos la versión aditiva del lema de Karlsson-Margulis (lema 2.18) para acotar superiormente los valores de $a(0, n)$ en un conjunto de probabilidad positiva.

Para mostrar que el resto no afecta esta acotación utilizaremos el siguiente hecho elemental de probabilidad:

Lema 3.9. *Si* $X_1, \ldots, X_n, \ldots$ *es una sucesión de variables que tienen la misma distribución y son tales que* $\mathbb{E}(|X_1|) < +\infty$ *entonces en un conjunto de probabilidad* 1 *se tiene:*

$$\frac{1}{n} X_n \to 0$$

Demostración: La demostración consiste en notar que para todo $\epsilon > 0$ la variable truncada:

$$\sum_{k=1}^{+\infty} k\epsilon 1_{\{|X_1| \in [k\epsilon, (k+1)\epsilon)\}} \leq |X_1|$$

tiene esperanza finita y por lo tanto:

$$\sum_{k=1}^{+\infty} k\epsilon\mathbb{P}(|X_1| \in [k\epsilon, (k+1)\epsilon)) < +\infty$$

reordenando esta suma absolutamente convergente y cambiando probabilidades del tipo $\mathbb{P}(X_1 \in A)$ por $\mathbb{P}(X_n \in A)$ (que son iguales porque las variables tienen igual distribución) se obtiene:

$$\sum_{k=1}^{+\infty} \mathbb{P}(\frac{1}{k}|X_k| > \epsilon) < +\infty$$

Por último se utiliza Borel-Cantelli (lema 1.2) para concluir que en un conjunto de probabilidad 1 se tiene $\frac{1}{k}|X_k| \geq \epsilon$ sólamente para un número finito de valores de $k \in \mathbb{N}$. □

El teorema de esta sección es el siguiente:

Teorema 3.10 (Lema de Acotación Superior)**.** *Si a es un proceso subaditivo estacionario tal que $a(m,n) \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ para todo $m, n \in \mathbb{Z}_+$ y además*

$$\lim_n \frac{1}{n}\mathbb{E}(a(0,n)) < 0$$

entonces

$$\mathbb{P}(\exists N \in \mathbb{N} : a(0,n) < 0 \text{ para todo } n > N) > 0$$

Demostración: Tomemos $m \in \mathbb{N}$ tal que $\mathbb{E}(a(0,m)) < t < 0$. Por el lema 2.18 aplicado a la sucesión estacionaria $t-a(0,m), \ldots, t-a(km,(k+1)m), \ldots$ se tiene que:

$$\mathbb{P}(\sum_{k=0}^{n-1} a(km,(k+1)m) < nt \ \forall n \in \mathbb{N}) > 0$$

Ahora para cada $n \in \mathbb{N}$ escribimos $n = qm+r$ con $0 \leq r < m$ y utilizando la subaditividad obtenemos en el conjunto de probabilidad positiva obtenido anteriormente:

$$\begin{aligned} a(0,n) &\leq \sum_{k=0}^{q-1} a(km,(k+1)m) + a(qm, qm+r) \\ &< qt + a(qm, qm+r) \\ &\leq q(t + \frac{a(qm,qm+1)^+ + \cdots + a(qm+m-1, qm+m)^+}{q}) \end{aligned}$$

Por el lema anterior en un conjunto de probabilidad 1:

$$\frac{a(qm, qm+1)^+ + \cdots + a(qm+m-1, qm+m)^+}{q} \to 0 \text{ cuando } q \to +\infty$$

con lo cual se obtiene la tesis. □

Notemos que la variable $a(0,1)$ podría ser siempre estríctamente positiva sin contradecir las hipótesis, por lo tanto no es posible mejorar el resultado demostrando que con probabilidad positiva se cumple $a(0,n) < 0 \forall n \in \mathbb{N}$.

3.5. Lema de Karlsson-Margulis

Modificaremos ahora los métodos utilizados en la versión aditiva del lema de Karlsson-Margulis (lema 2.18) para acotar inferiormente los valores de $a(0,n)$ en un conjunto de probabilidad positiva. En particular mostraremos el siguiente resultado:

Corolario 3.11. *Si a es un proceso subaditivo estacionario tal que $a(m,n) \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ para todo $m, n \in \mathbb{Z}_+$ y además*

$$\lim_n \frac{1}{n}\mathbb{E}(a(0,n)) > 0$$

entonces

$$\mathbb{P}(a(0,n) > 0 \text{ para todo } n \in \mathbb{N}) > 0$$

El resultado siguiente, que llamaremos el lema de Karlsson-Margulis (versión subaditiva) implica inmediatamente el anterior mediante la desigualdad:

$$a(0,n) + a(n,k) \leq a(0,k) \ \forall k \leq n$$

Teorema 3.12 (Lema de Karlsson-Margulis)**.** *Si a es un proceso subaditivo estacionario tal que $a(m,n) \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ para todo $m, n \in \mathbb{Z}_+$ y además*

$$\lim_n \frac{1}{n}\mathbb{E}(a(0,n)) > 0$$

entonces

$$\mathbb{P}(\text{ para infinitos } n \text{ se cumple } a(0,n) + a(n,k) > 0 \text{ para todo } k \in \{1, \ldots, n\}) > 0$$

DEMOSTRACIÓN: Definamos b mediante:

$$b(m,n) = a(m,n) + a(n,m+1)$$

Se tiene que:

$$a(0,n) = \sum_{k=0}^{n-1} b(k,n)$$

Aplicando el lema de los líderes a $b(0,n),\ldots,b(n-1,n)$ obtenemos:

$$\sum_{k=0}^{n-1} b(k,n) 1_{\{b(k,n) \text{ lider en } b(k,n),\ldots,b(n-1,n)\}} \leq 0$$

tomando esperanza

$$\sum_{k=0}^{n-1} \mathbb{E}(b(k,n) 1_{\{b(k,n) \text{ lider en } b(k,n),\ldots,b(n-1,n)\}}) \leq 0$$

usando la estacionariedad

$$\sum_{k=0}^{n-1} \mathbb{E}(b(0,n-k) 1_{\{b(0,n-k) \text{ lider en } b(0,n-k),\ldots,b(n-k-1,n-k)\}}) \leq 0$$

reordenando los índices

$$\sum_{k=1}^{n} \mathbb{E}(b(0,k) 1_{\{b(0,k) \text{ lider en } b(0,k),\ldots,b(k-1,k)\}}) \leq 0$$

Definiendo

$$\begin{aligned} A_k &= \{b(0,k) \text{ lider en } b(0,k),\ldots,b(k-1,k)\} \\ &= \{\exists j \in \{0,\ldots,k-1\} : b(0,k) + \cdots + b(j,k) \leq 0\} \\ &= \{\exists j \in \{1,\ldots,k\} : a(0,k) + a(k,j) \leq 0\} \end{aligned}$$

hemos obtenido

$$\sum_{k=1}^{n} \mathbb{E}(b(0,k) 1_{A_k}) \leq 0 \ \forall n \in \mathbb{N}$$

Definamos por conveniencia los complementos de los A_k:

$$B_k = \{a(0,k) + a(k,j) > 0 \ \forall j \in \{1,\ldots,k\}\}$$

Nuestro objetivo es mostrar que la probabilidad de pertenecer a infinitos B_k es positiva, es decir:

$$\mathbb{P}(\bigcap_n \bigcup_{k>n} B_k) > 0$$

Para esto mostraremos simplemente que $\mathbb{P}(B_k)$ no tiende a cero cuando $k \to +\infty$.

Utilizando la estacionariedad se obtiene lo siguiente

$$\sum_{k=1}^{n} \mathbb{E}(b(0,k)) = \sum_{k=0}^{n-1} \mathbb{E}(b(k,n)) = \mathbb{E}(a(0,n)) = nC + o(n)$$

donde $C = \text{lím}_n \frac{1}{n}\mathbb{E}(a(0,n)) > 0$.

Esto junto con la igualdad

$$\sum_{k=1}^{n} \mathbb{E}(b(0,k)) = \sum_{k=1}^{n} \mathbb{E}(b(0,k)1_{A_k}) + \sum_{k=1}^{n} \mathbb{E}(b(0,k)1_{B_k})$$

y el hecho de que el primer sumando es menor o igual a cero muestra que

$$nC + o(n) \leq \sum_{k=1}^{n} \mathbb{E}(b(0,k)1_{B_k})$$

Ahora de lo anterior y la desigualdad $b(0,k) \leq a(0,1)$ se obtiene (dividiendo entre n):

$$C + o(1) \leq \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}(a(0,1)1_{B_k})$$

Dado que $a(0,1) \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ la función $\nu(A) = \mathbb{E}(a(0,1)1_A)$ es una medida signada y es absolutamente continua respecto a $\mathbb{P}$. Por lo tanto si cuando $k \to +\infty$ se tiene que $\mathbb{P}(B_k) \to 0$ entonces $\nu(B_k) \to 0$. Esto implicaría que el lado derecho de la desigualdad anterior tendería a cero.

Como por hipótesis $C > 0$ esto no puede suceder y por lo tanto:

$$\mathbb{P}(\bigcap_n \bigcup_{k>n} B_k) > 0$$

□

3.6. El caso Ergódico

Recordemos el espacio D de distancias dirigidas con la topología de la convergencia puntual y el shift $T : D \to D$ dada por:

$$T(a)(m,n) = a(m+1,n+1)$$

Un proceso subaditivo estacionario es un elemento aleatorio de D tal que su distribución es invariante bajo T. Si además T es ergódica para la distribución entonces decimos que el proceso es ergódico.

La hipótesis de ergodicidad simplifica la descripción del límite de $\frac{1}{n}a(0,n)$ gracias al siguiente lema:

Lema 3.13. *Si a es un proceso subaditivo y estacionario entonces:*

$$\mathbb{P}(\liminf_n \frac{1}{n}a(0,n) = \liminf_n \frac{1}{n}a(1,n+1)) = 1$$

Demostración: La desigualdad:

$$a(0,n) \leq a(0,1) + a(1,n)$$

implica que

$$\begin{aligned}\liminf_n \frac{1}{n}a(0,n) \leq \liminf_n \frac{1}{n}a(1,n) &= \liminf_n \frac{1}{n-1}a(1,n) \\ &= \liminf_n \frac{1}{n}a(1,n+1)\end{aligned}$$

Por lo tanto para todo $a \in \mathbb{R}$ se tiene:

$$\{\liminf_n \frac{1}{n}a(1,n+1) \leq a\} \subset \{\liminf_n \frac{1}{n}a(0,n) \leq a\}$$

pero por la estacionariedad la probabilidad de ambos conjuntos es la misma. Por lo tanto coinciden a menos de un conjunto de probabilidad nula. En el complemento de la union de los conjuntos de medida nula que se obtienen para $a \in \mathbb{Q}$ se cumple la tesis. □

El lema anterior implica que $\liminf_n \frac{1}{n}a(0,n)$ coincide en un conjunto de probabilidad 1 con una función invariante. Por lo tanto bajo la hipótesis de ergodicidad debe ser casi seguramente constante. Claramente la afirmación análoga para $\limsup_n \frac{1}{n}a(0,n)$ tiene la misma demostración. Hemos mostrado entonces el siguiente resultado:

Lema 3.14. *Si a es un proceso subaditivo, estacionario y ergódico entonces las variables,* $\limsup_n \frac{1}{n}a(0,n)$ *y* $\liminf_n \frac{1}{n}a(0,n)$*, son casi seguramente constantes.*

Con esto podemos dar una demostración del caso ergódico del Teorema Ergódico de Kingman.

Teorema 3.15 (Teorema Ergódico de Kingman: Caso Ergódico)**.** *Si a es un proceso subaditivo, estacionario y ergódico, y $C = \lim_n \frac{1}{n}\mathbb{E}(a(0,n)) \neq -\infty$ entonces:*

$$\mathbb{P}(\lim_n \frac{1}{n}a(0,n) = C) = 1$$

Demostración: Fijado $t < C$ el lema de Karlsson-Margulis (lema 3.12) aplicado al proceso subaditivo $a(m,n) - (n-m)t$ implica:

$$\mathbb{P}(\frac{1}{n}a(0,n) > t \ \forall n \in \mathbb{N}) > 0$$

y en particular

$$\mathbb{P}(\liminf_n \frac{1}{n}a(0,n) \geq t) > 0$$

Por la ergodicidad esto implica

$$\mathbb{P}(\liminf_n \frac{1}{n}a(0,n) \geq t) = 1$$

y luego intersectando en númerables t menores que C se obtiene:

$$\mathbb{P}(\liminf_n \frac{1}{n}a(0,n) \geq C) = 1$$

Fijando ahora $t > C$ y aplicando el lema de acotación superior (lema 3.10) al proceso subaditivo $a(m,n) - (n-m)t$ se obtiene:

$$\mathbb{P}(\exists N \in \mathbb{N} : \frac{1}{n}a(0,n) < t \ \forall n > N) > 0$$

y en particular

$$\mathbb{P}(\limsup_n \frac{1}{n}a(0,n) \leq t) > 0$$

Utilizando la ergodicidad e intersectando en numerables valores de t se obtiene:

$$\mathbb{P}(\limsup_n \frac{1}{n}a(0,n) \leq C) = 1$$

□

3.7. Descripción del Límite

Al igual que hicimos en el capítulo sobre el Teorema Ergódico de Birkhoff, cambiaremos a la notación dinámica para el caso no ergódico.

Supongamos que $\lim_n f_n/n = g$ existe c.t.p. Hemos mostrado en el caso ergódico que $g = \lim_n \frac{1}{n}\int f_n \mathrm{d}\mu$. En el caso general uno esperaría que al menos se cumpla:

$$\int_A g\mathrm{d}\mu = \lim_n \frac{1}{n}\int_A f_n \mathrm{d}\mu$$

en cada conjunto invariante $A \in \mathcal{X}_T$ (recordemos la notación $\mathcal{X}_T$ para la σ-álgebra de conjuntos medibles e invariantes).

Mostraremos que existe una función invariante g que cumple lo anterior, y luego mostraremos que efectivamente es el límite de f_n/n.

Recordamos antes de empezar la notación $\mathbb{E}(f|\mathcal{X}_T)$ (esperanza condicional de f respecto a la σ-álgebra $\mathcal{X}_T$) para la función (definida de manera única a menos de modificaciones en conjuntos invariantes de medidad nula) medible e invariante que integra lo mismo que f en cada conjunto invariante $A \in \mathcal{X}_T$.

Lema 3.16 (Descripción del Límite)**.** *Sea $(X, \mathcal{X}, \mu, T)$ un sistema dinámico que preserva probabilidad y $\{f_n\}_{n\in\mathbb{N}} \subset L^1(X, \mathcal{X}, \mu)$ una sucesión de funciones integrables a valores reales que cumple:*

$$f_{n+m} \leq f_n + f_m \circ T^n \ \forall m, n \in \mathbb{N}$$

$$\lim_n \frac{1}{n}\int f_n \mathrm{d}\mu > -\infty$$

Entonces existe una función $g \in L^1(X, \mathcal{X}_T, \mu)$ (i.e. medible, invariante, y μ-integrable) que cumple:

$$\int_A g\mathrm{d}\mu = \lim_n \frac{1}{n}\int_A f_n \mathrm{d}\mu \ \forall A \in \mathcal{X}_T$$

Además g es única a menos de modificaciones en conjuntos invariantes de medida nula.

Demostración: Fijemos versiones de $\mathbb{E}(f_n|\mathcal{X}_T)$ para cada $n \in \mathbb{N}$. Y notemos que puede tomarse $\mathbb{E}(f_n \circ T^m|\mathcal{X}_T) = \mathbb{E}(f_n|\mathcal{X}_T)$ para todo $m \in \mathbb{Z}_+$ (esto puede hacerse porque $\int_A f_n \circ T^m \mathrm{d}\mu = \int_A f_n \mathrm{d}\mu$ para todo $A \in \mathcal{X}_T$).

La desigualdad

$$f_{n+m} \leq f_n + f_m \circ T^n \ \forall m, n \in \mathbb{N}$$

implica que en un conjunto invariante de probabilidad 1 se cumple

$$\mathbb{E}(f_{n+m}|\mathcal{X}_T) \leq \mathbb{E}(f_n|\mathcal{X}_T) + \mathbb{E}(f_m \circ T^n|\mathcal{X}_T) \ \forall m, n \in \mathbb{N}$$

y por lo tanto

$$\mathbb{E}(f_{n+m}|\mathcal{X}_T) \leq \mathbb{E}(f_n|\mathcal{X}_T) + \mathbb{E}(f_m|\mathcal{X}_T) \ \forall m, n \in \mathbb{N}$$

Trabajaremos a partir de ahora en el conjunto en el cual se cumple la desigualdad anterior.

La sucesión $\mathbb{E}(f_n|\mathcal{X}_T)$ es subaditiva y por el lema de Fekete (lema 3.2) existe el límite:

$$g = \lim_n \frac{1}{n}\mathbb{E}(f_n|\mathcal{X}_T) = \inf_n \frac{1}{n}\mathbb{E}(f_n|\mathcal{X}_T)$$

que es una función medible e invariante por serlo cada $\mathbb{E}(f_n|\mathcal{X}_T)$.

Se tiene $\frac{1}{n}\mathbb{E}(f_n|\mathcal{X}_T) \leq \mathbb{E}(f_1|\mathcal{X}_T) \in L^1(X, \mathcal{X}_T, \mu)$ y por lo tanto podemos usar el lema de Fatou (para lím sup) y deducir que en cada conjunto invariante $A \in \mathcal{X}_T$ se cumple:

$$\int_A g \mathrm{d}\mu \geq \lim_n \int_A \frac{1}{n}\mathbb{E}(f_n|\mathcal{X}_T)\mathrm{d}\mu = \lim_n \frac{1}{n}\int_A f_n \mathrm{d}\mu$$

Esto muestra que $g \in L^1(X, \mathcal{X}_T, \mu)$ y, dado que además se tiene $g \leq \frac{1}{n}\mathbb{E}(f_n|\mathcal{X}_T)$ para todo n se obtiene la propiedad deseada para la integral de g en cada conjunto invariante.

Si g' cumple las mismas propiedades que g se tiene que $g - g'$ es invariante e integra 0 en cada conjunto invariante. Esto muestra que el conjunto invariante $\{x \in X : |g(x) - g'(x)| > \epsilon\}$ tiene medida nula para todo $\epsilon > 0$. Con esto se obtiene la unicidad buscada. □

Daremos ahora una demostración del Teorema Ergódico de Kingman:

Teorema 3.17. *Sea $(X, \mathcal{X}, \mu, T)$ un sistema dinámico que preserva probabilidad y $\{f_n\}_{n\in\mathbb{N}} \subset L^1(X, \mathcal{X}, \mu)$ una sucesión de funciones integrables a valores reales que cumple:*

$$f_{n+m} \leq f_n + f_m \circ T^n \ \forall m, n \in \mathbb{N}$$

$$\lim_{n\to+\infty} \frac{1}{n}\int f_n \mathrm{d}\mu > -\infty$$

Entonces $\lim_{n\to+\infty} f_n/n$ *existe c.t.p. y coincide con la función* g *dada por 3.16 en un conjunto invariante de medida* 1.

Demostración: Sea $g \in L^1(X, \mathcal{X}_T, \mu)$ dada por 3.16.

Considerando la sucesión de funciones $f_n - ng$ se deduce que podemos asumir

$$\lim_n \frac{1}{n}\int_A f_n \mathrm{d}\mu = \inf_n \frac{1}{n}\int_A f_n \mathrm{d}\mu = 0 \ \forall A \in \mathcal{X}_T$$

Nos restringimos al conjunto de medida 1 provisto por el lema 3.13 en el cual las funciones $\liminf_n \frac{1}{n} f_n$ y $\limsup_n \frac{1}{n} f_n$ son invariantes. Se busca demostrar que ambas valen cero c.t.p..

Definamos para cada $\epsilon > 0$ el conjunto medible e invariante

$$A_\epsilon = \{x \in X : \liminf_n \frac{1}{n} f_n(x) < -\epsilon\}$$

y supongamos por absurdo que para cierto $\epsilon > 0$ se tiene $\mu(A_\epsilon) > 0$.

El lema de Karlsson-Margulis 3.12 aplicado al espacio de probabilidad $(A_\epsilon, \mathcal{X} \cap A_\epsilon, \frac{\mu}{\mu(A_\epsilon)})$ y el proceso definido mediante $a(m,n) = f_{n-m} \circ T^m_{/A_\epsilon}$ para todo $m < n$, implica la existencia de un conjunto medible $A \subset A_\epsilon$ con $\mu(A) > 0$ tal que:

$$\frac{1}{n} f_n(x) > -\epsilon \ \forall x \in A \text{ y todo } n \in \mathbb{N}$$

y en particular

$$\liminf_n \frac{1}{n} f_n(x) \geq -\epsilon \ \forall x \in A \text{ y todo } n \in \mathbb{N}$$

Lo cual contradice la definicion de A_ϵ.

Análogamente, si definimos:

$$B_\epsilon = \{x \in X : \limsup_n \frac{1}{n} f_n(x) > \epsilon\}$$

El Lema de Acotación Superior 3.10 muestra que se tiene $\mu(B_\epsilon) = 0$ para todo $\epsilon > 0$. Y con esto, se obtiene el resultado. □

Capítulo 4

Geometría de las Matrices Positivas

En este capítulo dotamos a las matrices positivas de una estructura de variedad riemanniana y estudiamos algunas propiedades de esta estructura (e.g. sus geodésicas y propiedades métricas de los tríangulos).

Este espacio es un ejemplo de variedad de curvatura semi-negativa (o no-positiva) en la cual cualquier par de puntos está unido por una única geodésica (i.e. espacios de Cartan-Hadamard). Y, a través de la descomposición polar, juega un papel central en el teorema de Oseledets que demostraremos más adelante.

Empezamos introduciendo la norma de Frobenius que usaremos de aquí en más en todos los espacios de matrices. Luego demostramos algunas propiedades básicas del mapa exponencial, en particular mostramos la diferenciabilidad utilizando resultados de series de potencias en varias variables reales (ver [KP02]) y daremos formas exactas para el diferencial en distintos puntos.

El teorema de descomposición polar sirve para adelantar trabajo hacia la demostración del teorema de Oseledets, y también para introducir las matrices positivas y la acción natural de $\mathrm{GL}(n)$ sobre ellas.

Hemos preferido ver $\mathrm{Pos}(n)$ como una subvariedad y no como el cociente $\mathrm{GL}(n)/O(n)$ porque de este modo es más obvia la estructura diferencial y se pueden hacer cuentas más fácilmente.

El primer trabajo que estudiamos (que fué [KL06]) tenía la frase “Las matrices invertibles actúan como isometrías sobre las positivas” entre paréntesis sin ninguna referencia o explicación adicional. Luego pasaba a explicar que esto implicaba que el resultado del trabajo (que era sobre los valores de ciertas funciones de Busseman en sucesiones aleatorias en ciertos espacios métricos) implicaba en particular el teorema de Oseledets.

Fué un gran dolor de cabeza tratar de entender esa frase. Y un gran placer cuando finalmente dimos con las referencias adecuadas (que además son altamente recomendables por estar muy bien escritas) que son [Lan99] y [Bha07].

4.1. Norma de Frobenius

Denotamos por $\mathbb{R}^{m\times n}$ las matrices de m filas y n columnas con entradas reales.

Si $A = (a_{ij}) \in \mathbb{R}^{m\times n}$ denotamos la matriz transpuesta por $A^* = (a_{ji}) \in \mathbb{R}^{n\times m}$.

Identificando $\mathbb{R}^n$ con las matrices de n filas y 1 columna, el producto interno usual en $\mathbb{R}^n$ está dado por:

$$\langle v, w\rangle = v^* w \ \forall v, w \in \mathbb{R}^n$$

Si $A = (a_{ij}) \in \mathbb{R}^{n\times n}$ denotamos la traza de A por $\mathrm{Tr}(A) = \sum_{1\le i\le n} a_{ii}$. Con esto se puede definir un producto interno en $\mathbb{R}^{m\times n}$ mediante:

$$\langle A, B\rangle = \mathrm{Tr}(A^* B) \ \forall A, B \in \mathbb{R}^{m\times n}$$

Puede verificarse mediante un cálculo directo que este producto interno coincide con el usual si identificamos $\mathbb{R}^{n\times n}$ con $\mathbb{R}^{n^2}$.

La norma que proviene de este producto interno es la llamada norma de Frobenius. Explícitamente si $A = (a_{ij})$ se tiene:

$$\|A\|^2 = \sum_{i,j} a_{ij}^2$$

Supongamos $A, B \in \mathbb{R}^{n\times n}$ y denotemos $A = (v_1, \ldots, v_n)^*$ donde $v_i^* \in \mathbb{R}^n$ es la i-ésima fila de A. Análogamente fijemos $B = (w_1, \ldots, w_n)$ donde $w_i \in \mathbb{R}^n$ es la i-ésima columna de B. Se tiene por lo anterior:

$$\|A\|^2 = \sum_i \|v_i\|^2$$

$$\|B\|^2 = \sum_j \|w_i\|^2$$

La igualdad $AB = (\langle v_i, w_j\rangle)$ y la desigualdad de Cauchy-Schwarz implican lo siguiente:

$$\|AB\|^2 = \sum_{1\le i,j\le n} |\langle v_i, w_j\rangle|^2 \le \sum_{1\le i,j\le n} \|v_i\|^2\|w_j\|^2 = \|A\|^2\|B\|^2$$

Por lo tanto hemos mostrado que:

$$\|AB\| \le \|A\|\|B\| \ \forall A, B \in \mathbb{R}^{n\times n}$$

Otra observación que utilizaremos sin mención es que el valor absoluto de cada entrada de una matriz A está acotada por $\|A\|$.

4.2. Mapa Exponencial

Definimos el mapa exponencial

$$\exp : \mathbb{R}^{d\times d} \to \mathbb{R}^{d\times d}$$

mediante la siguiente serie de potencias de matrices (con la convención $A^0 = I$ donde I es la matriz identidad)

$$\exp(A) = \sum_{n\ge 0} \frac{1}{n!} A^n$$

Para cada n la función $p_n(A) = \sum_{k=0}^{n} \frac{1}{k!} A^k$ es continua (simplemente porque $A \mapsto A^n$ es continua como función de las entradas de A) y la estimación:

$$\|p_n(A) - p_{n-1}(A)\| \le \frac{1}{n!}\|A\|^n$$

permite ver que p_n converge uniformemente en compactos cuando $n \to +\infty$ y por lo tanto exp es una función continua.

Tomemos $x = (x_1, \ldots, x_{d^2}) \in \mathbb{R}^{d^2}$ un vector que contiene las entradas de la matriz $A \in \mathbb{R}^{d\times d}$.

Definamos para cada multiíndice $k = (k_1, \ldots, k_{d^2}) \in \mathbb{Z}_+^{d^2}$ la norma de la suma $|k| = k_1 + \cdots + k_{d^2}$. Y utilicemos la notación:

$$x^k = x_1^{k_1} \cdots x_{d^2}^{k_{d^2}}$$

Cualquier entrada de A^n es un polinomio homogéneo de grado n en las d^2 coordenadas de x. Por lo tanto si $f(x)$ es una cierta entrada de la matriz $\exp(A)$ se tiene:

$$f(x) = \sum_{k \in \mathbb{Z}_+^{d^2}} a_k x^k$$

Para ciertos coeficientes $a_k \geq 0$. Además por las estimaciones anteriores esta serie es uniformemente absolutamente convergente en compactos de $\mathbb{R}^{d^2}$.

Mostraremos a continuación un resultado general sobre funciones analíticas reales que implica que existen las derivadas parciales de f de todos los ordenes y se calculan derivando la serie término a término. En particular la función exp es C^∞.

Teorema 4.1. *Supongamos que la serie*

$$\sum_{k \in \mathbb{Z}_+^N} a_k x^k$$

es uniformemente convergente en los compactos de $\mathbb{R}^N$*. Entonces define una función* C^∞ *cuyas derivadas parciales de todos los ordenes se calculan derivando la serie término a término.*

DEMOSTRACIÓN: Supongamos que cierta serie $f = \sum f_n$ de funciónes continuas converge uniformemente en compactos de $\mathbb{R}^N$ y definamos (la t se coloca en la i-ésima coordenada):

$$g(x) = \sum_n \int_o^{x_i} f_n(x_1, \ldots, t, \ldots, x_N) \mathrm{d}t \text{ donde } x = (x_1, \ldots, x_N)$$

La convergencia uniforme y el teorema fundamental del cálculo implican que $\frac{\partial}{\partial x_i} g = f$.

Por lo tanto para mostrar nuesto resultado alcanza con probar que la derivada formal respecto a x_i de la serie

$$\sum_{k \in \mathbb{Z}_+^N} a_k x^k$$

es uniformemente convergente en compactos para todo i (y luego proceder por inducción para derivadas de mayor orden).

Fijemos entonces $l = (0, \ldots, 1, \ldots, 0)$ un multi-índice con un 1 en la coordenada i. Utilicemos también la notación $(k+l)_i$ para la i-ésima coordenada de la suma coordenada a coordenada de los multiindices k y l.

En esta notación la derivada formal está dada por:

$$\sum_{k \in \mathbb{Z}_+^N} (k+l)_i a_{k+l} x^k$$

Necesitaremos la siguiente identidad:

$$\sum_{|k|=n} (k+l)_i = \frac{n+1}{N} \begin{pmatrix} n+N-1 \\ N-1 \end{pmatrix}$$

La identidad se deduce observando que la cantidad de N-úplas con $|k| = n$ está en biyección con las maneras de poner $N-1$ "separadores" en un total de $n+N-1$ lugares (e.g. $--/---//-$ donde hay 9 lugares y 3 separadores se corresponde con $k = (2, 3, 0, 1)$). De esto se deduce que

$$\sum_{|k|=n} \sum_j (k+l)_j = \sum_{|k|=n} n+1 = (n+1) \begin{pmatrix} n+N-1 \\ N-1 \end{pmatrix}$$

y por lo tanto si se suma únicamente la entrada i-ésima de cada N-upla se obtiene la N-ésima parte del total.

Probaremos ahora la convergencia uniforme en compactos de la serie derivada.

Fijemos $T > 0$. Las hipótesis implican que $\sum_{k \in \mathbb{Z}_+^N} a_k T^{|k|}$ es convergente y por lo tanto existe $C > 0$ tal que $|a_k| \leq C/T^{|k|}$ para todo k.

Para cierto $0 < \rho < 1$ fijo consideremos los $x = (x_1, \ldots, x_N) \in \mathbb{R}^N$ tales que $|x_i| \leq \rho T$ para todo i. La siguiente desigualdad demuestra la convergencia absoluta uniforme de la serie derivada en este compacto, y dado que es válida para todo T se deduce el enunciado:

$$\begin{aligned} \sum_{k \in \mathbb{Z}_+^N} |(k+l)_i a_{k+l} x^k| &\leq \sum_n \sum_{|k|=n} (k+l)_i \frac{C}{T^{|k|+1}} \rho^{|k|} T^{|k|} \\ &= \frac{C}{T} \sum_n \rho^n \sum_{|k|=n} (k+l)_i \\ &= \frac{C}{T} \sum_n \rho^n \frac{n+1}{N} \begin{pmatrix} n+N-1 \\ N-1 \end{pmatrix} < +\infty \end{aligned}$$

□

Las siguientes propiedades adicionales se deducen operando directamente con la serie que define la función exp:

1. $\exp(A^{-1}BA) = A^{-1}\exp(B)A$ para todo par $A, B \in \mathbb{R}^{n\times n}$ con A invertible.

2. Si $A, B \in \mathbb{R}^{n\times n}$ conmutan entonces $\exp(A+B) = \exp(A)\exp(B) = \exp(B)\exp(A)$.

Denotemos por $\mathrm{GL}(n) \subset \mathbb{R}^{n\times n}$ el subconjunto formado por las matrices invertibles. La primer propiedad implica que:

$$I = \exp(A - A) = \exp(A)\exp(-A) = \exp(-A)\exp(A)$$

y por lo tanto $\exp(A) \in \mathrm{GL}(n)$ para todo $A \in \mathbb{R}^{n\times n}$.

El diferencial del mapa exponencial en un punto A evaluado en una matriz B es por definición:

$$D_A \exp \cdot B = \lim_{t\to 0} \frac{\exp(A+tB) - \exp(A)}{t}$$

Si A y B conmutan se tiene

$$D_A \exp \cdot B = \exp(A)B$$

mientras que en el caso general (i.e. no conmutativo) se obtiene

$$D_A \exp \cdot B = B + \sum_{m\geq 2} \frac{1}{m!} \sum_{0\leq k\leq m-1} A^k B A^{m-1-k}$$

En particular $D_0 \exp : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ es la función identidad.

4.3. Descomposición Polar

Denotemos por $\mathrm{Sym}(n)$ el subconjunto de las matrices simétricas (i.e. que cumplen $A^* = A$) de $\mathbb{R}^{n\times n}$. Este subconjunto es un subespacio vectorial de dimensión $n(n+1)/2$.

Además usaremos $\mathrm{O}(n)$ para el conjunto de matrices ortogonales (i.e. que preservan el producto interno o equivalentemente son invertibles y cumplen $A^* = A^{-1}$).

El teorema espectral implica que si $A \in \mathrm{Sym}(n)$ entonces A se diagonaliza en una base ortonormal de $\mathbb{R}^n$. Equivalentemente existe $B \in \mathrm{O}(n)$ tal que $B^{-1}AB$ es diagonal.

Por último, denotamos por $\mathrm{Pos}(n) \subset \mathrm{GL}(n)$ el conjunto de matrices simétricas que además tienen todos sus valores propios positivos.

Por ejemplo idéntificando $\mathrm{Sym}(2)$ con $\mathbb{R}^3$ mediante el mapa:

$$\begin{pmatrix} a & c \\ c & b \end{pmatrix} \mapsto \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

El subconjunto $\mathrm{Pos}(2)$ queda idéntificado con la componente del cono abierto

$$\{(x, y, z)^* \in \mathbb{R}^3 : xy - z^2 > 0\}$$

que contiene a $(1, 1, 0)^*$.

En general $\mathrm{Pos}(n)$ es un subconjunto abierto en $\mathrm{Sym}(n)$ que es un "cono" en el sentido de que si $P, Q \in \mathrm{Pos}(n)$ se tiene $\alpha P \in \mathrm{Pos}(n)$ para todo $\alpha > 0$ y $P + Q \in \mathrm{Pos}(n)$.

El teorema espectral muestra que toda $P \in \mathrm{Pos}(n)$ tiene una única raiz cuadrada positiva (i.e. $Q \in \mathrm{Pos}(n)$ tal que $Q^2 = P$). También muestra que toda matriz $P \in \mathrm{Pos}(n)$ tiene un único logaritmo simétrico (i.e. $A \in \mathrm{Sym}(n)$ tal que $\exp(A) = P$).

Nuestro interés en el subconjunto $\mathrm{Pos}(n)$ es a causa del siguiente resultado clásico de álgebra lineal:

Teorema 4.2 (Teorema de Descomposición Polar)**.** *Si $A \in GL(n)$ entonces existen $P \in Pos(n)$ y $O \in O(n)$ tales que $A = OP$. Además esta descomposición es única.*

Demostración: Tomemos la matriz simétrica $P^2 = A^*A \in \mathrm{Sym}(n)$. Mostraremos que, de hecho, esta matriz es positiva. En efecto, sea $v \in \mathbb{R}^n$ un vector propio de P^2 se tiene

$$0 < \|Av\|^2 = \langle Av, Av\rangle = \langle P^2 v, v\rangle = \lambda\|v\|$$

de modo que todo valor propio de P^2 es positivo (en particular P^2 es invertible).

Tomemos entonces $P = (A^*A)^{\frac{1}{2}}$ la única raiz cuadrada positiva de A^*A (que existe gracias al teorema espectral). Se tiene que $O = AP^{-1}$ es ortogonal por el siguiente cálculo (que consiste únicamente en pasar matrices de un lado al otro del producto interno tomando transpuestas):

$$\|Ov\|^2 = \langle AP^{-1}v, AP^{-1}v\rangle = \|v\|^2$$

Hemos mostrado entonces la existencia de $P \in \mathrm{Pos}(n)$ y $O \in \mathrm{O}(n)$ tales que $A = OP$. La unicidad de la desomposición se deduce simplemente notando que si $A = O'P'$ se tiene $(P')^2 = A^*A$ y por lo tanto $P' = P$, además $O' = A(P')^{-1} = AP^{-1} = O$. □

El siguiente lema muestra la acción natural de las matrices invertibles sobre las positivas.

Lema 4.3. *Si $A \in GL(n)$ y $P \in Pos(n)$ entonces $A^*PA \in Pos(n)$.*

Demostración: Tomemos $Q \in \mathrm{Pos}(n)$ la única raiz cuadrada positiva de P. Se tiene que $A^*PA = (QA)^*QA$ es el cuadrado de la parte positiva de la descomposición polar de QA. □

El propósito de este capítulo es dotar a $\mathrm{Pos}(n)$ de una estructura riemanniana que hace que los mapas de la forma $P \mapsto A^*PA$ sean isometrías, y estudiar las propiedades de esta estructura (e.g. sus geodésicas y su curvatura).

4.4. Estructura Riemanniana en Pos(n)

En la discusión sobre la diferenciabilidad del mapa exponencial implícitamente hemos dado a $\mathbb{R}^{n\times n}$ la estructura diferencial que viene de identificarlo con $\mathbb{R}^{n^2}$.

Dado que $\mathrm{Sym}(n)$ es un subespacio de $\mathbb{R}^{n\times n}$ hereda una estructura diferencial, que coincide con la que obtenemos al identificarlo mediante cualquier base con $\mathbb{R}^{\frac{n(n+1)}{2}}$.

El conjunto de matrices positivas $\mathrm{Pos}(n)$ es un subconjunto abierto de $\mathrm{Sym}(n)$ [1] y por lo tanto tiene una estructura diferencial obvia en la cual el espacio tangente $T_P\mathrm{Pos}(n)$ en cualquier matriz positiva es $\mathrm{Sym}(n)$.

[1]Es la componente conexa que contiene a la matriz identidad de la intersección del subespacio $\mathrm{Sym}(n)$ con el abierto $\{A : \det(A) \neq 0\}$.

Una estructura riemanniana en $\mathrm{Pos}(n)$ consiste por lo tanto en dar un producto interno en $\mathrm{Sym}(n)$ para cada matriz $P \in \mathrm{Pos}(n)$. Utilizaremos el producto interno de Frobenius en el tangente a la matriz identidad:

$$\langle A, B\rangle_I = \langle A, B\rangle = \mathrm{Tr}(A^*B) = \mathrm{Tr}(AB)\ \forall A, B \in \mathrm{Sym}(n)$$

En el resto de las matrices positivas definiremos el producto de manera que sea invariante bajo la operación $P \mapsto QPQ$ donde $P, Q \in \mathrm{Pos}(n)$. Por lo tanto no queda otra opción más que definir:

$$\begin{aligned}\langle A, B\rangle_P &= \langle P^{-\frac{1}{2}}AP^{-\frac{1}{2}}, P^{-\frac{1}{2}}BP^{-\frac{1}{2}}\rangle_I \\ &= \mathrm{Tr}(P^{-\frac{1}{2}}AP^{-\frac{1}{2}}P^{-\frac{1}{2}}BP^{-\frac{1}{2}}) \\ &= \mathrm{Tr}(P^{-1}AP^{-1}B)\end{aligned}$$

Dado que las operaciones de multiplicar, tomar traza, y de invertir una matriz son funciones C^∞ de las matrices involucradas (esto último puede verse a partir de la expresión de la inversa en función de determinantes de submatrices) este producto interno también lo es.

Por lo tanto hemos dado una estructura riemanniana a $\mathrm{Pos}(n)$. Mostraremos ahora la propiedad de interés, con respecto a la forma polar de un producto de matrices:

Teorema 4.4. *Para toda matriz $A \in GL(n)$ el mapa $\rho_A : Pos(n) \to Pos(n)$ definido mediante:*

$$\rho_A(P) = A^*PA$$

es una isometría.

Además ρ_A es invertible y $(\rho_A)^{-1} = \rho_{A^{-1}}$.

Demostración: El mapa ρ_A puede extenderse con la misma definición a una transformación lineal de todo $\mathbb{R}^{n\times n}$. Por lo tanto su diferencial en todo punto es ρ_A. El siguiente cálculo verifica que efectivamente el diferencial de ρ_A preserva el producto interno.

$$\begin{aligned}\langle A^*BA, A^*CA\rangle_{A^*PA} &= \mathrm{Tr}((A^*PA)^{-1}A^*BA(A^*PA)^{-1}A^*CA) \\ &= \mathrm{Tr}(A^{-1}P^{-1}BP^{-1}CA) \\ &= \mathrm{Tr}(P^{-1}BP^{-1}C) \\ &= \langle B, C\rangle_P\end{aligned}$$

La verificación de que $(\rho_A)^{-1} = \rho_{A^{-1}}$ es trivial. $\square$

4.5. Geodésicas y Distancia

En el espacio vectorial $\mathrm{Sym}(n)$ tenemos el producto interno de Frobenius.

En $\mathrm{Pos}(n)$ hemos colocado el producto interno de Frobenius en el tangente a la matriz identidad, y luego hemos definido la métrica de modo que la operación $P \mapsto A^*PA$ sea una isometría para toda $A \in \mathrm{GL}(n)$.

Ya hemos mostrado que el mapa exponencial es C^∞ e invertible entre $\mathrm{Sym}(n)$ y $\mathrm{Pos}(n)$. Mostraremos a continuación que, de hecho, exp es un difeomorfismo.

Lema 4.5. *Si $A = (a_{ij}) \in Sym(n)$ es diagonal con valores propios $\lambda_i = a_{ii}$ entonces definiendo:*

$$C = D_A \exp \cdot B = (c_{ij})$$

se tiene

$$c_{ij} = \begin{cases} \frac{e^{\lambda_i} - e^{\lambda_j}}{\lambda_i - \lambda_j} b_{ij} & si\ \lambda_i \neq \lambda_j \\ e^{\lambda_i} b_{ij} & si\ \lambda_i = \lambda_j b_{ij} \end{cases}$$

Además para todo $A \in Sym(n)$ si elegimos $U \in O(n)$ y $A' \in Sym(n)$ diagonal de modo que $A = U^{-1}A'U$. En este caso se tiene para todo $B \in Sym(n)$:

$$D_A \exp \cdot B = U^{-1} D_{A'}(\exp \cdot B) U$$

En particular $D_A \exp$ es invertible para todo $A \in Sym(n)$ y por lo tanto exp *es un difeomorfismo.*

Demostración: Fijemos entonces $A = (a_{ij})$ con $a_{ii} = \lambda_i$ y $a_{ij} = 0$ si $i \neq j$. Para todo $B \in \mathrm{Sym}(n)$ se tiene:

$$D_A \exp \cdot B = B + \sum_{m \geq 2} \frac{1}{m!} \sum_{k=0}^{m-1} A^k B A^{m-1-k} = (c_{ij})$$

y por lo tanto:

$$\begin{aligned} c_{ij} &= b_{ij} + \sum_{m \geq 2} \frac{1}{m!} \sum_{k=0}^{m-1} \lambda_i^k b_{ij} \lambda_j^{m-1-k} \\ &= f(\lambda_i, \lambda_j) b_{ij} \end{aligned}$$

donde

$$f(x, y) = 1 + \sum_{m \geq 2} \frac{1}{m!} \sum_{k=0}^{m-1} x^k y^{m-1-k}$$

Es fácil verificar que $f(x,x) = e^x$ para todo $x \in \mathbb{R}$. Además se tiene:

$$(x-y)f(x,y) = (x-y) + \sum_{m\geq 2} \frac{1}{m!}(x^m - y^m) = e^x - e^y$$

Por lo tanto hemos demostrado la afirmación sobre c_{ij}.

En general el teorema espectral asegura que para todo $A \in \mathrm{Sym}(n)$ existen $A' \in \mathrm{Sym}(n)$ diagonal y $U \in \mathrm{O}(n)$ tales que $A = U^{-1}A'U$. Tomemos el mapa $\rho_U : \mathrm{Sym}(n) \to \mathrm{Sym}(n)$ definido mediante $\rho_U(S) = U^{-1}SU$. Se tiene $\rho_U \circ \exp = \exp \circ \rho_U$. Y dado que ρ_U es lineal la regla de la cadena demuestra la afirmación sobre $D_A \exp$.

La afimación de que exp es un difeomorfismo se sigue del teorema de la función inversa ya que mostramos que su diferencial es invertible en todo punto de $\mathrm{Sym}(n)$. Hacemos notar que el mismo teorema implica que la inversa tiene la misma diferenciabilidad que exp, es decir la función $\log : \mathrm{Pos}(n) \to \mathrm{Sym}(n)$ que asocia a cada matriz positiva su único logaritmo simétrico es C^∞. $\square$

El siguiente paso natural es utilizar el difeomorfismo $\exp : \mathrm{Sym}(n) \to \mathrm{Pos}(n)$ para comparar las estructuras métricas en ambos espacios.

Lema 4.6. *Si $A, B \in \mathrm{Sym}(n)$ conmutan entonces:*

$$\|D_A \exp \cdot B\|_{\exp(A)} = \|B\|$$

donde la norma sin subindice indica (como siempre) la norma de Frobenius.

En general para todo $A, B \in \mathrm{Sym}(n)$ se cumple:

$$\|D_A \exp \cdot B\|_{\exp(A)} \geq \|B\|$$

Demostración: En el caso conmutativo se tiene:

$$\begin{aligned}\|D_A \exp \cdot B\|_{\exp(A)} &= \|\exp(A)^{-\frac{1}{2}}(D_A \exp \cdot B)\exp(A)^{-\frac{1}{2}}\| \\ &= \|\exp(A)^{-\frac{1}{2}}\exp(A)B\exp(A)^{-\frac{1}{2}}\| \\ &= \|B\|\end{aligned}$$

Supongamos ahora que $A \in \mathrm{Sym}(n)$ es diagonal de valores propios $\lambda_i = a_{ii}$ y definamos

$$f(x,y) = \begin{cases} \frac{e^x - e^y}{x-y} & \text{si } x \neq y \\ e^x & \text{si } x = y \end{cases}$$

Hemos mostrado anteriormente que si $B = (b_{ij}) \in \mathrm{Sym}(n)$ se cumple que $D_A \exp \cdot B = (f(\lambda_i, \lambda_j) b_{ij})$ por lo tanto en este caso:

$$\begin{aligned}
\|D_A \exp \cdot B\|^2_{\exp(A)} &= \| \exp(A)^{-\frac{1}{2}} (D_A \exp \cdot B) \exp(A)^{-\frac{1}{2}} \|^2 \\
&= \| \exp(A)^{-\frac{1}{2}} \exp(A) B \exp(A)^{-\frac{1}{2}} \|^2 \\
&= \sum_{i,j} \left(e^{-\frac{1}{2}\lambda_i} f(\lambda_i, \lambda_j) b_{ij} e^{-\frac{1}{2}\lambda_j} \right)^2 \\
&= \sum_{i,j} (g(\lambda_i, \lambda_j) b_{ij})^2
\end{aligned}$$

Alcanza entonces con mostrar $g(x, y)^2 \geq 1$ para todo $x, y \in \mathbb{R}$.

Explícitamente se tiene:

$$g(x, y) = \begin{cases} \frac{e^{\frac{x-y}{2}} - e^{-\frac{x-y}{2}}}{x-y} & \text{si } x \neq y \\ 1 & \text{si } x = y \end{cases}$$

En el caso $x = y$ claramente $g(x, x)^2 \geq 1$. Por lo tanto alcanza discutir el caso $x \neq y$ en el cual $g(x, y) = \frac{\sinh((x-y)/2)}{(x-y)/2}$. En este caso se demuestra la desigualdad notando que por el teorema del valor medio $\frac{\sinh(t)}{t} = \sinh'(\theta t) = \cosh(\theta t)$ para cierto $\theta \in [0, 1]$. Y que $\cosh(t) \geq 1$ para todo $t \in \mathbb{R}$.

En el caso general escribimos $A = U^{-1} A' U$ con $U \in \mathrm{O}(n)$ y $A' \in \mathrm{Sym}(n)$ diagonal y el resultado se obtiene del caso anterior utilizando que el mapa $\rho_U : \mathrm{Pos}(n) \to \mathrm{Pos}(n)$ definido mediante $\rho_U(P) = U^{-1} P U = U^* P U$ es una isometría. □

Un corolario importante:

Corolario 4.7 (Comparación entre $\mathrm{Sym}(n)$ y $\mathrm{Pos}(n)$)**.** *Para todo $P, Q \in$ Pos(n) se tiene:*

$$d(P, Q) \geq \| \log(P) - \log(Q) \|$$

Demostración: Sea $\beta : [0, 1] \to \mathrm{Pos}(n)$ diferenciable y tal que $\beta(0) = P, \beta(1) = Q$. Definamos $\alpha : [0, 1] \to \mathrm{Sym}(n)$ mediante:

$$\alpha(t) = \log(\beta(t)) \ \forall t \in [0, 1]$$

El lema anterior implica que la longitud de β es mayor o igual a la de α. A su vez la longitud de α es mayor o igual a la distancia entre sus extremos que es exáctamente $\| \log(P) - \log(Q) \|$. □

A través de este corolario y las isometrías $P \mapsto A^*PA$ obtendremos una fórmula explícita para las geodésicas en $\mathrm{Pos}(n)$ y para la distancia entre dos puntos:

Corolario 4.8. *El mapa* $\exp : Sym(n) \to Pos(n)$ *es el mapa exponencial asociado a la métrica en* $Pos(n)$ *en el tangente de la matriz identidad.*

Si denotamos por $d(P,Q)$ *la distancia entre dos matriz positivas* P *y* Q *se tiene:*

$$d(P,I) = \| \log(P) \|$$

$$d(P,Q) = \| \log(Q^{-\frac{1}{2}} P Q^{-\frac{1}{2}}) \| = \| \log(P^{-\frac{1}{2}} Q P^{-\frac{1}{2}}) \|$$

La curva dada por $\alpha(t) = P^t$ *para* $t \in [0,1]$ *es la única geodésica (a menos de reparametrizaciones) entre la matriz identidad y* $P \in Pos(n)$*. En general para* $P, Q \in Pos(n)$ *la única geodésica de* P *a* Q *está dada por:*

$$\beta(t) = P^{\frac{1}{2}} \left(P^{-\frac{1}{2}} Q P^{-\frac{1}{2}} \right)^t P^{\frac{1}{2}} \text{ para } t \in [0,1]$$

En particular esto implica que la métrica en $Pos(n)$ *es completa.*

Demostración: Fijemos $A \in \mathrm{Sym}(n)$ y $P = \exp(A) \in \mathrm{Pos}(n)$. Dada una curva diferenciable $\beta : [0,1] \to \mathrm{Pos}(n)$ con $\beta(0) = I$ y $\beta(1) = P$ definimos $\alpha : [0,1] \to \mathrm{Sym}(n)$ mediante $\beta(t) = \exp(\alpha(t))$.

Cuando $\alpha(t) = tA$ se obtiene que $\beta([0,t])$ tiene longitud $t\|A\|$, y por el corolario 4.7 la curva β el la de menor longitud que une I con P^t para todo t. En particular β es geodésica.

Con esto hemos obtenido todas las geodésicas por la matriz identidad (ya que tenemos una para cada valor inicial de la derivada). Más explicitamente, la geodésica entre la matriz identidad y una matriz $P \in \mathrm{Pos}(n)$ es única a menos de reparametrizaciones y puede definirse mediante:

$$\beta(t) = \exp(t \log(P)) = P^t \text{ para todo } t \in [0,1]$$

Esto muestra la afirmación:

$$d(P,I) = \| \log(P) \|$$

Y también muestra que exp es el mapa exponencial de la métrica en la matriz identidad.

Para $P, Q \in \mathrm{Pos}(n)$ la geodésica y la distancia entre ellos se obtienen a partir de lo anterior utilizando la isometría $X \mapsto P^{\frac{1}{2}} X P^{\frac{1}{2}}$ que mapea la identidad a P.

La métrica es completa porque las geodésicas están definidas para todo tiempo. $\square$

4.6. Paralelogramos y Convexidad

En términos intuitivos, lo que hemos mostrado hasta ahora es que $\mathrm{Pos}(n)$ como espacio métrico, se obtiene del espacio vectorial con producto interno $\mathrm{Sym}(n)$ aplicando un estiramiento (dado por exp) en casi todas las direcciones (una excepción notable son las rectas por 0, que no son estiradas).

Es razonable entonces, imaginar a $\mathrm{Pos}(n)$ como un espacio hipérbolico en el sentido de que las distancias son mayores y en general "hay más espacio" que en un espacio euclideo. Esto puede demostrarse rigurosamente con varias definiciones formales distintas (por ejemplo puede mostrarse que en $\mathrm{Pos}(n)$ todas las curvaturas secionales son menores o iguales a cero).

Es claro que la hiperbolicidad que posee $\mathrm{Pos}(n)$ no es uniforme. En particular el subespacio de las matrices positivas diagonales es totalmente geodésico, e isométrico a las matrices simétricas diagonales (con el producto interno de Frobenius) a través de exp. Por lo tanto este subespacio tiene curvatura cero.

Mostraremos en esta seccion algunas propiedades métricas de $\mathrm{Pos}(n)$ que reflejan hiperbolicidad.

Obtuvimos, hasta el momento, suficientes isometrías de $\mathrm{Pos}(n)$ como para enviar cualquier punto a cualquier otro. Agregaremos una isometría más que puede definirse básicamente como una simetría central.

Definición 4.9. Fijado un punto $P \in \mathrm{Pos}(n)$ definimos la simetría central con centro P como la transformación que mapea $\alpha(t)$ a $\alpha(-t)$ para toda geodésica α con $\alpha(0) = P$.

Lema 4.10. *La simetría central con respecto a cualquier punto es una isometría.*

DEMOSTRACIÓN: Conjugando cualquier simetría central con una isometría que evía el centro a la matriz identidad, vemos que alcanza con demostrarlo para la simetría con respecto a la matriz idéntidad. Esta es la transformación dada por:

$$P \mapsto P^{-1} \ \forall P \in \mathrm{Pos}(n)$$

Para cada $P, Q \in \text{Pos}(n)$ fijando $A = P^{-\frac{1}{2}} Q P^{-\frac{1}{2}}$ se tiene:

$$d(P,Q) = \| \log(A)\| = \| -\log(A)\| = \| \log(A^{-1})\| = d(P^{-1}, Q^{-1})$$

□

Para seguir demostraremos un análogo a la ley del paralelogramo. Nos tomaremos la pequeña molestia de definir explícitamente los términos básicos a utilizar, aunque asumimos que todo lector provería definiciones equivalentes.

Definición 4.11 (Punto Medio)**.** El punto medio entre $A, B \in \text{Pos}(n)$ es el único $M \in \text{Pos}(n)$ tal que $d(A,M) = d(M,B) = \frac{1}{2}d(A,B)$.

Definición 4.12 (Paralelogramo)**.** Un parelogramo es una cuatro-úpla ordenada $(A_1, A_2, A_3, A_4) \in \text{Pos}(n)^4$ de vértices que cumple que el punto medio de A_1 y A_3, coincide con el de A_2 y A_4. Los lados son los segmentos geodésicos entre A_i y A_j donde $(i,j) \in \{(1,2),(2,3),(3,4),(4,1)\}$. Las diagonales son los segmentos entre A_1 y A_3, y entre A_2 y A_4. Dos lados son opuestos si no comparten ningún vértice.

El lema 4.10 muestra que los lados opuestos de un paralelogramo miden lo mismo.

La siguiente "Ley del semi-paralelogramo" muestra que los lados de un paralelogramo en $\text{Pos}(n)$ crecen más en relación a la longitud de las diagonales que en el caso euclideo.

Lema 4.13 (Ley del semi-paralelogramo)**.** *Para todo paralelogramo* (A,B,C,D) *en Pos*(n) *se cumple:*

$$d(A,C)^2 + d(B,D)^2 \leq d(A,B)^2 + d(B,C)^2 + d(C,D)^2 + d(D,A)^2$$

Demostración: Mediante una isometría llevamos el punto medio de las diagonales del paralelogramo a la matriz identidad. Por lo tanto nos reducimos al caso (P, Q, P^{-1}, Q^{-1}).

La desigualdad se obtiene utilizando la ley del paralelogramo en el espacio euclideo $\text{Sym}(n)$ y la comparación entre este espacio y $\text{Pos}(n)$ dado por 4.7 como sigue:

$$\begin{aligned} d(P,P^{-1})^2 + d(Q,Q^{-1})^2 &= 4\|\log(P)\|^2 + 4\|\log(Q)\|^2 \\ &= \|\log(P) - \log(Q)\|^2 + \|\log(Q) - \log(P^{-1})\|^2 \\ &+ \|\log(P^{-1}) - \log(Q^{-1})\|^2 + \|\log(Q^{-1}) - \log(P)\|^2 \\ &\leq d(P,Q)^2 + d(Q,P^{-1})^2 + d(P^{-1},Q^{-1})^2 + d(Q^{-1},P)^2 \end{aligned}$$

□

Siguiendo con las comparaciones con espacios euclideos demostraremos ahora una desigualdad de la paralela media para triangulos.

Definición 4.14. Un tríangulo en Pos(n) es un terna de puntos $A, B, C \in$ Pos(n) que se llamarán vértices. Los lados del tríangulo son los segmentos geodésicos entre los vértices.

Lema 4.15 (Desigualdad de la paralela media)**.** *Dado un triángulo $A, B, C \in$ Pos(n) si se toma M el punto medio de A y B, y N el punto medio de A y C se cumple:*

$$d(M,N) \leq \frac{1}{2} d(B,C)$$

DEMOSTRACIÓN: Completando un paralelogramo (A, C, B, C') donde C' es la imagen de C por la simetría central de centro M se obtiene utilizando la ley del semi-paralelogramo:

$$d(A,B)^2 + d(C,C')^2 \leq d(A,C)^2 + d(C,B)^2 + d(B,C')^2 + d(C',A)^2$$

Dado que los lados opuestos miden lo mismo y que M es punto medio del segmento CC':

$$d(A,B)^2 + 4d(C,M)^2 \leq 2d(A,C)^2 + 2d(B,C)^2$$

Tomando M' la imagen de M por la simetría central de centro N se tiene un paralelogramo (A, M, C, M') y análogamente se obtiene:

$$d(A,C)^2 + 4d(M,N)^2 \leq 2d(A,M)^2 + 2d(M,C)^2 = \frac{1}{2} d(A,B)^2 + 2d(M,C)^2$$

Substituyendo una desigualdad en la otra obtenemos:

$$4d(M,N)^2 \leq d(B,C)^2$$

□

La desigualdad de la paralela media puede reescribirse como sigue:

Lema 4.16 (Desigualdad de la paralela media)**.** *Para todo par de geodésicas $\alpha, \beta : \mathbb{R} \to Pos(n)$ tales que $\alpha(0) = \beta(0)$ se cumple:*

$$d(\alpha(\frac{1}{2}t), \beta(\frac{1}{2}t)) \leq \frac{1}{2} d(\alpha(t), \beta(t)) \ \forall t \in \mathbb{R}$$

Recordemos que una función $f : \mathbb{R} \to \mathbb{R}$ se dice que es convexa si para todo $a, b \in \mathbb{R}$ se cumple:

$$f(ta + (1-t)b) \leq tf(a) + (1-t)f(b) \ \forall t \in [0,1]$$

La desigualdad de la paralela media tiene como consecuencia la convexidad de la distancia entre geodésicas.

Dado que en el caso euclideo las distancia entre el tiempo t de dos geodésicas es lineal en t es razonable asumir que en un espacio hiperbólico sea "supra-lineal". Todo encaja, si observamos que toda función convexa es el supremo de numerables funciones lineales.

Lema 4.17 (Convexidad de la distancia geodésica)**.** *Para todo par de geodésicas* $\alpha, \beta : \mathbb{R} \to Pos(n)$ *la función* $f : \mathbb{R} \to \mathbb{R}$ *tal que:*

$$t \overset{f}{\mapsto} d(\alpha(t), \beta(t))$$

es convexa.

Demostración: Fijados $a, b \in \mathbb{R}$ y $c = \frac{1}{2}(a+b)$ comenzaremos mostrando que $f(c) \leq \frac{1}{2}(f(a) + f(b))$.

Para esto consideremos una geodésica γ tal que $\gamma(a) = \alpha(a)$ y $\gamma(b) = \beta(b)$. Por la desigualdad triangular y la de la paralela media tenemos:

$$\begin{aligned} f(c) = d(\alpha(c), \beta(c)) &\leq d(\alpha(c), \gamma(c)) + d(\gamma(c), \beta(c)) \\ &\leq \frac{1}{2} d(\alpha(b), \gamma(b)) + \frac{1}{2} d(\beta(a), \gamma(a)) = \frac{1}{2}(f(a) + f(b)) \end{aligned}$$

La demostración se completa con un hecho totalmente general: toda función continua que cumple $f(\frac{1}{2}(a+b)) \leq \frac{1}{2}(f(a) + f(b))$ para todo $a, b \in \mathbb{R}$ es convexa.

Dado que la hipótesis y la tesis de esta afirmación son invariantes pre o post componer f con una transformación afin $x \mapsto Ax + B$ con $A > 0$, alcanza con demostrar la afirmación en el caso $a = 0, b = 1, f(0) = 0$.

Lo que se quiere demostrar en este caso es $f(t) \leq f(1)t \ \forall t \in [0,1]$.

Para esto se utiliza inducción en n para demostrar que la afirmación es cierta para $t = \frac{k}{2^n}$ con $k \in \{0, \ldots, n\}$. Y luego el caso general se obtiene utilizando la continuidad. $\square$

Capítulo 5

Hiperbolicidad y Escoltas Geodésicas

En [Kaĭ89] Kaimanovich notó que el teorema de Oseledets sobre productos de matrices aleatorias puede interpretarse diciendo que la parte positiva del producto de las primeras n matrices está cerca de la potencia n-ésima de cierta matriz fija. Y que esto es una afirmación sobre la cercanía de una geodésica a la sucesión de partes positivas.

Con esto él fue capaz de dar una nueva demostración del teorema y extenderlo a otros espacios simétricos distintos de $\mathrm{Pos}(n)$.

En [KM99] A.Karlsson y G.Margulis encuentran una manera de demostrar el mismo teorema apelando únicamente a la distancia (en contraste con utilizar la estructura de variedad o métrica riemanniana) y esto abrió la puerta a demostrar el mismo tipo de teoremas sobre "escoltas geodésicas" para sucesiones aleatorias, pero en espacios métricos mucho más generales (e.g. se han usado estos resultados para demostrar resultados sobre paseos aleatorios y funciones armónicas en ciertos grupos).

En este capítulo aislamos el concepto de "escolta geodésica" de Kaimanovich, y mostramos con ejemplos por qué este concepto aparece vinculado a espacios hiperbólicos. En esta discusión la palabra "hiperbólico" básicamente significa algo así como de curvatura seccional menor o igual a cero, o CAT(0), pero no necesitaremos darle una defininción exacta.

Aprovechamos para dar algunos ejemplos especialmente sencillos en el disco hiperbólico y en el plano de los teoremas ergódicos que se pueden demostrar con estas técnicas. Hacemos notar que ésta es un área de investigación activa (ver por ejemplo [KL06]).

Utilizaremos algunos hechos de geometría en el disco hiperbólico que pueden encontrarse demostrados en [And05].

5.1. Escape Lineal y Geodésicas Escoltas

Empecemos con las definiciones.

Definición 5.1 (Escape Lineal)**.** Sea $\{x_n\}_{n\in\mathbb{Z}_+}$ una sucesión en un espacio métrico.

Se dice que la sucesión tiene velocidad de escape A si existe y es finito el siguiente límite:

$$A = \lim_n \frac{1}{n} d(x_0, x_n)$$

Si además se cumple que $A > 0$ entonces decimos que la sucesión se escapa linealmente a infinito.

Definición 5.2 (Geodésica Escolta)**.** En una variedad riemanniana: Dada una sucesión $\{x_n\}_{n\in\mathbb{Z}_+}$ decimos que es escoltada por una geodésica α si $\alpha(0) = x_0$ y existe $A \in [0, +\infty)$ tal que:

$$\frac{1}{n} d(\alpha(An), x_n) \to 0 \text{ cuando } n \to +\infty$$

Sucesiones que cumplen $\lim_n d(x_0, x_n)/n = 0$ son trivialmente escoltadas por cualquier geodésica α con $\alpha(0) = x_0$.

En el caso de sucesiones que se escapan linealmente a infinito tenemos la siguiente definición equivalente de geodésica escolta:

Lema 5.3. *Una sucesión $\{x_n\}_{n\in\mathbb{Z}_+}$ con velocidad de escape $A > 0$ es escoltada por una geodésica α parametrizada por longitud de arco si y sólo si $\alpha(0) = x_0$ y:*

$$\frac{1}{n} d(y_n, x_n) \to 0 \textit{ cuando } n \to +\infty$$

donde $y_n = \alpha(d(x_0, x_n))$.

La demostración consiste en usar la desigualdad triangular en el tríangulo $x_n, y_n, \alpha(An)$.

5.2. Escoltas en el disco de Poincaré

Sea $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\} \subset \mathbb{C}$ el disco de Poincaré con la métrica riemanniana usual dada por:

$$\langle v, w\rangle_z = \left(\frac{2}{1-|z|^2}\right)^2 \langle v, w\rangle \ \forall v, w \in \mathbb{R}^2, z \in \mathbb{D}$$

donde el producto interno del lado derecho es el usual en $\mathbb{R}^2$.

Se tiene entonces:

$$\|v\|_z = \frac{2}{1-|z|^2}\|v\| \ \forall v \in \mathbb{R}^2, z \in \mathbb{D}$$

Con esto se calcula fácilmente:

$$d(0,z) = \log\left(\frac{1+|z|}{1-|z|}\right) = \operatorname{arctanh}(|z|) \ \forall z \in \mathbb{D}$$

Utilizando las isometrías del disco de la forma $\phi_\alpha(z) = (z-\alpha)/(1-\bar{\alpha}z)$ un cálculo un poco más engorroso conduce a:

$$\frac{|z-w|^2}{(1-|z|^2)(1-|w|^2)} = \sinh\left(\frac{1}{2}d(z,w)\right)^2 \ \ \forall z, w \in \mathbb{D}$$

En particular cuando $|z| = |w|$ se obtiene:

$$d(z,w) = 2\operatorname{arcsinh}\left(\frac{|z-w|}{1-|z|^2}\right)$$

Una muestra de la hiperbolicidad de $\mathbb{D}$ es que el perímetro del círculo de radio r (i.e. $\{z \in \mathbb{D} : d(0,z) = r\}$) crece exponencialmente a medida que r tiende a $+\infty$. De hecho vale:

$$[\text{perímetro del círculo de radio } r] = 2\pi\sinh(r)$$

Intuitivamente el crecimiento exponencial de los perimetros de los círculos, implica que si una sucesión $\{z_n\}_{n\in\mathbb{Z}_+}$ con $z_0 = 0$ se escapa linealmente a infinito y no lo hace en una dirección dada (i.e. $z_n/|z_n|$ no tiene límite cuando $n \to +\infty$) debe tener saltos $d(z_n, z_{n+1})$ arbitrariamente grandes.

De hecho podemos permitir saltos de crecimiento sub-lineal en n y obtener el siguiente resultado:

Lema 5.4. *Si una sucesión $\{z_n\}_{n\in\mathbb{Z}_+} \subset \mathbb{D}$ se escapa linealmente a infinito y cumple $d(z_n, z_{n+1}) = o(n)$, entonces, es escoltada por una geodésica.*

Demostración: Podemos suponer aplicando una isometría que $z_0 = 0$. Sea $A > 0$ la velocidad de escape de la sucesión $\{z_n\}$.

Consideremos el tríangulo geodésico $0, z_n, z_{n+1}$ (notemos que dos de los lados son segmentos del plano) y denotemos por θ el ángulo en el vértice en 0. Tomemos:

$$x_n = \begin{cases} z_n & \text{si } |z_n| \leq |z_{n+1}| \\ z_{n+1} & \text{si } |z_n| > |z_{n+1}| \end{cases}$$

Definimos y_n tal que $|y_n| = |x_n|$ de la siguiente manera: Si $x_n = z_n$ tomamos y_n sobre el segmento entre 0 y z_{n+1} en caso contrario tomamos y_n sobre el segmento entre 0 y z_n.

Consideremos el triangulo isóceles formado por $0, x_n, y_n$. Denotemos $a = d(0, x_n) = d(0, y_n)$ y $b = d(x_n, y_n)$. Se tiene $b \leq d(z_n, z_{n+1})$.

La ley ley del coseno para triángulos hiperbólics (para una demostración ver [And05]) en este caso nos da:

$$\cos(\theta) = \frac{\cosh(a)^2 - \cosh(b)}{\sinh(a)^2}$$

Restando 1 a ambos lados obtenemos:

$$\cos(\theta) - 1 = \frac{1 - \cosh(b)}{\sinh(a)^2}$$

Utilizando que $\sinh(x) > \frac{e^x}{4}$ y que $\cosh(b) - 1 \leq e^x$ para x grande, obtenemos para n suficientemente grande:

$$1 - \cos(\theta) \leq 4e^{b-2a} \leq 4e^{-2An+o(n)}$$

Con esto ya se ve que $\theta \to 0$ cuando $n \to +\infty$ y por lo tanto para n suficientemente grande, podemos aplicar la desigualdad $x^2 \leq 4(1 - cos(x))$ válida para $x > 0$ suficientemente chico. Al hacerlo obtenemos lo siguiente (siempre para n suficientemente grande):

$$\theta \leq 4e^{-An+o(n)}$$

Esto implica que θ es sumable en n y por lo tanto existe $\lim_n z_n/|z_n|$. Sea α la geodésica parametrizada por longitud de arco tal que $\alpha(0) = 0$ y $\lim_n \alpha(n) = \lim_n z_n/|z_n|$. Mostraremos que α escolta a $\{z_n\}$.

Fijado $\epsilon > 0$ existe $N > 0$ de modo que para todo $n > N$ se cumple:

$$\theta < e^{-(A-\epsilon)n}$$

$$d(0, z_n) \leq (A + \epsilon)n$$

Esto implica que para todo $n > N$ se tiene que el ángulo ϕ entre el segmento $0, z_n$ y α cumple:

$$\phi \leq \frac{1}{1 - e^{-(A-\epsilon)}} e^{-(A-\epsilon)n} = Ce^{-(A-\epsilon)n}$$

Fijemos ahora $w_n = \alpha(d(0, z_n))$, $a = d(0, z_n) = d(0, w_n)$. Utilizando $|z_n| = |w_n|$ obtenemos para $n > N$:

$$\begin{aligned} d(z_n, w_n) &= 2 \operatorname{arcsinh}\left(\frac{|z_n - w_n|}{1 - |z_n|^2}\right) \leq 2 \operatorname{arcsinh}\left(\frac{Ce^{-(A-\epsilon)n}}{1 - \tanh(\frac{a}{2})^2}\right) \\ &= 2 \operatorname{arcsinh}\left(Ce^{-(A-\epsilon)n} \cosh(\frac{a}{2})^2\right) \leq 2 \operatorname{arcsinh}\left(Ce^{-(A-\epsilon)n+a}\right) \\ &\leq 2 \operatorname{arcsinh}\left(Ce^{-(A-\epsilon)n+(A+\epsilon)n}\right) = 2 \operatorname{arcsinh}\left(2Ce^{2\epsilon n}\right) \\ &\leq 2\log(2C) + 4\epsilon n \end{aligned}$$

En particular esto implica que $\limsup_n d(z_n, w_n)/n \leq 4\epsilon$. Como esto es válido para todo $\epsilon > 0$ obtenemos que α escolta a $\{z_n\}$. $\square$

Estamos (por fin) en condiciones de dar una muestra del objetivo central de esta tesis. Hemos definido en 3.4 lo que es un elemento aleatorio de un espacio topológico. Pongamos en el grupo de isometrías de $\mathbb{D}$ la topología de la convergencia uniforme en compactos. El siguiente resultado significa que si elegimos isometrías al azar y aplicamos la composición de ellas a un punto en $\mathbb{D}$, casi seguramente, la sucesión que se obtiene es escoltada por una geodésica (hay una hipótesis técnica que dice que no elegimos isometrías demasiado grandes).

Teorema 5.5 (Paseos al Azar en $\mathbb{D}$)**.** *Si $\psi_1, \ldots, \psi_n, \ldots$ es una sucesión estacionaria de isometrías aleatorias de $\mathbb{D}$ que cumplen:*

$$\mathbb{E}\left(d(0, \psi_n(0))\right) < +\infty$$

entonces definiendo $z_n = (\psi_1 \circ \cdots \circ \psi_n)(0)$ *se obtiene:*

$$\mathbb{P}\left(\frac{d(0,z_n)}{n} \text{ tiene límite cuando } n \to +\infty\right) = 1$$

$$\mathbb{P}\left(\{z_n\} \text{ es escoltada por una geodésica}\right) = 1$$

Demostración: Observemos que si $m < n$ se cumple que:

$$\begin{aligned} d(z_m, z_n) &= d((\psi_1 \circ \cdots \circ \psi_m)(0), (\psi_1 \circ \cdots \circ \psi_n)(0)) \\ &= d(0, (\psi_{m+1} \circ \cdots \circ \psi_n)(0)) \end{aligned}$$

Esto, junto con la estacionariedad de las ψ_n implican que a definido mediante:

$$a(m,n) = \begin{cases} d(z_m, z_n) & \text{si } m \leq n \\ -d(z_m, z_n) & \text{si } m > n \end{cases}$$

es un proceso subaditivo estacionario.

Además se tiene:

$$\mathbb{E}(a(0,1)) = \mathbb{E}(d(0, \psi_1(0))) < +\infty$$

y esto junto a la subaditividad implica $\mathbb{E}(|a(m,n)|) < +\infty$ para todo $m, n \in \mathbb{Z}_+$.

De modo que estamos en hipótesis del Teorema ergódico de Kingman 3.7 que asegura la existencia del límite:

$$\lim_n \frac{1}{n} d(0, z_n) = A(\omega)$$

para casi todo $\omega \in \Omega$ donde $(\Omega, \mathcal{F}, \mathbb{P})$ es el espacio de probabilidad del que provienen las ψ_n.

En aquellos puntos en los que $A = 0$ la sucesión es escoltada trivialmente.

Para el resto, observamos que por el lema 3.9 existe un conjunto de probabilidad 1 en el cual $d(z_n, z_{n+1}) = o(n)$. Y por lo tanto podemos aplicar el lema 5.4 y obtenemos el resultado. □

El lema 5.4 tiene también una aplicación dinámica.

Teorema 5.6 (Vector de Rotación). *Sea* Γ *un subgrupo de isometrías de* $\mathbb{D}$ *que actúa discontinuamente, de modo que* $M = \mathbb{D}/\Gamma$ *es una superficie compacta.*

Sea $f : M \to M$ un difeomorfismo isotópico a la identidad que preserva una medida de probabilidad boreliana μ.

Fijemos $F : \mathbb{D} \to \mathbb{D}$ el levantado de f que también es isotópico a la identidad, y fijemos para cada $x \in M$ un levantado arbitrario $\tilde{x} \in \mathbb{D}$.

Para μ casi todo x la sucesión $\{F^n(\tilde{x})\}_{n\in\mathbb{Z}_+}$ es escoltada por una geodésica.

Demostración: Notemos que, como f es isotópica a la identidad, F conmuta con todo elemento de Γ y por lo tanto se tiene que $d(\tilde{x}, F\tilde{x})$ no depende del levantado de x.

Definamos para cada $x \in M$:

$$a(m,n) = \begin{cases} d(F^m\tilde{x}, F^n\tilde{x}) & \text{si } m \leq n \\ -d(F^m\tilde{x}, F^n\tilde{x}) & \text{si } m > n \end{cases}$$

Esto es una distancia dirigida para cada $x \in M$ y dado que $(M, \mathcal{B}(M), \mu)$ es un espacio de probabilidad (la σ-álgebra es la de Borel) hemos definido un proceso subaditivo.

Para cada $n \in \mathbb{N}$, y cada familia de borelianos $\{A_{ij}\}_{0\leq i,j\leq n}$ se cumple:

$$\begin{aligned}
\mu(\{x \in M : a(i,j) \in A_{ij}\ \forall 0 \leq i,j \leq n\}) &= \mu(f^{-1}\{x \in M : a(i,j) \in A_{ij}\ \forall 0 \leq i,j \leq n\}) \\
&= \mu(\{x \in M : d(F^i\widetilde{f(x)}, F^j\widetilde{f(x)}) \in A_{ij}\ \forall 0 \leq i,j \leq n\}) \\
&= \mu(\{x \in M : d(F^{i+1}\tilde{x}, F^{j+1}\tilde{x}) \in A_{ij}\ \forall 0 \leq i,j \leq n\}) \\
&= \mu(\{x \in M : a(i+1,j+1) \in A_{ij}\ \forall 0 \leq i,j \leq n\})
\end{aligned}$$

lo cual muestra que a es un proceso estacionario (en penúltima igualdad utilizamos que f es isotópica a la identidad).

Además dado que $a(m,n)$ es acotada para cada $m, n \in \mathbb{Z}_+$ estamos en condiciones de utilizar el Teorema Ergódico de Kingman 3.7, y obtenemos que el límite:

$$\lim_n \frac{1}{n} d(\tilde{x}, F^n\tilde{x}) = A(x)$$

existe para μ-casi todo x.

Por último se tiene:

$$\begin{aligned}
\int d(F^n\tilde{x}, F^{n+1}\tilde{x})\mathrm{d}\mu(x) &= \int d(\widetilde{f^n x}, F\widetilde{f^n x})\mathrm{d}\mu(x) \\
&= \int d(\tilde{x}, F\tilde{x})\mathrm{d}\mu(x) < +\infty
\end{aligned}$$

de modo que por el lema 3.9 para μ casi todo $x \in M$ se cumple $d(F^n\tilde{x}, F^{n+1}\tilde{x}) = o(n)$.

En esas condiciones podemos aplicar el lema 5.4 y obtener el resultado. □

5.3. Escoltas en el plano

Consideramos ahora sucesiones en el plano $\mathbb{C}$ con la distancia usual $d(x, y) = |x - y|$.

Desde el punto de vista geométrico el hecho de que el perímetro del círculo de radio r sea una función lineal de r (en lugar de una función con crecimiento exponencial como en $\mathbb{D}$) hace que sea razonable pensar que en el caso plano es más dificil lograr la existencia de geodésicas escoltas.

De hecho, la sucesión dada por

$$x_n = e^{i\log(n)}n$$

se escapa a infinito a velocidad 1 y además cumple $d(x_n, x_{n+1})/n \to 0$ pero sin embargo no es escoltada por ninguna geodésica. Esto muestra que el lema 5.4 no es válido en el plano.

Para garantizar la existencia de una geodésica escolta es necesario utilizar alguna hipótesis que controle cuanto puede variar el ángulo que forma x_n con x_m. Pero a su vez si buscamos demostrar un teorema ergodico análogo a 5.5 con métodos similares a los ya usados, la condición debe darse únicamente en términos de las distancias relativas entre los puntos de la sucesión.

Definición 5.7 (ϵ-ángulo)**.** Sea X un espacio métrico. Decimos que $y \in X$ está en el ϵ-ángulo de x a z para cierto par $x, z \in X$ si se cumple:

$$e^{-\epsilon}d(x, y) + d(y, z) \leq d(x, z)$$

Denotemos el ϵ-ángulo entre x y z mediante:

$$[x, z]_\epsilon = \{y \in X : e^{-\epsilon}d(x, y) + d(x, z) \leq d(y, z)\}$$

Observamos que la definición no es simétrica en x y z.

Para $x, z \in \mathbb{C}$, el 0-ángulo entre ellos es un segmento, mientras que el $+\infty$-ángulo es un disco centrado en z, explícitamente:

$$[x, z]_0 = [x, z] = \{tx + (1 - t)z : t \in [0, 1]\}$$

$$[x, z]_{+\infty} = \{y \in \mathbb{C} : |z - y| \leq |x - z|\}$$

Además las homotecias del plano envían ϵ-ángulos en ϵ-ángulos. Más explícitamente, si tomamos $\lambda > 0$ y $x, z \in \mathbb{C}$ se tiene:

$$[\lambda x, \lambda z]_\epsilon = \lambda[x, z]_\epsilon = \{\lambda y : y \in [x, z]_\epsilon\}$$

$$[0, \lambda x]_\epsilon = \lambda[0, x]_\epsilon = \{\lambda y : y \in [0, x]_\epsilon\}$$

Claramente las isometrías del plano preservan los ϵ-ángulos, así que alcanza con entender los ϵ-ángulos entre 0 y 1 para entender todos los posibles en $\mathbb{C}$.

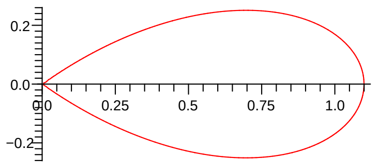


Figura 5.1: El borde de $[0, 1]_{0,2}$, es decir el 0,2-ángulo entre 0 y 1 en $\mathbb{C}$.

El siguiente lema, muestra que si $y \in [x, z]_\epsilon$ y $\epsilon > 0$ es chico, entonces el ángulo entre los segmentos $[x, y]$ y $[x, z]$ es chico. Utilizamos $\langle z, w\rangle = \mathrm{Re}(z\bar{w})$ para el producto interno que se obtiene identificando $\mathbb{C}$ con $\mathbb{R}^2$.

Lema 5.8. *Si $x, y, z \in \mathbb{C}$ cumplen $z \in [x, y]_\epsilon$ entonces:*

$$\langle z - x, y - x\rangle \geq 0$$

$$|\langle z - x, i(y - x)\rangle| \leq \sqrt{e^{2\epsilon} - 1}\langle z - x, y - x\rangle$$

Demostración: Podemos asumir $x = 0, y = 1$.

En este caso tenemos $e^{-\epsilon}|z| + |1 - z| \leq 1$. Utilizando la desigualdad $1 - \mathrm{Re}(z) \leq |1 - z|$ se obtiene:

$$e^{-\epsilon}|z| \leq \mathrm{Re}(z)$$

esto implica que $\mathrm{Re}(z) \geq 0$. Además elevando al cuadrado se obtiene:

$$\mathrm{Im}(z)^2 + \mathrm{Re}(z)^2 \leq e^{2\epsilon}\mathrm{Re}(z)^2$$

lo cual implica fácilmente el resultado. $\square$

Un poco de geometría analítica aplicada al estudio de los triangulos iscóceles, lleva al siguiente corolario que claramente nos acerca a demostrar un resultado sobre escoltas.

Corolario 5.9. *Si $x, y, z \in \mathbb{C}$ cumplen $z \in [x, y]_\epsilon$ y w sobre la geodésica que va de x a y cumple $d(w, x) = d(z, x)$ entonces:*

$$d(w, z)^2 \leq 2(e^\epsilon - 1)d(x, z)^2$$

DEMOSTRACIÓN: Aplicando una isometría nos reducimos al caso $x = 0, y \in [0, +\infty)$ con $\mathrm{Im}(z) \geq 0$. En este caso tenemos $w = |z|$.

El caso $z = 0$ es obvio.

En el caso restante el lema 5.8 nos asegura que $\mathrm{Re}(z) > 0$ y $0 \leq \mathrm{Im}(z)/\mathrm{Re}(z) = \lambda \leq \sqrt{e^{2\epsilon} - 1}$. Por lo tanto:

$$z = \frac{|z|}{\sqrt{1 + \lambda^2}} + i\frac{\lambda|z|}{\sqrt{1 + \lambda^2}}$$

Y calculando a fuerza bruta se obtiene:

$$|z - w|^2 = \frac{2}{\sqrt{1 + \lambda^2}}(\sqrt{1 + \lambda^2} - 1)|z|^2 \leq 2(\sqrt{1 + \lambda^2} - 1)|z|^2 \leq 2(e^\epsilon - 1)|z|^2$$

$\square$

Vimos hasta ahora que si $x, y, z \in \mathbb{C}$ cumplen $z \in [x, y]_\epsilon$ se tiene un control sobre el ángulo entre $z - x$ y $y - x$, y también se tiene un control sobre la distancia entre z y $w = x + |z - x|(y - x)/|y - x|$.

El control obtenido es bueno, porque las cotas tienden a cero cuando ϵ lo hace.

Definición 5.10 (Sucesión ϵ-alineada, Sucesión alineada)**.** Sea X un espacio métrico. Una sucesión $x_0, \ldots, x_n, \ldots \in X$ se dice que está ϵ-alineada si existe un valor $K \in \mathbb{N}$ e infinitos valores de $n \in \mathbb{N}$ tales que:

$$x_k \in [x_0, x_n]_\epsilon \ \forall K \leq k \leq n$$

Decimos que una sucesión está alineada si está ϵ-alineada para todo $\epsilon > 0$.

La importancia de esta definición viene dada básicamente por dos razones.

La primera es que el lema de Karlsson-Margulis 3.12 permite mostrar que una gran cantidad de sucesiones que se obtienen en forma aleatoria están alineadas.

La segunda razón es que podemos mostrar que en muchos espacios las sucesiones alineadas son escoltadas por una geodésica.

El siguiente lema muestra el vínculo entre alineamiento y escoltas geodésicas en el caso plano (utilizaremos el hecho de que una sucesión $\{x_n\}$ en el plano es escoltada por una geodésica si y sólo si $\frac{1}{n}x_n$ tiene límite[1]):

Lema 5.11. *Sea* $\{x_n\}_{n\in\mathbb{Z}_+} \subset \mathbb{C}$ *una sucesión que se escapa linealmente a infinito.*

Si $\{x_n\}$ *está alineada entonces es escoltada por una geodésica.*

DEMOSTRACIÓN: Aplicando una isometría podemos suponer $x_0 = 0$.

Fijado $\epsilon > 0$ existe $K \in \mathbb{N}$ y una sucesión $n_i \to +\infty$ tal que $x_k \in [0, x_{n_i}]_\epsilon$ para todo $K \leq k \leq n_i$. Tomemos para cada uno de estos valores de n_i la geodésica $\alpha_i : [0, +\infty) \to \mathbb{C}$ parametrizada por longitud de arco, tal que, $\alpha_i(0) = 0$ y $\alpha_i(|x_{n_i}|) = x_{n_i}$. Se tiene por el lema 5.9 que:

$$d(\alpha_i(|x_k|), x_k) \leq f(\epsilon)|x_k| \ \forall K \leq k \leq n_i$$

donde $f(\epsilon) = \sqrt{2(e^\epsilon - 1)}$.

Tomamos una subsucesión convergente puntualmente de α_i (alcanza con tomarla para que sea convergente en un valor $|x_k|$). El límite es una cierta geodésica α que cumple $\alpha(0) = 0$ y:

$$d(\alpha(|x_k|), x_k) \leq f(\epsilon)|x_k| \ \forall K \leq k$$

Esto implica que:

$$d(\alpha(1), x_k/|x_k|) < f(\epsilon) \ \forall k \geq K$$

En particular se tiene:

$$d(\frac{x_k}{|x_k|}, \frac{x_{k'}}{|x_{k'}|}) \leq 2f(\epsilon) \ \forall k, k' \geq K$$

[1] Esto sale directamente de la definición al conocer que las geodésicas son de la forma $t \mapsto x + ty$.

El razonamiento anterior y el hecho que $\lim_{\epsilon\to 0} f(\epsilon) = 0$, muestran que $x_n/|x_n|$ es de Cauchy y por lo tanto tiene límite.

De lo anterior se deduce:

$$\lim_n \frac{1}{n} x_n = \lim_n \frac{|x_n|}{n} \frac{x_n}{|x_n|} = A \lim_n \frac{x_n}{|x_n|}$$

donde A es la velocidad de escape de $\{x_n\}$. Por lo tanto $\{x_n\}$ es escoltada por una geodésica. □

Como aplicación de este resultado demostraremos el resultado análogo a 5.5[2].

Teorema 5.12 (Dinámica de Isometrías en $\mathbb{C}$)**.** *Si $\psi_1, \ldots, \psi_n, \ldots$ es una sucesión estacionaria de isometrías aleatorias de $\mathbb{C}$ que cumplen:*

$$\mathbb{E}\left(d(0, \psi_n(0))\right) < +\infty$$

entonces definiendo $z_n = (\psi_1 \circ \cdots \circ \psi_n)(0)$ se obtiene:

$$\frac{1}{n} z_n \text{ tiene límite cuando } n \to +\infty$$

Demostración: Definimos para $m \leq n$

$$a(m, n) = d(z_m, z_n)$$

y extendemos cambiando de signo al caso $m > n$.

De este modo obtuvimos un proceso subaditivo. Se muestra que este proceso es estacionario y tiene esperanza finita con el mismo argumento usado en 5.5.

Mostraremos en 5.14 que esto implica que con probabilidad 1 o bien $d(z_0, z_n)/n \to 0$ o la sucesión está alineada.

Asumiendo esto por el lema 5.11 se obtiene el resultado. □

[2] El resultado análogo a 5.6 puede deducirse usando únicamente el teorema ergódico de Birkhoff.

5.4. Alineamiento de sucesiones aleatorias

Lo primero que haremos en esta sección es reducir la condición de que una sucesión esté alineada a la afirmación de que cierto número es mayor o igual que la velocidad de escape de la sucesión.

Luego, vamos a utilizar el Teorema de Kingman 3.17 y el lema de Karlsson-Margulis 3.12 para demostrar que ciertas sucesiones aleatorias están alineadas casi-seguramente.

Lema 5.13. *Sea $\{x_n\}_{n\in\mathbb{Z}_+}$ una sucesión en un espacio métrico que se escapa linealmente a infinito con velocidad de escape $A > 0$.*

Si se cumple:

$$L = \lim_K \limsup_n \min_{K\leq k\leq n}\{\frac{d(x_0,x_n)-d(x_n,x_k)}{k}\} \geq A$$

entonces $\{x_n\}$ está alineada.

Demostración: Notemos que el límite L existe porque la función de K involucrada es creciente.

Supongamos que $L \geq A$. Por la definición de L se tiene que fijado $\epsilon > 0$, existe $K \in \mathbb{N}$ tal que:

$$\limsup_n \min_{K\leq k\leq n}\{\frac{d(x_0,x_n)-d(x_n,x_k)}{k}\} > e^{-\epsilon}A$$

Además (por la definición de A) podemos elegir K de modo que:

$$d(x_0,x_k) \leq Ae^{\epsilon}k \ \forall k \geq K$$

Juntando ambas cosas se tiene que para infinitos $n \in \mathbb{N}$ se cumple:

$$d(x_0,x_n) - d(x_n,x_k) > ke^{-\epsilon}A \geq e^{-2\epsilon}d(x_0,x_k) \ \forall K \leq k \leq n$$

de modo que:

$$x_k \in [x_0,x_n]_{2\epsilon} \ \forall K \leq k \leq n$$

En otras palabras $\{x_n\}$ está 2ϵ-alineada. Como esto es verdad para todo $\epsilon > 0$ se obtiene el resultado. □

Teorema 5.14 (Alineamiento de Sucesiones Aleatorias)**.** *Sea $x_0, \ldots, x_n, \ldots$ una sucesión de elementos aleatorios de un espacio métrico X tal que:*

$$a(m,n) = d(x_m, x_n)\ \forall m \leq n$$

es un proceso subaditivo estacionario con $\mathbb{E}(|a(m,n)|) < +\infty$ para todo $m, n \in \mathbb{Z}_+$.

Entonces casi seguramente, existirá el siguiente límite:

$$\lim_{n\to+\infty} \frac{d(x_0, x_n)}{n}$$

y cuando es positivo la sucesión estará alineada.

Demostración: Podemos suponer (tomando la transformación shift en $\mathbb{R}^{\mathbb{Z}_+^2}$, la distribución del proceso a como medida de probabilidad, y f_n la evaluación en el par $(0,n)$) que hay un sistema dinámico que preserva probabilidad $(\Omega, \mathcal{F}, \mathbb{P}, T)$ y una sucesión de funciones medibles $f_n : \Omega \to \mathbb{R}$ tales que:

$$a(m,n) = f_{n-m} \circ T^m\ \forall m < n, m, n \in \mathbb{Z}$$

Consideremos la función invariante $A : \Omega \to \mathbb{R}$ dada por el lema 3.16 que cumple:

$$\mathbb{E}(A 1_E) = \lim_n \frac{1}{n}\mathbb{E}(a(0,n)1_E)$$

para todo conjunto invariante E. Por el teorema de kingman 3.17 se tiene $A = \lim_n \frac{1}{n} a(0,n)$ casi seguramente.

Definamos $L : \mathbb{R}^{\mathbb{Z}_+^2} \to [0, +\infty]$ mediante:

$$L(x) = \max(0, \lim_K \limsup_n \min_{K \leq k \leq n} \{\frac{x(0,n) + x(n,k)}{k}\})$$

El lema 5.13 implica que nos alcanza con demostrar $\mathbb{P}(L(a) \geq A) = 1$.

Supongamos que $L(a) \geq t > 0$ en cierto punto de Ω. Esto implica que para todo $\epsilon > 0$ existe K e infinitos n tales que:

$$a(0,n) + a(n,k) > te^{-\epsilon}k\ \forall K \leq k \leq n$$

Agrandando K un poco podemos asegurar:

$$\begin{aligned} a(1,n) + a(n,k) &\geq a(1,0) + a(0,n) + a(n,k) \\ &> a(1,0) + te^{-\epsilon}k > te^{-2\epsilon}(k-1) \forall K \leq k \leq n \end{aligned}$$

Lo cual implica $L(a) \circ T \geq te^{-2\epsilon}$.

Dado que este razonamiento es válido para todo $\epsilon > 0$ hemos demostrado que $\{L(a) \geq t\} \subset \{L(a) \circ T \geq t\}$ para todo $t \in \mathbb{R}$ (para $t \leq 0$ es trivial). Pero la estacionariedad de a (o lo que es lo mismo, el hecho de que T deja invariante a la medida) implica que ambos conjuntos miden lo mismo y por lo tanto coinciden a menos de un conjunto de probabilidad nula. Como esto es válido para todo $t \in \mathbb{R}$ tenemos $\mathbb{P}(L(a) \circ T = L(a)) = 1$ y por lo tanto podemos asumir que $L(a)$ es invariante.

Supongamos por absurdo que para cierto $\epsilon > 0$ el conjunto invariante $E = \{L(a) < e^{-\epsilon}A\}$ tiene probabilidad positiva (notemos que en este conjunto $A > 0$).

Tomemos el proceso $b(m,n) = (a(m,n) - (n-m)e^{-\epsilon}A)1_E$. Este proceso cumple:

$$\lim_n \frac{1}{n}\mathbb{E}(b(0,n)) = (1 - e^{-\epsilon})\mathbb{E}(A1_E) > 0$$

Por lo tanto aplicando el lema de Karlsson-Margulis 3.12 se obtiene:

$$\mathbb{P}(\text{ para infinitos } n \text{ se cumple } b(0,n) + b(n,k) > 0 \text{ para todo } k \in \{1,\ldots,n\}) > 0$$

Dado que $b(0,n) + b(n,k) = (a(0,n) + a(n,k))1_E - e^{-\epsilon}Ak$ esto implica:

$$\mathbb{P}(\{L(a) \geq e^{-\epsilon}A\} \cap E) > 0$$

contrario a la definición de E.

Por lo tanto $\mathbb{P}(L(a) \geq A) = 1$ como se buscaba. $\square$

5.5. Escoltas en Pos(n)

Como último ejemplo de los conceptos de alineamiento y escoltas, trabajaremos con el espacio de matrices Pos(n). Mostraremos que, también en este espacio, las sucesiones alineadas que se escapan linealmente a infinito son escoltadas por una geodésica.

Lema 5.15. *Sean $x, y, z \in Pos(n)$ tales que $z \in [x,y]_\epsilon$.*

Y fijemos $w = \alpha(d(x,z))$, donde α es la geodésica parametrizada por longitud de arco que cumple $\alpha(0) = x$ y $\alpha(d(x,y)) = y$.

Entonces se cumple

$$d(w,z)^2 \leq 4(1 - e^{-2\epsilon})d(x,z)^2$$

Demostración: Fijamos $A = d(x,z) = d(x,w), B = d(x,y) - A, C = d(z,y)$. Hacemos notar que $B = d(w,y)$ si $d(x,z) < d(x,y)$ y $B = -d(w,y)$ en caso contrario. En ambos casos $|B| \leq C$ (en el primero porque $d(x,y) = A + |B| \leq A + C = d(x,z) + d(z,y)$, y en el segundo porque $C = d(z,y) \geq d(z,x) - d(x,y) = A - (A - |B|)$).

Con la notación anterior la condición $z \in [x,y]_\epsilon$ equivale a:

$$e^{-\epsilon}A + C \leq A + B$$

o lo que es lo mismo

$$A + B - C \geq e^{-\epsilon}A$$

Sea m el punto medio del segmento $[z,w]$. La convexidad de la distancia entre geodésicas dada por 4.17 (el caso límite cuando una de las dos es constante) nos da:

$$d(y,m) \leq \text{máx}(|B|, C) = C$$

lo cual implica

$$d(x,m) \geq d(x,y) - d(y,m) \geq A + B - C \geq e^{-\epsilon}A$$

Ahora completando un paralelogramo x, w, x', z (con x' el punto simétrico de x respecto a m), la ley del semi-paralelogramo 4.13 da:

$$d(w,z)^2 + d(x,x')^2 = d(w,z)^2 + 4d(x,m)^2 \leq 4A^2$$

por lo tanto se obtiene:

$$d(w,z)^2 \leq 4(A^2 - d(x,m)^2) \leq 4(1 - e^{-2\epsilon})A^2$$

como se buscaba. □

Demostraremos ahora la correspondencia entre alineación y existencia de una escolta.

Lema 5.16. *Sea $\{x_n\}_{n \in \mathbb{Z}_+}$ una sucesión de matrices positivas que se escapa linealmente a infinito a velocidad $A > 0$.*

Si $\{x_n\}$ está alineada entonces es escoltada por una geodésica.

Demostración: Fijemos $f : [0, +\infty) \to [0, +\infty)$ dada por $f(\epsilon) = \sqrt{4(1 - e^{-2\epsilon})}$.

Para cada $\epsilon > 0$ existe $K_\epsilon \in \mathbb{N}$ e infinitos n tales que $x_k \in [x_0, x_n]_\epsilon$ para todo $K_\epsilon \leq k \leq n$. Tomando la geodésica α_n tal que $\alpha_n(0) = 0$ y $\alpha_n(d(x_0, x_n)) = x_n$ se tiene por 5.15 que:

$$d(x_k, \alpha_n(d(x_0, x_k))) \leq f(\epsilon)d(x_0, x_k)\ \forall K_\epsilon \leq k \leq n$$

Tomando límite de α_n en una subsucesión de valores de n en los que se cumple lo anterior[3] se obtiene una geodésica α_ϵ parametrizada por longitud de arco que cumple $\alpha(0) = x_0$ y tal que:

$$d(x_k, \alpha_\epsilon(d(x_0, x_k))) \leq f(\epsilon)d(x_0, x_k)\ \forall k \geq K_\epsilon$$

Dados $\epsilon, \epsilon' > 0$ y tomando $K = \text{máx}(K_\epsilon, K_{\epsilon'})$ obtenemos:

$$d(x_K, \alpha_\epsilon(d(x_0, x_K))) \leq f(\epsilon)d(x_0, x_K)$$

$$d(x_K, \alpha_{\epsilon'}(d(x_0, x_K))) \leq f(\epsilon')d(x_0, x_K)$$

lo cual implica:

$$d(\alpha_\epsilon(d(x_0, x_K)), \alpha_{\epsilon'}(d(x_0, x_K))) \leq (f(\epsilon) + f(\epsilon))d(x_0, x_K)$$

Utilizando la convexidad de la distancia entre geodésicas 4.17 obtenemos:

$$d(\alpha_\epsilon(1), \alpha_{\epsilon'}(1)) \leq f(\epsilon) + f(\epsilon')$$

Por lo tanto $\alpha_\epsilon(1)$ es de Cauchy y tiene límite cuando $\epsilon \to 0$. Esto implica que α_ϵ tiene un límite puntual que llamaremos α. Se tiene que α es una geodésica parametrizada por longitud de arco que cumple $\alpha(0) = 0$ y además (triangulando con α_ϵ):

$$d(\alpha(d(x_0, x_k)), x_k) \leq (d(\alpha(1), \alpha_\epsilon(1)) + f(\epsilon))d(x_0, x_k)\ \forall k \geq K_\epsilon$$

lo cual implica que α escolta a $\{x_n\}$. $\square$

[3] Alcanza con tomar una subsucesión tal que $\alpha_n(t)$ converge para algún t, como puede verificarse utilizando el mapa exponencial. Otra opción es utilizar el teorema de Arsela-Ascoli que garantiza la existencia de una subsucesión uniformemente convergente en compactos.

Capítulo 6

Regularidad Lyapunov

En este capítulo obtenemos a partir de lo anterior una versión del famoso teorema de Oseledets para productos de matrices invertibles aleatorias. Una demostración alternativa del mismo resultado basado sólamente en el teorema de Kingman y en el producto exterior de espacios vectoriales puede encontrarse en [Kre85].

6.1. Versión Geodesica

Teorema 6.1 (Teorema de Oseledets: Versión Geodésica)**.** *Sea $A_1, \ldots, A_n, \ldots$ una sucesión estacionaria de matrices aleatorias invertibles de $N \times N$ que cumplen:*

$$\mathbb{E}(\| \log(A_n^* A_n) \|) < +\infty \ \forall n \in \mathbb{N}$$

Si $A_n \cdots A_1 = O_n P_n$ donde $P_n \in Pos(N)$ y $O_n \in O(N)$ y fijamos $P_0 = I$ entonces se cumple:

$$\mathbb{P}(\{P_n^2\} \text{ es escoltada por una geodésica }) = 1$$

Demostración: Se tiene para todo $n \in \mathbb{N}$

$$\begin{aligned} P_n^2 &= (A_n \cdots A_1)^*(A_n \cdots A_1) \\ &= A_1^* \cdots A_n^* A_n \cdots A_1 \\ &= (\rho_{A_1} \circ \cdots \circ \rho_{A_n})(I) \end{aligned}$$

donde para cada $A \in \mathrm{GL}(N)$ el mapa $\rho_A : \mathrm{Pos}(N) \to \mathrm{Pos}(N)$ es la isometría dada por $\rho_A(P) = A^* P A$ para todo $P \in \mathrm{Pos}(N)$.

Consideremos el proceso subaditivo a que cumple:

$$a(m,n) = d(P_m^2, P_n^2)\ \forall 0 \leq m < n$$

La igualdad siguiente (que se deduce simplemente de que las ρ_A son isometrías) implica que a es un proceso estacionario:

$$\begin{aligned} a(m,n) &= d((\rho_1 \circ \cdots \circ \rho_m)(I), (\rho_1 \circ \cdots \circ \rho_n)(I)) \\ &= d(I, (\rho_m \circ \cdots \circ \rho_n)(I)) \\ &= d(I, A_m^* \cdots A_n^* A_n \cdots A_m)\ \forall 0 \leq m < n \end{aligned}$$

Por hipótesis tenemos $\mathbb{E}(a(0,1)) < +\infty$, esto implica por subaditividad $\mathbb{E}(|a(m,n)|) < +\infty$ para todo $m, n \in \mathbb{Z}_+$. Por lo tanto estamos en condiciones de aplicar el teorema 5.14.

Se obtiene que casi seguramente $\{P_n^2\}_{n\geq 0}$ o bien cumple

$$\frac{d(I, P_n^2)}{n} \to 0 \text{ cuando } n \to +\infty$$

o se escapa linealmente a infinito y está alineada.

En el primer caso cualquier geodésica por la identidad escolta a $\{P_n\}^2$ en el segundo caso se tiene que existe una escolta por 5.16. □

6.2. Regularidad Lyapunov

Definición 6.2. Una sucesión $B_1, \ldots, B_n, \ldots$ de matrices de $\mathrm{GL}(N)$ se dice que es Lyapunov regular si se cumple:

- Existen una filtración de subespacios[1] $\{0\} = V_0 \subsetneq V_1 \subsetneq \cdots \subsetneq V_m = \mathbb{R}^N$ y números $\chi_1 < \cdots < \chi_m$ tales que para todo $1 \leq i \leq m$ se cumple:

$$\lim_n \frac{1}{n} \log(\|B_n v\|) = \chi_i\ \forall v \in V_i \setminus V_{i-1}$$

- $$\lim_n \frac{1}{n} \log(|\det(B_n)|) = \sum_{1\leq i\leq m} (\dim(V_i) - \dim(V_{i-1}))\chi_i$$

[1]La palabra "filtración" significa en este contexto, cadena creciente respecto la inclusión.

Los χ_i son llamados "exponentes de Lyapunov".

El teorema 6.1 implica que la sucesión de productos parciales es casi-seguramente Lyapunov regular, como mostraremos a continuación. Utilizamos a partir de ahora $\|A\|_o = \max_{\|v\|=1} \|Av\|$ para denotar la norma de operadores de una matriz A.

Necesitaremos el siguiente lema:

Lema 6.3. *Para toda matriz $A \in GL(N)$ se cumple:*

$$|\log(\|A\|_o)| \leq \|\log(A^*A)\|$$

DEMOSTRACIÓN: Notemos que A^*A es una matriz positiva y que su máximo valor propio es exáctamente $\|A\|_o^2$. De esto se deduce el enunciado utilizando que la norma de Frobenius de la matriz simétrica $\log(A^*A)$ es simplemente la raiz cuadrada de la suma de los cuadrados de sus valores propios. □

El resultado que vincula la regularidad Lyapunov con la existencia de una escolta geodésica es entonces el siguiente:

Teorema 6.4 (Regularidad Lyapunov)**.** *Sea B_n una sucesión de matrices invertibles de $N \times N$ con $B_0 = I$, y fijemos $B_n = O_n P_n$ donde $P_n \in Pos(N)$ y $O_n \in O(N)$. Si existe $Q \in Pos(N)$ tal que:*

$$d(Q^{2n}, P_n^2) = o(n)$$

Entonces $\{B_n\}$ es Lyapunov Regular. Además fijando $e^{\chi_1} < \cdots < e^{\chi_k}$ los valores propios de Q, y $W_1, \ldots, W_k$ los subespacios propios asociados, se tiene que los exponentes de Lyapunov de B_n son $\chi_1 < \ldots < \chi_k$ y la filtración asociada es

$$V_0 = \{0\} \subsetneq V_1 = W_1 \subsetneq V_2 = W_1 \oplus W_2 \subsetneq \cdots \subsetneq V_k = W_1 \oplus \cdots \oplus W_k = \mathbb{R}^N$$

DEMOSTRACIÓN: Notemos que $|\det(B_n)| = \det(P_n)$ y para todo $v \in \mathbb{R}^N$ se tiene $\|B_n v\| = \|P_n v\|$. Por lo tanto la regularidad Lyapunov es una afirmación sobre la sucesión P_n únicamente.

Además tenemos por hipótesis:

$$d(P_n^2, Q^{2n}) = o(n)$$

Calculando explícitamente la distancia obtenemos:

$$d(P_n^2, Q^{2n}) = \| \log(Q^{-n} P_n P_n Q^{-n})\| = o(n)$$

Aplicando 6.3 a $P_n Q^{-n}$ obtenemos:

$$\frac{1}{n} \log(\|P_n Q^{-n}\|_o) \to 0 \text{ cuando } n \to +\infty$$

Supongamos entonces que $e^{\chi_1} < \cdots < e^{\chi_m}$ son los valores propios de Q mientras que W_i es el subespacio propio asociado a e^{χ_i}. Si $v \in W_i$ se obtiene:

$$\frac{1}{n} \log(\|P_n v\|) - \chi_i = \frac{1}{n} \log(\|P_n Q^{-n} v\|) \to 0 \text{ cuando } n \to +\infty$$

o

$$\frac{1}{n} \log(\|P_n v\|) \to \chi_i \text{ cuando } n \to +\infty$$

Ahora tomemos la filtración:

$$\{0\} = V_0 \subsetneq V_1 = W_1 \subsetneq V_2 = W_1 \oplus W_2 \subsetneq \cdots \subsetneq V_m = W_1 \oplus \cdots \oplus W_m = \mathbb{R}^N$$

Supongamos que para algún $i \in \{1, \ldots, m\}$ se tiene $v \in V_i \setminus V_{i-1}$. En este caso $v = v_1 + \cdots + v_i$ donde $v_j \in W_j$ para todo $j \leq i$ y $v_i \neq 0$. Por lo tanto:

$$\log(\|P_n v\|) = \log(\|P_n v_1 + \cdots P_n v_i\|)$$

y esto da para n suficientemente grande (cosa de no salirnos del dominio del logaritmo)

$$\log(\|P_n v_i\|) + \log(1 - \sum_{j<i} \frac{\|P_n v_j\|}{\|P_n v_i\|}) \leq \log(\|P_n Q^{-n} v\|) \leq \log(\|P_n v_i\|) + \log(1 + \sum_{j<i} \frac{\|P_n v_j\|}{\|P_n v_i\|})$$

lo cual, dado que si $v_j \neq 0$ se tiene que $\|P_n v_j\|$ es equivalente a $e^{n\chi_j}$, nos da:

$$\frac{1}{n} \log(\|P_n v\|) \to \chi_i \ \forall v \in V_i \setminus V_{i-1}$$

lo cual equivale a “media” regularidad Lyapunov.

La mitad que nos falta demostrar equivale a $\log(\det(P_n))/n \to \det(Q)$. Se deduce fácilmente como sigue:

$$|\frac{1}{n} \log(\det(P_n)) - \det(Q)| = \frac{1}{n} |\log(\det(P_n Q^{-n}))| \leq \frac{N}{n} |\log(\|P_n Q^{-n}\|_o)| \to 0$$

□

Notemos que la demostración anterior implica la siguiente versión del teorema de Oseledets:

Teorema 6.5 (Teorema de Oseledets: Versión Regularidad Lyapunov)**.** *Sea* $A_1, \ldots, A_n, \ldots$ *una sucesión estacionaria de matrices aleatorias invertibles de* $N \times N$ *que cumplen:*

$$\mathbb{E}(\| \log(A_n^* A_n) \|) < +\infty \ \forall n \in \mathbb{N}$$

Sea $B_n = A_n \cdots A_1$ *y fijemos* $B_0 = I$ *entonces casi seguramente la sucesión* $\{B_n\}_{n \in \mathbb{Z}_+}$ *es Lyapunov Regular.*

Capítulo 7

Teorema de Oseledets: Versión Invertible

En este capítulo demostraremos la versión del teorema de Oseledets más utilizada en dinámica. En contraste con el capítulo anterior trabajaremos con sucesiones bi-infinitas de matrices $\{A_n\}_{n\in\mathbb{Z}}$.

El teorema garantiza la existencia de ciertos subespacios tales que $V_1 \oplus \cdots \oplus V_k = \mathbb{R}^N$ con un comportamiento específico de las normas de los vectores al aplicar los productos parciales de la sucesión $\{A_n\}$ en cada subespacio.

Con respecto a 6.5, la dificultad adicional escencial es que la afirmación ya no refiere únicamente a la parte positiva de los productos parciales de la sucesión. En particular los subespacios V_i no son ortogonales.

En [Mañ83] se obtiene una demostración en el caso de que $\{A_n\}$ esté casi segurmente acotada (en realidad Mañe trabaja con el diferencial de un difeomorfismo de una variedad riemanniana compacta). Un resumen, escrito por V.Oseledets, de las distintas versiones del teorema puede encontrarse en [Ose08].

7.1. Exponentes a futuro y pasado

En todo el capítulo $\{A_n\}_{n\in\mathbb{Z}}$ denotara una sucesión estacionaria bi-infinita de matrices invertibles aleatorias de $N \times N$ que cumple:

$$\mathbb{E}(\|\log(A_n^* A_n)\|) < \infty \ \forall n \in \mathbb{Z}$$

Fijaremos también la sucesión de productos parciales:

$$B(m,n) = \begin{cases} A_{n-1}\cdots A_m & m < n \\ I & m = n \\ B(n,m)^{-1} & m > n \end{cases}$$

Para $v \in \mathbb{R}^N$ definimos el exponente de Lyapunov a futuro $\chi_+(v)$ y el exponente de Lyapunov a pasado $\chi_-(v)$ como los siguientes límites en caso de existir:

$$\chi_+(v) = \lim_{n\to+\infty} \frac{1}{n}\log(\|B(0,n)v\|)$$

$$\chi_-(v) = \lim_{n\to-\infty} \frac{1}{n}\log(\|B(0,n)v\|)$$

El siguiente teorema resume los resultados de la sección anterior, y agrega la medibilidad de la matriz positiva cuyas potencias acompañan a la parte positiva de los productos que se consideran.

Teorema 7.1. *Sea $\{A_n\}_{n\in\mathbb{Z}}$ una sucesión estacionaria de matrices aleatorias inveribles de $N \times N$ tal que*

$$\mathbb{E}(\|\log(A_n^* A_n)\|) < \infty \ \forall n \in \mathbb{Z}$$

Entonces existen matrices positivas aleatorias Q_+ y Q_- tales que casi seguramente:

1. $$d(Q_+^{2n}, B(0,n)^*B(0,n)) = o(n) \text{ cuando } n \to +\infty$$
   $$d(Q_-^{2n}, B(0,n)^*B(0,n)) = o(n) \text{ cuando } n \to -\infty$$

2. *Si $e^{\chi_1^+} < \cdots < e^{\chi_k^+}$ son los valores propios de Q_+ y $V_1^+,\ldots,V_k^+$ los subespacios propios respectivos, y además se define $e^{\chi_1^-} < \ldots < e^{\chi_{k'}^-}$ y $V_1^-,\ldots,V_{k'}^-$ análogamente para Q_- se cumple:*

   $$\chi_+(v) = \chi_i^+ \text{ si y sólo si } v \in V_1^+ \oplus \cdots \oplus V_i^+ \setminus V_1^+ \oplus \cdots \oplus V_{i-1}^+$$

   $$\chi_-(v) = \chi_i^- \text{ si y sólo si } v \in V_i^- \oplus \cdots \oplus V_{k'}^- \setminus V_{i+1}^- \oplus \cdots \oplus V_{k'}^-$$

DEMOSTRACIÓN: Fijemos

$$B(0,n) = O_n P_n,\ \ P_n \in \mathrm{Pos}(N), O_n \in \mathrm{O}(N)\ \forall n \in \mathbb{Z}$$

Por 6.1, casi seguramente existe una matriz $Q_+ \in \mathrm{Pos}(N)$ tal que:

$$d(Q_+^{2n}, P_n^2) = o(n) \text{ cuando } n \to +\infty$$

además la convexidad de la distancia entre geodésicas 4.17 implica:

$$d(Q_+, P_n^{\frac{1}{n}}) \leq \frac{1}{2n} d(Q_+^{2n}, P_n^2) \to 0 \text{ cuando } n \to +\infty$$

de modo que Q_+ es una matriz aleatoria (i.e. es medible por ser límite puntual de funciones medibles).

Las afirmaciones respecto a los subespacios V_i^+ estan dados por 6.4.

La matriz aleatoria Q_- se obtiene con exáctamente el mismo método y sus propiedades se deducen de la misma manera (notemos que $d(Q_-^{-2n}, P_{-n}^2) = o(n)$ cuando $n \to +\infty$ de modo que 6.4 se aplica a Q_-^{-1}). $\square$

7.2. Invariancia de χ_+

Notemos que 7.1 applicado a $\{A_{n-m}\}_{n\in\mathbb{Z}}$ garantiza la existencia casi segura de matrices positivas $Q_+(m)$ para cada $m \in \mathbb{Z}$ tales que si $e^{\chi_1^+(m)} < \cdots < e^{\chi_k^+(m)}$ son los valores propios de $Q_+(m)$ y $V_i^+(m)$ son los subespacios propios asociados se cumple:

$$\lim_{n\to+\infty} \frac{1}{n} \log(\|B(m, m+n)v\|) = \chi_i^+(m)$$

si y sólo si

$$v \in V_1^+(m) \oplus \cdots \oplus V_i^+(m) \setminus V_1^+(m) \oplus \cdots \oplus V_{i-1}^+(m)$$

El siguiente resultado muestra que, casi seguramente, los exponentes $\chi_i^+(m)$ no dependen de m y también muestra que $V_i^+(m) = B(0,m)V_i^+$.

Lema 7.2. *Casi seguramente $Q_+(m)$ es ortogonalmente conjugada a Q_+ para todo $m \in \mathbb{Z}$ y además:*

$$\chi_+(v) = \chi_i^+$$

si y sólamente si para algún m se cumple:

$$B(0,m)v \in V_1^+(m) \oplus \cdots \oplus V_i^+(m) \setminus V_1^+(m) \oplus \cdots \oplus V_{i-1}^+(m)$$

DEMOSTRACIÓN: Aplicando 7.1 a las sucesiones $\{A_{n-m}\}_{n\in\mathbb{Z}}$ obtenemos un conjunto de probabilidad total en el cual están definidas todas las matrices aleatorias $Q_+(m)$.

En dicho conjunto los valores propios de $Q_+(m)$ y sus subespacios propios quedan determinados por los límites:

$$\lim_{n\to+\infty}\frac{1}{n}\log(\|B(m,m+n)v\|)$$

Notemos además que dos matrices positivas son ortogonalmente conjugadas si y sólamente si tienen los mismos valores propios (contando multiplicidades).

Pero se tiene:

$$\begin{aligned}\lim_{n\to+\infty}\frac{1}{n}\log(\|B(m,m+n)v\|) &= \lim_{n\to+\infty}\frac{1}{n}\log(\|B(0,m+n)B(m,0)v\|)\\ &= \lim_{n\to+\infty}\frac{1}{m+n}\log(\|B(0,m+n)B(m,0)v\|)\\ &= \chi_+(B(m,0)v)\end{aligned}$$

Lo cual demuestra el enunciado. □

7.3. Relación entre Q_+ y Q_-

En esta sección mostraremos que Q_+ y Q_- tienen los mismos valores propios con las mismas multiplicidades. Esto implica que los valores posibles para $\chi_+(v)$ y $\chi_-(v)$ son los mismos exponentes $\chi_1 < \ldots < \chi_k$.

Lema 7.3. *Existe una sucesión $m_l \to +\infty$ tal que casi seguramente:*

$$d(Q_+(-m_l)^{2m_l}, B(-m_l,0)^*B(-m_l,0)) = o(m_l)$$

DEMOSTRACIÓN: Se obtiene por la estacionariedad de la sucesión $\{A_n\}_{n\in\mathbb{Z}}$ que cuando $m\to+\infty$:

$$\begin{aligned}&\mathbb{P}(d(Q_+(-m)^{2m}, B(-m,0)^*B(-m,0)) > \epsilon m)\\ &\qquad= \mathbb{P}(d(Q_+(0)^{2m}, B(0,m)^*B(0,m))\\ &\qquad= \mathbb{P}(d(Q_+^{2m}, P_m^2) > \epsilon m) \to 0\end{aligned}$$

Por lo tanto es posible, mediante un procedimiento diagonal, obtener una sucesión $m_l \to +\infty$ tal que para todo $\epsilon > 0$ se cumple:

$$\sum_{l\in\mathbb{N}} \mathbb{P}(d(Q_+(-m_l)^{2m_l}, B(-m_l,0)^*B(-m_l,0)) > \epsilon m_l) < +\infty$$

Y esto implica por el lema de Borel-Cantelli 1.2 que esta sucesión cumple el enunciado. □

Lema 7.4. *Casi seguramente Q_- es ortogonalmente conjugada a Q_+.*

Demostración: Fijemos la descomposición polar $B(0,-m) = O_{-m}P_{-m}$. Se cumple:

$$B(-m,0)^*B(-m,0) = (B(0,-m)B(0,-m)^*)^{-1} = O_{-m}P_{-m}^{-2}O_{-m}^{-1}$$

Ahora por 7.1 se cumple casi seguramente:

$$d(Q_-^{-2m}, P_{-m}^2) = o(m) \text{ cuando } m \to +\infty$$

Utilizando que, invertir y luego conjugar por una matriz ortogonal, es una isometría de $\mathrm{Pos}(n)$ se obtiene

$$d(O_{-m}Q_-^{2m}O_{-m}^{-1}, B(-m,0)^*B(-m,0)) = o(m) \text{ cuando } m \to +\infty$$

Si $\{m_l\}_{l\in\mathbb{N}}$ es la sucesión provista por 7.3 se obtiene por la desigualdad triangular que casi seguramente:

$$d(O_{-m_l}Q_-^{2m_l}O_{-m_l}^{-1}, Q_+(-m_l)^{2m_l}) = o(m_l) \text{ cuando } l \to +\infty$$

Utilizando 4.17 y el hecho de que conjugar por una matriz ortogonal es una isometría se obtiene:

$$\begin{aligned} d(Q_-, O_{-m_l}^{-1}Q_+(-m_l)O_{-m_l}) &= d(O_{-m_l}Q_-O_{-m_l}^{-1}, Q_+(-m_l)) \\ &\leq \frac{1}{2m_l} d(O_{-m_l}Q_-^{2m_l}O_{-m_l}^{-1}, Q_+(-m_l)^{2m_l}) \to 0 \text{ si } l \to +\infty \end{aligned}$$

Dado que la clase de conjugación por matrices ortogonales de $Q_+(m)$ es un conjunto cerrado, que no depende de m (por 7.2), se tiene que Q_- debe pertenecer a dicha clase. □

7.4. Splitting

Siguiendo con la notación de la sección anterior, fijemos las variables aleatorias k y:

$$\chi_1 < \ldots < \chi_k$$

los logaritmos de los valores propios de Q_- y Q_+. Estos valores coinciden gracias a 7.4, que además asegura que los subespacios propios asociados tienen la misma dimensión.

Sea entonces para cada $i \in \{1, \ldots, k\}$, V_i^+ el subespacio de valor propio e^{χ_i} de Q_+, y V_i^- el subespacio del mismo valor propio para Q_-.

Notemos que por 7.2, casi seguramente para cada $m \in \mathbb{Z}$ se tiene que $Q_+(m)$ tiene los mismos valores propios y multiplicidades que Q_+.

Gracias a esto último podemos definir $i \in \{1, \ldots, k\}$, $V_i^+(m)$ el subespacio propio de valor propio χ_i de la transformación $Q_+(m)$ que tendrá la misma dimensión que V_i^+ casi seguramente.

Mostraremos ahora que los subespacios V_i^+ y V_j^- no pueden estar dispuestos de cualquier manera en $\mathbb{R}^n$.

Lema 7.5. *Casi seguramente se cumple*

$$\chi_-(v) \leq \chi_+(v) \ \forall v \in \mathbb{R}^N$$

DEMOSTRACIÓN: Trabajemos en la intersección de los conjuntos de probabilidad total dados por 7.1, 7.2,7.3, y 7.4.

Suponiendo que $\chi_-(v) \geq \chi_i$ mostraremos que $\chi_+(v) \geq \chi_i$.

Con esta suposición, se cumple:

$$\|B(0, -m)v\| \leq e^{-m(\chi_i + o(1))} \text{ cuando } m \to +\infty$$

Denotemos ahora $w_m = B(0, -m)v/\|B(0, -m)v\|$ que tiene norma 1. Se cumple:

$$\|B(-m, 0)w_m\| = \frac{\|v\|}{\|B(0, -m)v\|} \geq \|v\|e^{m(\chi_i + o(1))} \ m \to +\infty$$

Fijemos $Q_m = \sqrt{B(-m, 0)^* B(-m, 0)}$, se tiene:

$$\|Q_m w_m\| \geq \|v\|e^{m(\chi_i + o(1))} \ m \to +\infty$$

Por 7.2 alcanza con demostrar que para cierto m se tiene $w_m \notin V_1^+(-m) \oplus \cdots \oplus V_{i-1}^+(-m)$, para obtener $\chi_+(v) \geq \chi_i$.

Tomando la subsucesión de valores de m dada por 7.3 se cumple:

$$d(Q_+(-m)^{2m}, Q_m^2) = o(m)$$

o explícitamente

$$\|\log(Q_+(-m)^{-m}Q_mQ_mQ_+(-m)^{-m})\| = o(m)$$

Por 6.3 esto implica

$$\|Q_mQ_+(-m)^{-m}\|_o = e^{o(m)}$$

Si se tuviera $w_m \in V_1^+(-m)\oplus\cdots\oplus V_{i-1}^+(-m)$ para todo m en la subsucesión, se cumpliría:

$$\begin{aligned}\|Q_mw_m\| &= \|Q_mQ_+(-m)^{-m}Q_+(-m)^m w_m\|\\ &\le \|Q_mQ_+(-m)^{-m}\|_o\|Q_+(-m)^m w_m\| \le e^{m(\chi_{i-1}+o(1))}\end{aligned}$$

lo cual es falso. Por lo tanto hemos mostrado que $w_m \notin V_1^+(-m)\oplus\cdots\oplus V_{i-1}^+(-m)$ para algún m y por lo tanto $\chi_+(v) \ge \chi_i$. $\square$

El resultado anterior implica que las filtraciones "a pasado" y "a futuro" de la sucesión Lyapunov regular $B(0,n)$ deben intersectarse para dar una descomposición (o "splitting") de $\mathbb{R}^n$ en suma directa de subespacios en los cuales $\chi_+ = \chi_-$. Esta es otra versión del teorema de Oseledets.

Teorema 7.6 (Oseledets: Versión Invertible)**.** *Sea $\{A_n\}_{n\in\mathbb{Z}}$ una sucesión estacionaria de matrices aleatorias invertibles de $N\times N$ tal que:*

$$\mathbb{E}(\|\log(A_n^*A_n)\|) < +\infty \ \forall n\in\mathbb{Z}$$

Defininamos para $m,n\in\mathbb{Z}$:

$$B(m,n) = \begin{cases} A_{n-1}\cdots A_m & m<n\\ I & m=n\\ B(n,m)^{-1} & m>n\end{cases}$$

Entonces existen Q_+ y Q_- matrices positivas aleatorias tales que casi seguramente se cumple:

1.
$$d(Q_+^{2n}, B(0,n)^*B(0,n)) = o(n) \text{ cuando } n\to+\infty$$
$$d(Q_-^{2n}, B(0,n)^*B(0,n)) = o(n) \text{ cuando } n\to-\infty$$

2. *Q_+ y Q_- son ortogonalmente conjugadas.*

3. *Si tomamos las variables aleatorias k y $\chi_1 < \ldots < \chi_k$ tales que $e^{\chi_1} < \cdots < e^{\chi_k}$ son los valores propios de Q_+ y Q_-, y definimos para cada i, V_i^+ y V_i^- los subespacios propios de valor propio e^{χ_i} para Q_+ y Q_- respectivamente se tiene:*

$$V_1^+ \oplus \cdots \oplus V_{i-1}^+ \cap V_i^- \oplus \cdots \oplus V_k^- = \{0\}$$

4. *Definiendo $V_i = V_1^+ \oplus \cdots \oplus V_i^+ \cap V_i^- \oplus \cdots \oplus V_k^-$ se tiene $\mathbb{R}^N = V_1 \oplus \cdots \oplus V_k$ y además:*

$$\chi_+(v) = \chi_-(v) = \chi_i \ \forall v \in V_i$$

Demostración: Tomamos Q_+ y Q_- dados por 7.1 que garantiza que son medibles (i.e. que son matrices aleatorias) y que cumplen la propiedad 1.

Por 7.4 se tiene que casi seguramente se cumple 2. Mientras que 7.5 implica 3, dado que un vector no nulo $v \in V_1^+ \oplus \cdots \oplus V_{i-1}^+ \cap V_i^- \oplus \cdots \oplus V_k^-$ tendría $\chi_+(v) \le \chi_{i-1} < \chi_i \le \chi_-(v)$.

Falta demostrar que $\mathbb{R}^N = V_1 \oplus \cdots \oplus V_k$. Para esto fijemos $d_i = \dim(V_i^+) = \dim(V_i^-)$. Se tiene por lo anterior que:

$$V_1^+ \oplus \cdots \oplus V_{i-1}^+ \cap V_i^- \oplus \cdots \oplus V_k^- = \{0\}$$

lo cual implica (usando $\dim(X \cap Y) = \dim(X) + \dim(Y) - \dim(X+Y)$)

$$V_1^+ \oplus \cdots \oplus V_{i-1}^+ \oplus V_i^- \oplus \cdots \oplus V_k^- = \mathbb{R}^n$$

y (por la misma fórmula para la dimensión de una intersección) $\dim(V_i) = d_i$.

Si $0 \neq v = v_1 + \cdots + v_i$ con $v_j \in V_j$ para todo j se tiene $\chi_+(v) \le \chi_i$, por lo tanto:

$$(V_1 + \cdots + V_i) \cap V_{i+1} = \{0\}$$

Esto implica que:

$$\dim(V_1 + \cdots + V_{i+1}) = \dim(V_1 + \cdots + V_i) + \dim(V_{i+1}) = \dim(V_1 + \cdots + V_i) + d_{i+1}$$

Dado que $V_1 = V_1^+$ se obtiene por inducción que:

$$V_1 \oplus \cdots \oplus V_k = \mathbb{R}^N$$

□

7.5. Versión Diferenciable

Dado un difeomorfismo de una variedad riemanniana, los vectores tangentes tienen asociados exponentes de Lyapunov como se define a continuación.

Definición 7.7 (Exponentes de Lyapunov)**.** Sea M una variedad riemanniana y $f : M \to M$ un difeomorfismo. Para cada $(x, v) \in TM$ definimos los exponentes de Lyapunov a futuro y a pasado (χ_+ y χ_- respectivamente) como los siguientes límites en caso de que existan:

$$\chi_+(v) = \lim_{n \to +\infty} \frac{1}{n} \log(\|D_x f^n v\|)$$

$$\chi_-(v) = \lim_{n \to -\infty} \frac{1}{n} \log(\|D_x f^n v\|)$$

En este contexto demostraremos la siguiente versión del teorema de Oseledets[1].

Teorema 7.8. *Sea M una variedad riemanniana compacta, $f : M \to M$ un difeomorfismo, y μ una probabilidad boreliana en M que es f-invariante.*

Entonces para μ casi todo $x \in M$ existe $1 \leq k(x) \leq dim(M)$, subespacios $V_1(x) \oplus \cdots \oplus V_k(x) = T_x M$, y exponentes $\chi_1(x) < \ldots < \chi_k(x)$ tales que:

$$\chi_+(v) = \chi_-(v) = \chi_i \ \forall v \in V_i(x) \setminus \{0\}$$

$$\lim_{n \to \infty} \frac{1}{n} \log\left(|det(D_x f^n)|\right) = \sum_i dim(V_i(x))\chi_i(x)$$

Demostración: Cubrimos M con numerables conjuntos medibles disjuntos tales que en cada uno de ellos tenemos una base ortornormal de $T_x M$ que varia continuamente respecto a x. Utilizando dichas bases ortornormales identificamos cada espacio tangente $T_x M$ con $\mathbb{R}^N$ donde $N = \dim(M)$.

Definamos para cada $n \in \mathbb{Z}$, la matriz $A_n(x)$ asociada a $D_{f^n(x)} f$. La función $A_n : M \to \mathrm{GL}(N)$ es medible para cada $n \in \mathbb{Z}$. Además se cumple que $A_n \circ f = A_{n+1}$ para todo $n \in \mathbb{Z}$.

[1]Notemos que está bien definido el determinante de cualquier transformación lineal entre dos espacios vectoriales con producto interno de la misma dimensión finita (simplemente fijando bases ortonormales en ambos espacios).

Sean ahora $E_1, \ldots, E_n \subset \mathrm{GL}(N)$ conjuntos borelianos. Para todo $k \in \mathbb{Z}$ se cumple:

$$\begin{aligned}
&\mu(\{x : A_{k+1} \in E_1, \ldots, A_{k+n} \in E_k\}) \\
&\qquad = \mu(f^{-1}\{x : A_{k+1} \in E_1, \ldots, A_{k+n} \in E_k\}) \\
&\qquad = \mu(\{x : A_{k+1} \circ f \in E_1, \ldots, A_{k+n} \circ f \in E_k\}) \\
&\qquad = \mu(\{x : A_{k+2} \in E_1, \ldots, A_{k+n+1} \in E_k\})
\end{aligned}$$

De modo que $\{A_n\}_{n \in \mathbb{Z}}$ es estacionaria.

El teorema se deduce directamente de 6.5 y 7.6. $\square$

Bibliografía

[And05] James W. Anderson. *Hyperbolic geometry*. Springer Undergraduate Mathematics Series. Springer-Verlag London Ltd., London, second edition, 2005.

[Ash00] Robert B. Ash. *Probability and measure theory*. Harcourt/Academic Press, Burlington, MA, second edition, 2000. With contributions by Catherine Doléans-Dade.

[Bha07] Rajendra Bhatia. *Positive definite matrices*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ, 2007.

[Bir31] G. D. Birkhoff. Proof of the ergodic theorem. *Proc. Nat. Acad. Sci. U.S.A.*, 17:656–660, 1931.

[BP01] L. Barreira and Ya. Pesin. Lectures on Lyapunov exponents and smooth ergodic theory. In *Smooth ergodic theory and its applications (Seattle, WA, 1999)*, volume 69 of *Proc. Sympos. Pure Math.*, pages 3–106. Amer. Math. Soc., Providence, RI, 2001. Appendix A by M. Brin and Appendix B by D. Dolgopyat, H. Hu and Pesin.

[FK60] H. Furstenberg and H. Kesten. Products of random matrices. *Ann. Math. Statist*, 31:457–469, 1960.

[Fur81] H. Furstenberg. *Recurrence in ergodic theory and combinatorial number theory*. Princeton University Press, Princeton, N.J., 1981. M. B. Porter Lectures.

[Kaĭ89] V. A. Kaĭmanovich. Lyapunov exponents, symmetric spaces and a multiplicative ergodic theorem for semisimple Lie groups. *J. Soviet Math.*, 47:2387–2398, 1989.

[Kin68] J. F. C. Kingman. The ergodic theory of subadditive stochastic processes. *J. Roy. Statist. Soc. Ser. B*, 30:499–510, 1968.

[KL06] Anders Karlsson and François Ledrappier. On laws of large numbers for random walks. *Ann. Probab.*, 34(5):1693–1706, 2006.

[KM99] Anders Karlsson and Gregory A. Margulis. A multiplicative ergodic theorem and nonpositively curved spaces. *Comm. Math. Phys.*, 208(1):107–123, 1999.

[KP02] Steven G. Krantz and Harold R. Parks. *A primer of real analytic functions*. Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks]. Birkhäuser Boston Inc., Boston, MA, second edition, 2002.

[Kre85] Ulrich Krengel. *Ergodic theorems*, volume 6 of *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 1985. With a supplement by Antoine Brunel.

[Lan99] Serge Lang. *Fundamentals of differential geometry*, volume 191 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1999.

[Mañ83] Ricardo Mañé. *Introdução à teoria ergódica*, volume 14 of *Projeto Euclides [Euclid Project]*. Instituto de Matemática Pura e Aplicada (IMPA), Rio de Janeiro, 1983.

[Ose08] V. Oseledets. Oseledets theorem. *Scholarpedia*, 3(1):1846, 2008.

[Shi96] A.Ñ. Shiryaev. *Probability*, volume 95 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1996. Translated from the first (1980) Russian edition by R. P. Boas.

[Tal96] Michel Talagrand. A new look at independence. *Ann. Probab.*, 24(1):1–34, 1996.