Trabajo Monográfico

Cadenas de Markov con restricciones y aplicación a la composición automática de música tonal

Verónica Rumbo

Orientadores: Ernesto Mordecki (CMAT, FCIEN) y Paola Bermolen (IMERL, FING)

Licenciatura en Matemática, Facultad de Ciencias

Universidad de la República, Uruguay

11 de marzo de 2017

Índice general

Introducción

En el presente trabajo se explorarán posibles usos de cadenas de Markov para la composición algorítmica de música tonal. Dicha música se caracteriza por tener una estructura fuertemente jerarquizada en sus notas y acordes, donde los roles se distribuyen en función de una nota principal que se denomina *tónica* y las transiciones entre notas y acordes se encuentran en gran medida arbitradas por el rol que le corresponde a los mismos. El problema puede enmarcarse en el contexto de la modelación matemática de fenómenos musicales, en el cual se tienen diversos tópicos que han despertado interés a lo largo de los años, con especial impulso tras el desarrollo de los sistemas computacionales.

Así, la composición algorítmica, el reconocimiento de sonidos y melodías y la síntesis de sonido emulando ciertos timbres son algunos de los temas de interés en el área, siendo el primero de éstos nuestro principal interés.

La automatización total o parcial del proceso compositivo, con o sin elementos aleatorios, data de la antigua Grecia[12], aunque un caso paradigmático puede encontrarse varios siglos después en la *Musikalisches Würfelspiel* (Música de dados) del siglo XVIII, siendo el "Juego de dados" de W.A. Mozart una de las creaciones más conocidas bajo esta premisa [1]. En este tipo de piezas se dispone de ciertos fragmentos pre-compuestos entre los cuales se elige según el resultado de un evento aleatorio (por ejemplo, tirar un dado). Cabe observar que, en este caso, la incidencia del azar en el resultado es escasa ya que la estructura y estilo de la música quedan determinados por los bloques preexistentes.

Finalizado el siglo XVIII la *Musikalisches Würfelspiel* perdió popularidad, y es a partir del siglo XX donde la composición algorítmica encuentra su mayor auge. El surgimiento del dodecafonismo por un lado, y la denominada música aleatoria por otro son casos emblemáticos de estilos musicales en los cuales lo automático y a veces lo aleatorio se utilizan como elemento estético.

El dodecafonismo consiste esencialmente en considerar una secuencia con 12 notas con distinta clase de octava[2] a la cual se le aplican ciertas transformaciones. Utilizando dichas transformaciones (inversión, reversión, transporte, y combinaciones de éstas) sobre la secuencia original se

[1] En realidad la obra fue publicada por su editor Nikolaus Simrock, y no se ha podido autenticar la alegada autoría de Mozart.

[2] Al comienzo del capítulo 5 puede encontrarse una descripción de las clases de octava.

obtienen 48 secuencias de 12 notas a partir de las cuales se componen las obras. Este método es completamente determinista.

Por otra parte la música aleatoria incorpora elementos azarosos o indeterminados en sus composiciones. Éstos pueden aparecer explícitamente (en forma de indicaciones vagas al intérprete o indicaciones de reordenar libremente las notas en ciertos pasajes), o estar involucrados durante el proceso de composición de la obra pero no durante la interpretación.

Ambos estilos se contraponen, en tanto el primero intenta controlar lo más posible los parámetros de la música y el segundo, en cambio, procura dejar elementos librados al azar. Sin embargo ambos son casos de música atonal, que no es el estilo que procuramos emular en nuestro trabajo.

El caso que nos atañe, que es el de la música tonal, también ha motivado intentos de composición algorítmica más allá de la simple selección aleatoria de módulos precompuestos de modo convencional. En este sentido cabe destacar el relativamente reciente trabajo de David Cope[5], cuyas composiciones logran emular razonablemente, al menos para un oído inexperto, estilos y compositores tonales. Su enfoque inicial puede verse como una variante más sofisticada de la *Musikalisches Würfelspiel* basada en redes de transición aumentadas, donde considerando un conjunto de obras de similar estilo el modelo identifica automáticamente los fragmentos adecuados a utilizar para luego componer. Posteriormente el modelo fue adquiriendo mayor complejidad derivando en la creación de *Emily Howell* [4], un programa que "aprende" a partir de un conjunto de obras generadas por otro programa (su precedesor EMI -*Experiments in Musical Intelligence*-) y procura desarrollar su propio estilo.

En general, se han utilizado diversas técnicas con la intención de automatizar la generación de música, algunas de ellas son:

- Redes neuronales artificiales,
- Gramáticas generativas,
- Modelos de Markov.

Una descripción de de esas y otras técnicas puede encontrarse en [15]. Asimismo, en [9] pueden encontrarse ejemplos de música generada utilizando redes neuronales así como una descripción detallada del procedimiento utilizado para la composición. Este trabajo, sin embargo, pondrá énfasis en los modelos de Markov.

Algunas variantes de cadenas de Markov se emplean para la modelación de fenómenos musicales. Las cadenas de Markov de largo N (no necesariamente 1), o de largo variable, permiten modelar procesos de memoria un poco más larga que las cadenas de Markov ordinarias. Sin embargo considerar un largo demasiado grande puede llevar a que el modelo copie literalmente fragmentos del conjunto de obras (*corpus*) original. Otro aspecto a considerar es que aumentar

el orden de la cadena de Markov implica aumentar notoriamente el costo computacional, aunque pueden adoptarse estrategias de optimización almacenando únicamente las probabilidades no nulas.

Por otra parte se tienen las cadenas de Markov ocultas, en las cuales no se conocen las variables de la cadena sino que se observan otras variables, cuyas probabilidades de aparición dependen del comportamiento de la cadena subyacente (en particular, del estado -oculto- que ocurre en ese instante). Este enfoque, con eventuales variantes, es de utilidad por ejemplo para estimar acordes (la variable no observable) a partir del audio (ver, por ejemplo, [3]).

Por último destacamos las cadenas de Markov con restricciones, que serán el principal objeto de estudio en este trabajo. Las mismas son un caso particular de cadenas no homogéneas en el tiempo, resultantes de considerar una cadena homogénea e imponerle ciertas condiciones. En la práctica compositiva la aplicación de restricciones permite imponer o prohibir determinados comportamientos en las obras a componer basándose, por ejemplo, en las reglas que rigen la música tonal y su armonía.

Utilizando tales cadenas de Markov generamos ejemplos de música cuyas alturas fueron elegidas aleatoriamente logrando preservar varios rasgos característicos de la música tonal. Por un lado, se crearon "variaciones" sobre una melodía breve y conocida (Arroz con Leche), y por otro se exploró la aplicación de restricciones en el caso de música polifónica, tomando algunos corales de J.S.Bach como caso de estudio. La elección de dichos corales se debe, por una parte, a que se trata de un conjunto numeroso de obras que podemos considerar del mismo estilo y cuya conducción de voces está bastante reglamentada. Por otra parte veremos más adelante que la elección responde también a un criterio técnico: disponemos de los corales en un formato que permite trabajar fácilmente con ellos.

El problema se abordó desde varias vertientes: en primer lugar el estudio del modelo matemático elegido, las cadenas de Markov con restricciones propuestas por Pachet, Roy y Barbieri en [18], y su aplicación el el proceso compositivo. Utilizaremos dichas cadenas para modelar las alturas de las notas y utilizaremos las restricciones para fijar algunas de ellas. Para ello consideraremos un tipo de restricciones denominadas unitarias pues operan sobre un solo estado, y si bien presentaremos restricciones que operan sobre las transiciones (restricciones binarias) no las aplicaremos al momento de la implementación. Por otra parte se requieren algunos resultados de estadística para realizar la estimación de probabilidades de transición en las cadenas de Markov.

Claramente será necesario disponer de un trasfondo de teoría musical para las aplicaciones. Además de presentar algunas nociones básicas, será importante elegir las restricciones con criterio musical (más concretamente, basado en la armonía subyacente). En ese sentido cabe destacar el aporte de Luis Jure (Estudio de Música Electroacúsica, Escuela Universitaria de Música, UdelaR), que nos ayudó a mejorar la elección de las restricciones con lo cual se mejoraron algunos de los ejemplos realizados hasta el momento (en el caso de Arroz con Leche). Dicho enfoque también puede aplicarse para mejorar las simulaciones de Corales de Bach, lo cual si bien no fue realizado en este trabajo queda como un problema a explorar a futuro.

El otro punto a considerar es la programación y los aspectos técnicos. Por un lado se describirán e implementarán algoritmos para la simulación de cadenas de Markov (en particular con restricciones). Por otro, la elección de un medio adecuado para manipular la música: los dos formatos ubicuos, audio y partituras, suelen presentarse de un modo en el que no es posible operar con sus notas y la identificación automática de las mismas constituye un gran problema en sí mismo que queremos evitar. En este y otros aspectos contamos con el valioso aporte de Martín Rocamora (Grupo de Procesamiento de Audio, Instituto de Ingeniería Eléctrica, Facultad de Ingeniería, UdelaR) con quien mantuvimos un fluído intercambio. Finalmente optamos por utilizar la biblioteca de *python music21* [13] para resolver el problema del procesamiento de piezas musicales.

Finalmente, el contenido del trabajo se distribuye del siguiente modo: en el primer capítulo se presentan las cadenas de Markov homogéneas así como los resultados centrales referidos a ellas. El capítulo 2 introduce las cadenas de Markov con restricciones así como un algoritmo para construirlas. En el capítulo 3 se tratan estrategias para estimar parámetros (i.e. distribución inicial y probabilidades de transición) en cadenas de Markov homogéneas y se definen algoritmos para simular cadenas de Markov en general. El capítulo 5 describe la aplicación de las cadenas de Markov con restricciones a la composición de música y presenta los dos ejemplos antes mencionados.

Capítulo 1

Preliminares

1.1. Cadenas de Markov: definición y generalidades

Las sucesiones de variables aleatorias son un objeto de interés dado que conocer su comportamiento asintótico permite aproximar el comportamiento de la variable a partir de una secuencia de observaciones. Así, importantes resultados como las leyes de grandes números o el teorema del límite central y de Lindeberg refieren a la convergencia de las mismas. Por otra parte es usual que en un primer acercamiento a estas sucesiones (y en las formulaciones más usuales de los resultados antes mencionados) se requiera que las variables sean independientes.

Las cadenas de Markov en sus diversas variantes nos brindan un marco teórico en el cual en lugar de requerir que las variables en la sucesión sean independientes se requiere que dependan (a lo sumo) de la variable inmediata anterior o -ampliando un poco el modelo- de una cantidad N de variables inmediatamente anteriores.

Definición 1.1.1. Sean $E = \{e_i\}_{i\in I} \subset \mathbb{R}$ un conjunto finito o numerable con una numeración I, $\mu = (\mu_i)_{i\in I}$ un vector o familia numerable de reales positivos con $\sum_{i\in I} \mu_i = 1$ y $\{X_k\}_{k\in\mathbb{N}}$ una sucesión de variables aleatorias que toman valores en E. Diremos que $\{X_k\}$ es una cadena de Markov con espacio de estados E y distribución inicial μ si:

- X_0 tiene distribución μ, es decir, si $P(X_0 = e_i) = \mu_i$, $\forall i \in I$.
- Para todo $n \in \mathbb{N}$ y para toda n-upla $e_{i_0}, e_{i_1}, ..., e_{i_n}$ de elementos de E se cumple que

$$P(X_n = e_{i_n} | X_{n-1} = e_{i_{n-1}}, X_{n-2} = e_{i_{n-2}}, ..., X_0 = e_{i_0}) = P(X_n = e_{i_n} | X_{n-1} = e_{i_{n-1}}).$$

Al conjunto E lo denominamos *espacio de estados* mientras que a μ lo llamamos *distribución inicial*

Un caso particular de interés es el de las cadenas de Markov homogéneas en el tiempo, es decir, aquellas cuyas probabilidades de ir de un estado a otro en un tiempo dado dependen sólo de los estados involucrados (no así del tiempo). Buena parte de la teoría que se desarrolla a continuación está enfocada en cadenas de Markov homogéneas por lo que en el resto del capítulo las cadenas de Markov que se presenten serán homogéneas a menos que se especifique lo contrario.

1.2. Matriz de transición y probabilidades de orden n

Si en una cadena de Markov se tiene que para todo par de estados e_i, e_j las probabilidades $P(X_n = e_j | X_{n-1} = e_i)$ no dependen de n se dice que la cadena es *homogénea* (en el tiempo). En ese caso tiene sentido denominar a dicha probabilidad como p_{ij} (la probabilidad de ir de i a j en un paso). Si la cadena no es homogénea será necesario también indicar el instante considerado como se verá más adelante.

Para simplificar la notación, asumiremos en general que el espacio de estados es de la forma $\{1, \ldots, n\}$ (o bien $\mathbb{N}$, o eventualmente $\mathbb{Z}$ si el espacio es infinito). No se pierde generalidad, ya que simplemente identificamos el espacio de estados E con la numeración I. Asimismo, notaremos $p_i(j) := P(X_1 = j | X_0 = i)$.

Definición 1.2.1. Sea $\{X_k\}_{k\in\mathbb{N}}$ una cadena de Markov con espacio de estados E como en la definición 1.1.1. Definimos su matriz de transición P como la matriz $(p_{ij})_{i,j\in E}$ tal que $p_{ij} = P(X_n = j | X_{n-1} = i)$.

Observación 1.2.2.

- Como la cadena considerada es homogénea las entradas de la matriz no dependen de n. Más aún, con la notación incorporada se tiene que $p_{ij} = p_i(j)$. Mantendremos la notación $p_i(j)$ a fin de recordar que son probabilidades condicionales.
- Las entradas de una matriz de transición $p_i(j)$ verifican que $\sum_{j\in I} p_i(j) = 1$ para todo $i \in E$. Decimos que tales matrices son *estocásticas*.

De ese modo una cadena de Markov (homogénea) queda determinada por su espacio de estados, su distribución inicial y una matriz estocástica que será su matriz de transición.

Ejemplo 1.2.3. Sea E un conjunto como en la definición 1.1.1. Si $\{X_k\}_{k\in\mathbb{N}}$ es una sucesión de variables aleatorias aleatorias independientes e idénticamente distribuidas (i.i.d) que toman valores en E, entonces $\{X_k\}$ es una cadena de Markov con matriz de transición $(p_i(j))_{i,j}$ -siendo $p_i(j) = p_j := P(X_0 = j)$- y distribución inicial μ con $\mu_j = p_j \forall j \in I$.

Para probarlo basta calcular $P(X_n = i_n | X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, ..., X_0 = i_0)$ y $P(X_n = i_n | X_{n-1} = i_{n-1})$. Como las variables $X_n, X_{n-1}, ..., X_0$ son independientes se tiene:

$$P(X_n = i_n | X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, ..., X_0 = i_0) = P(X_n = i_n) = p_{i_n}.$$

Con la otra probabilidad condicional ocurre lo mismo:

$$P(X_n = i_n | X_{n-1} = i_{n-1}) = P(X_n = i_n) = p_{i_n}.$$

Con lo cual queda demostrado que $\{X_k\}$ es una cadena de Markov. Luego de la definición de los μ_j se desprende de inmediato que su distribución inicial es μ.

Además de las sucesiones de variables i.i.d podemos considerar las sumas parciales de algunas de estas sucesiones, obteniendo así otra familia de cadenas de Markov.

Ejemplo 1.2.4 (Paseos al azar). Si $\{X_k\}_{k\in\mathbb{N}}$ una sucesión de variables aleatorias i.i.d que toman valores en $\mathbb{Z}$ se tiene que $S_k = \sum_{i=1}^{k} X_i$ es una cadena de Markov. En efecto, si $i_1, \ldots, i_n$ son números enteros:

$$\begin{aligned} P(S_n = i_n | S_{n-1} = i_{n-1}, \ldots, S_0 = i_0) &= P(\sum_{i=1}^{n} X_k = i_n | \sum_{i=1}^{n-1} X_k = i_{n-1}, \ldots, X_0 = i_0) \\ &= P(X_n + i_{n-1} = i_n) = P(X_n = i_n - i_{n-1}). \end{aligned} \tag{1.1}$$

Lo mismo ocurre al calcular $P(S_n = i_n | S_{n-1} = i_{n-1})$ ya que por la independencia de $\{X_k\}$ lo único que se necesita es conocer el valor de la suma parcial en el instante $n-1$.

Además, dados $i, j \in \mathbb{Z}$ la probabilidad de transición de i a j es d_{j-i}, notando $d_k := P(X = k)$. Así, considerando los propios estados como índices la matriz de transición resulta ser $P = (p_i(j))_{i,j} = (d_{j-i})_{i,j}$.

Resulta de interés conocer también las probabilidades de orden n de una cadena de Markov. Esto es, dada $\{X_k\}_{k\in\mathbb{N}}$ cadena de Markov con espacio de estados E, matriz de transición P y distribución inicial μ definimos su *matriz de transición de orden n* como

$$P^{(n)} = (p_i^n(j))_{i,j\in E}, \quad \text{siendo } p_i^n(j) = P(X_n = j | X_0 = i),$$

donde las probabilidades $p_i^n(j)$ se denominan *probabilidades de transición de orden n*. Asimismo, llamaremos *distribución de probabilidad de orden n* al vector

$$\mu^n = (\mu_i^n)_{i\in E}, \quad \text{siendo } \mu_i^n = P(X_n = i).$$

Cabe notar que $P^{(n)}$ es por el momento sólo una notación y no refiere a P^n como potencia la matriz P. Sin embargo se puede probar que ambas coinciden.

Proposición 1.2.5. *Sea $\{X_k\}_{k\in\mathbb{N}}$ una cadena de Markov con espacio de estados E, matriz de transición $P = (p_i(j))$ y distribución inicial μ. Para todo $n \in \mathbb{N}$ se cumple:*

1. $P^{(n)} = P^n$.
2. $\mu^n = \mu \times P^n, \forall n \in \mathbb{N}$.

Demostración. 1. Si $n = 0$ se tiene que:

$$p_i^n(j) = P(X_0 = j | X_0 = i) = \begin{cases} 1, & \text{si } i = j, \\ 0, & \text{si } i \neq j, \end{cases}$$

con lo cual $P^{(0)}$ es la matriz identidad, al igual que P^0. Si además verificamos que $P^{(n+1)} = P \times P^{(n)}$, por inducción tenemos:

$$P^{(n+1)} = P \times P^{(n)} \overset{Induccion}{=} P \times P^n = P^{n+1}.$$

Para verificar que $P^{(n+1)} = P \times P^{(n)}$ probaremos que, dados m y n naturales, $P^{(n+m)} = P^{(n)}P^{(m)}$.

$$\begin{aligned} p_i^{n+m}(j) = P(X_{n+m} = j | X_0 = i) &= \sum_{k\in E} P(X_{n+m} = j, X_m = k | X_0 = i) \\ &= \sum_{k\in E} P(X_{n+m} = j | X_m = k, X_0 = i) P(X_m = k | X_0 = i) \\ &\overset{*}{=} \sum_{k\in E} P(X_n = j | X_0 = k) P(X_m = k | X_0 = i) \\ &= \sum_{k\in E} p_k^n(j) p_i^m(k). \end{aligned} \tag{1.2}$$

donde en la igualdad $*$ se utilizó la homogeneidad de la cadena. Notar que la entrada i, j-ésima de $P^{(n+m)}$ es, efectivamente la entrada i, j-ésima del producto $P^{(n)}P^{(m)}$.

Luego como por definición de $P^{(n)}$ sabemos que $P^{(1)} = P$, tomando $m = 1$ se tiene la igualdad buscada.

2. Utilizando lo probado en la parte anterior se verifica que

$$\mu_j^n = P(X_n = j) = \sum_{k\in E} P(X_n = j | X_0 = k) P(X_0 = k) = \sum_{k\in E} p_k^n(j)\mu_k = \mu \times P^n.$$

□

También resulta de interés conocer la probabilidad de una trayectoria finita arbitraria, es decir, las probabilidades de la forma $P(X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0)$. Descomponiendo en probabilidades condicionales y utilizando la definición 1.1.1 se tiene que

$$\begin{aligned} &P(X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \\ =&P(X_n = i_n | X_{n-1} = i_{n-1}) P(X_{n-1} = i_{n-1} | X_{n-2} = i_{n-2}) \dots P(X_1 = i_1 | X_0 = i_0) P(X_0 = i_0) = \\ =&p_{i_{n-1}}(i_n) p_{i_{n-2}}(i_{n-1}) \dots p_0(1) \mu_{i_0}. \end{aligned}$$

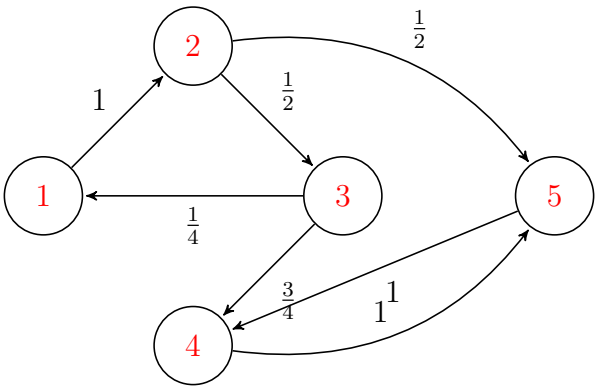
Además cabe observar que no es realmente relevante que la trayectoria comience en tiempo 0. El resultado vale también si consideramos las variables $X_m, X_{m+1}, \dots, X_{m+n}$ con $m \in \mathbb{N}$, utilizando μ^m en lugar de μ. En este caso se utiliza la homogeneidad en el tiempo, pero se verá también una expresión para el caso no homogéneo más adelante.

Hasta ahora hemos introducido notación y presentado herramientas que permiten calcular probabilidades en cadenas de Markov. Ahora nos interesa describir cierta forma de accesibilidad en la cadena: ¿Es siempre posible realizar una trayectoria entre dos estados dados? ¿Es posible hacerlo fijando además la cantidad de transiciones a utilizar? ¿Siempre se puede alcanzar un estado dado en tiempo finito? A continuación se presentarán algunos conceptos y ejemplos que permiten responder estas preguntas. Para ello nos será útil presentar los *grafos de transición*: son grafos cuyos vértices representan los estados de la cadena y sus aristas las probabilidades de transición.

1.3. Comunicación entre estados y reducibilidad

Una noción importante es la de *comunicación entre estados*. Comencemos con un ejemplo:

Ejemplo 1.3.1. Consideremos una cadena de Markov en $E = \{1, 2, 3, 4, 5\}$ con distribución inicial $\mu = (\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5})$ y probabilidades de transición indicadas en el siguiente grafo:



Puede apreciarse que no se puede, por ejemplo, ir desde el estado 5 hacia el 2. En cambio se puede transitar libremente entre los estados 5 y 4, así como entre 1,2 y 3. Asimismo, si tenemos una trayectoria situada en el estado 1 es posible alcanzar todos los estados, pero si nos situamos en los estados 4 y 5 ya no podemos salir de este bucle. Para formalizar un poco definiremos algunos conceptos.

Definición 1.3.2. Sea $\{X_k\}_{k\in\mathbb{N}}$ una cadena de Markov con espacio de estados E y matriz de transición $P = (p_i(j))$.

1. Si i, j son estados, decimos que de i *se llega a* j si $P(X_n = j \text{ para algún } n \in \mathbb{N}|X_0 = i) > 0$ y notaremos $i \to j$.
2. Consideremos un estado i. Si para todo estado j tal que $i \to j$ se tiene que $j \to i$, se dice que i es *esencial*.

De la definición anterior se desprende que dados dos estados esenciales i, j, si $i \to j$ se tiene que $j \to i$ y viceversa. En el conjunto de los estados esenciales tiene sentido decir que dos estados *se comunican* y se prueba fácilmente que dicha relación (que notaremos $i \leftrightarrow j$) es una relación de equivalencia. Llamaremos *clases irreducibles* a sus clases de equivalencia.

Definición 1.3.3. Sea $\{X_k\}_{k\in\mathbb{N}}$ una cadena de Markov con espacio de estados E y matriz de transición $P = (p_i(j))$. Si E es una clase irreducible decimos que la cadena es *irreducible*.

Observación 1.3.4. Cuando una cadena de Markov alcanza un estado esencial, queda "atrapada" en la clase irreducible correspondiente. De este modo cada clase irreducible puede pensarse como una cadena de Markov en sí misma y las cadenas irreducibles son, en ese sentido, indescomponibles. Interesa particularmente estudiar dichas cadenas.

En el ejemplo 1.3.1 se puede apreciar que hay estados no esenciales (1, 2 y 3) con lo cual la cadena no es irreducible. Los estados 4 y 5, por otra parte, sí son esenciales y forman la única clase irreducible.

Proposición 1.3.5. *Sea $\{X_k\}_{k\in\mathbb{N}}$ una cadena de Markov con espacio de estados E, matriz de transición $P = (p_i(j))_{i,j}$ e i, j dos estados distintos. Son equivalentes:*

1. $i \to j$.
2. *Existen estados $i_1, i_2, \ldots, i_{n-1}$ tales que $p_i(i_1)p_{i_1}(i_2)\ldots p_{i_{n-1}}(j) > 0$.*
3. $p_i^n(j) > 0$ *para algún* $n \in \mathbb{N}$.

Demostración.$1 \to 3$ Como $i \to j$ se tiene que $P(\bigcup_{n\in\mathbb{N}} \{X_n = j\}|X_0 = i) > 0$. Luego

$$0 < P(\bigcup_{n\in\mathbb{N}}\{X_n = j\}|X_0 = i) \leq \sum_{n\in\mathbb{N}} P(X_n = j|X_0 = i) = \sum_{n\in\mathbb{N}} p_i^n(j),$$

lo cual significa que al menos uno de los sumandos $p_i^n(j)$ debe ser positivo.

$3 \to 1$ Basta observar que para todo $n \in \mathbb{N}$, $P(\bigcup_{k\in\mathbb{N}}\{X_k = j\}|X_0 = i) \geq P(X_n = j|X_0 = i) = p_i^n(j)$. Como existe n de modo que $p_i^n(j) > 0$, se verifica $P(\bigcup_{n\in\mathbb{N}} X_n = j|X_0 = i) > 0$.

$2 \to 3$ Como $p_i(i_1)p_{i_1}(i_2)\dots p_{i_{n-1}}(j)$ es la probabilidad de **una** trayectoria que va de i a j se tiene $p_i^n(j) \geq p_i(i_1)p_{i_1}(i_2)\dots p_{i_{n-1}}(j) > 0$ con lo cual $p_i^n(j) > 0$.

$3 \to 2$ Considero n tal que $p_i^n(j) > 0$. Como $p_i^n = \sum_{k_1,k_2,\dots,k_{n-1}\in E} p_i(k_1)p_{k_1}(k_2)\dots p_{k_{n-1}}(j)$ al menos un término de la suma debe ser positivo, lo cual concluye la demostración.

□

Con este resultado podemos indicar si dos estados se comunican en términos de la matriz de transición. Como veremos a continuación, nos interesará particularmente observar las probabilidades de retorno a un estado dado, es decir, las probabilidades de la forma $p_i^n(i)$.

1.4. Periodicidad

Retomando el ejemplo 1.3.1 puede apreciarse que en ese caso $p_4(4) = 0$ pero $p_4^2(4) = p_4(5)p_5(4) > 0$. Es decir, no es posible retornar al estado 4 en un paso, pero sí en dos pasos. Más aún, se tiene que si k es impar $p_4^k(4) = 0$ pero si k es par $p_4^k(4) > 0$ y lo mismo ocurre con el estado 5. Por otra parte, si observamos los estados $i = 1$, 2 o 3 de dicho ejemplo tenemos que las únicas probabilidades de retorno no nulas son las $p_i^k(i)$ cuando k es múltiplo de 3. Esto motiva la siguiente definición:

Definición 1.4.1. Sean $\{X_k\}_{k\in\mathbb{N}}$ una cadena de Markov con espacio de estados E y matriz de transición $P = (p_i(j))$ y $i \in E$ un estado. Definimos $d(i)$ el *periodo* de i como

$$d(i) := m.c.d\{k \geq 1 : p_i^k(i) > 0\}.$$

Es decir, $d(i)$ es el mayor entero positivo que verifica que, si $p_i^n(i) > 0$, entonces dicha probabilidad es múltiplo de $d(i)$. En particular, sólo podemos retornar a i en una cantidad de pasos múltiplo de $d(i)$.

Si $d(i) = 1$ decimos que i es *aperiódico*. En caso contrario se dice que i es periódico y su periodo es $d(i)$.

En el ejemplo anterior no hay estados aperiódicos. Los estados 4 y 5 tienen período 2 y los estados 1, 2 y 3 son de período 3. En efecto, no es casual que los estados que se comunican tengan el mismo período, como puede verse a continuación:

Proposición 1.4.2. *Si dos estados i, j de una cadena de Markov se comunican (i.e., si $i \leftrightarrow j$), tienen el mismo período.*

Demostración. Para demostrar la propiedad veremos que $d(i)$ y $d(j)$ se dividen mutuamente. Para probar que $d(i)$ divide a $d(j)$ veamos primero que si $p_j^n(j) > 0$, $d(i)$ divide a n:

Como $i \to j$, existe $s \geq 1$ tal que $p_i^s(j) > 0$. Asimismo, como $j \to i$ existe $r > 0$ tal que $p_j^r(i) > 0$. En consecuencia

$$p_i^{r+s}(i) = \sum_{k \in E} p_i^s(k) p_k^r(i) \geq p_i^s(j) p_j^r(i) > 0,$$

con lo cual $d(i)$ divide a $r + s$. Además, como $p_j^n(j) > 0$, se tiene que

$$p_i^{r+n+s}(i) \geq p_i^s(j) p_j^n(j) p_j^r(i) > 0,$$

y por lo tanto $d(i)$ divide a $r + n + s$. Luego $d(i)$ divide a n y por la definicion 1.4.1 $d(j) = m.c.d\{n \geq 1 : p_j^n(j) > 0\}$, con lo cual $d(i)$ divide a $d(j)$.

Finalmente basta observar que se pueden intercambiar i y j en el argumento anterior, con lo cual se prueba que $d(j)$ divide a $d(i)$ y por lo tanto $d(i) = d(j)$. □

Como consecuencia de la proposición anterior todos los elementos de una clase irreducible tienen el mismo periodo, con lo cual quedan bien definidos el periodo de una clase o incluso del periodo de una cadena, si la misma es irreducible.

Definición 1.4.3. Sea $\{X_k\}_{k \in \mathbb{N}}$ una cadena de Markov y C una clase irreducible. Definimos $d(C)$ el *periodo* de C como $d(C) := d(i_C)$, con $i_C \in C$ un estado arbitrario. En particular si $d(C) = 1$ se dice que la clase es *aperiódica*.

Si la cadena es irreducible, decimos que su periodo es $d(i)$, siendo i un estado cualquiera de la cadena. Nuevamente, si dicho periodo es 1 diremos que la cadena es *aperiódica*.

Veamos ahora un ejemplo sobre el cual volveremos más adelante: el paseo al azar simple. Para ello consideramos el ejemplo 1.2.4 en el caso particular en que la sucesión de variables i.i.d utilizada en la construcción sólo toma valores 1 o -1. Obtenemos así el siguiente ejemplo:

Ejemplo 1.4.4 (Paseo al azar simple). Sea $\{X_k\}_{k\in\mathbb{N}}$ una cadena de Markov con espacio de estados $\mathbb{Z}$, distribución inicial arbitraria y matriz de transición $P = (p_i(j))_{i,j}$ tal que para cierto $p \in (0,1)$ se tiene que:

$$p_i(j) = \begin{cases} p, & \text{si j = i+1,} \\ 1-p, & \text{si j = i-1,} \\ 0, & \text{en otro caso.} \end{cases}$$

En este caso la cadena es irreducible pero no aperiódica. Para ver que es irreducible basta probar que todos los estados se comunican entre sí. Por la propiedad 1.3.5 esto equivale a verificar que para todo par de estados $i, j \in \mathbb{Z}$ existe una trayectoria con probabilidad positiva que comunica i con j.

Si $j > i$, una posible trayectoria es $i, i+1, \ldots, j$ cuya probabilidad es $p_i(i+1)p_{i+1}(i+2)\ldots p_{j-1}(j) = p^{j-i} > 0$. Análogamente si $j < i$ tenemos $i, i-1, \ldots, j+1, j$ con probabilidad $(1-p)^{i-j} > 0$. Si $i = j$ basta tomar la trayectoria $i, i+1, i$ cuya probabilidad también es positiva. Así, todos los estados se comunican y la cadena es irreducible.

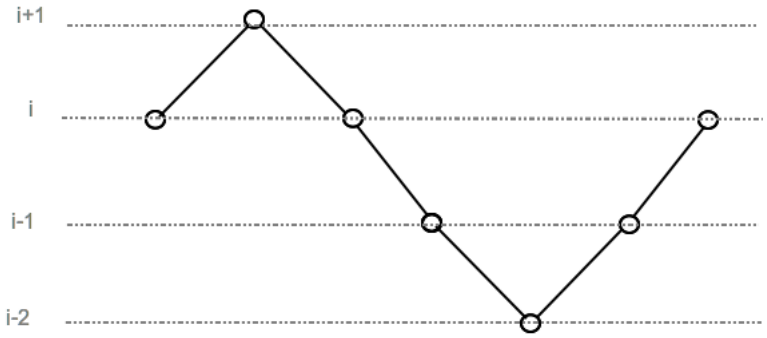


Figura 1.1: Ejemplo de trayectoria que va del estado i a i en 2 y 6 pasos. Observar que hay 2 transiciones involucrando los estados i e $i+1$, 2 que involucran a i e $i-1$ y 2 a $i-1$ e $i-2$.

Por otra parte se observa que $\{X_k\}$ tiene periodo 2, es decir, que solo es posible retornar a un estado dado en un número par de pasos. Para ello consideremos una trayectoria de i a i. Como sólo es posible moverse a los enteros más próximos, se tiene que para cada transición de k a $k+1$ tiene que haber un retorno de $k+1$ a k y viceversa, lo cual implica una cantidad siempre par de transiciones.

Una modificación posible a este ejemplo para obtener una cadena irreducible y aperiódica consiste en permitir transiciones de un estado a sí mismo, es decir, redefiniendo las entradas de la matriz de transición como

$$p_i(j) = \begin{cases} p, & \text{si j = i+1,} \\ q, & \text{si j = i-1,} \\ r, & \text{si i = j,} \\ 0, & \text{en otro caso,} \end{cases}$$

donde $p + q + r = 1$.

Es fácil probar en este caso que la cadena es irreducible y aperiódica.

Nótese que si bien en la modificación anterior permitimos transiciones de i a i para cualquier i, a los efectos de obtener una cadena irreducible y aperiódica hubiese bastado con permitirlas para un **único** estado cualquiera, como se verá a continuación.

Proposición 1.4.5. *Sea $\{X_k\}_{k\in\mathbb{N}}$ una cadena de Markov irreducible. Si existe un estado i tal que $p_i(i) > 0$, entonces la cadena es aperiódica.*

Demostración. Sea j un estado. Se quiere verificar que $d(j) = 1$. Como la cadena es irreducible se tiene que $i \leftrightarrow j$ y por lo tanto existen r y s positivos tales que $p_i^s(j) > 0$ y $p_j^r(i) > 0$. Luego como $p_j^{r+s}(j) \geq p_j^r(i)p_i^s(j) > 0$ podemos afirmar que $d(j)$ divide a $r + s$.

Por otra parte, como $p_i(i) > 0$, se cumple que $p_j^{r+s+1}(j) \geq p_j^r(i)p_i(i)p_i^s(j) > 0$ con lo cual $d(j)$ divide a $r + s + 1$. Es decir, $d(j)$ es divisor común de dos números coprimos y por lo tanto $d(j) = 1$.

□

1.5. Recurrencia y tiempos de pasaje

Una de las nociones centrales de este apartado es el tiempo de pasaje o retorno por un estado dado, que comenzaremos definiendo.

Definición 1.5.1. Sea $\{X_k\}_{k\in\mathbb{N}}$ una cadena de Markov con distribución inicial arbitraria.

- Denominamos *tiempo del primer pasaje por j* a la variable aleatoria

$$\tau_j := \inf_{n\in\mathbb{N}}\{n \geq 1 : X_n = j\}.$$

  Si dicho conjunto es vacío definimos $\tau_j = \infty$.

  En el caso particular en que $P(X_0 = j) = 1$ llamamos a dicha variable *tiempo del primer retorno a j.*

- La *probabilidad del primer pasaje por j en tiempo n, partiendo de i* se define como:

$$f_{ij}^n := P(\tau_j = n | X_0 = i).$$

  Nuevamente si $i = j$ hablaremos de *retorno* a j en vez de pasaje.

- Llamamos *probabilidad de visitar j partiendo de i* (sin importar el instante) a

$$f_{ij} = P(\bigcup_{n\geq 1} \tau_j = n | X_0 = i).$$

En términos de las trayectorias, es fácil ver que $f_{ij}^n = P(X_n = j, X_{n-1} \neq j, \ldots, X_1 \neq j | X_0 = i)$. Además como los sucesos $\{\tau_j = n\} := \{X_n = j, X_{n-1} \neq j, \ldots, X_1 \neq j\}$ son disjuntos, se tiene que

$$f_{ij} = P(\bigcup_{n\geq 1} \{\tau_j = n\} | X_0 = i) = \sum_{n\geq 1} P(\{\tau_j = n\} | X_0 = i) = \sum_{n\geq 1} f_{ij}^n.$$

Por otra parte los sucesos $\{\tau_j = n\}$ están incluidos en $\{X_n = j\}$ con lo cual $f_{ij}^n \leq p_i^n(j), \forall n \geq 1, i, j \in E$.

Las definiciones anteriores nos permiten clasificar los estados de acuerdo su tiempo de retorno.

Definición 1.5.2. Sean $\{X_k\}_{k\in\mathbb{N}}$ una cadena de Markov con espacio de estados E e $i \in E$ un estado. Diremos que i es *recurrente* si $f_{ii} = 1$. En caso contrario (es decir, $f_{ii} < 1$) se dice que i es *transitorio*.

Intuitivamente, los estados recurrentes son aquellos a los que, con probabilidad 1, se vuelve en algún tiempo finito.

Veamos ahora un par de resultados que permiten, además de vincular la noción de recurrencia/transitoriedad con los conceptos vistos previamente, estudiar ciertos aspectos del comportamiento asintótico de las cadenas de Markov. Para ello primero necesitaremos el siguiente lema:

Lema 1.5.3. *Sea $\{X_k\}_{k\in\mathbb{N}}$ una cadena de Markov con espacio de estados E y matriz de transición $P = (p_i(j))_{i,j\in E}$. Se verifica:*

$$p_i^n(j) = \sum_{k=1}^{k=n} f_{ij}^k p_j^{n-k}(j) \quad \forall n \geq 1, \forall i, j \in E.$$

En particular, observar que $f_{ij} > 0$ si y sólo si $i \to j$.

Demostración. Notemos primero que los sucesos $\{\tau_j = k\}_{k\in\{1,\ldots n\}}$ son disjuntos y cubren al suceso $\{X_n = j\}$ con lo cual

$$p_i^n(j) = P(X_n = j | X_0 = i) = \sum_{k=1}^{k=n} P(X_n = j, \tau_j = k | X_0 = i) =$$

$$= \sum_{k=1}^{k=n} P(X_n = j, | \tau_j = k, X_0 = i) P(\tau_j = k | X_0 = i) = p_j(j)^{n-k} f_{ij}^k.$$

□

Ahora estamos en condiciones de probar el siguiente teorema:

Teorema 1.5.4. *Sea $\{X_k\}_{k\in\mathbb{N}}$ una cadena de Markov con espacio de estados E y matriz de transición $P = (p_i(j))_{i,j}$. Se verifican:*

1. *Un estado i es recurrente si y sólo si $\sum_{n\geq 1} p_i^n(i) = +\infty$.*
2. *La recurrencia es invariante por clases irreducibles, es decir, dados i, j tales que $i \leftrightarrow j$ se tiene que i es recurrente si y sólo si j lo es.*
3. *Si $i \to j$ y j es recurrente, se tiene que $\sum_{n\geq 1} p_i^n(j) = +\infty$.*
4. *Si j es transitorio $\sum_{n\geq 1} p_i^n(j) < \infty, \forall i \in E$. En particular $\lim_{n\to+\infty} p_i^n(j) = 0$.*

Demostración. Una observación general de utilidad es que, si $\sum_{n\geq 1} p_i^n(j) < +\infty$, al escribirlo como $\sum_{n\geq 1}(\sum_{m=1}^{m=n} f_{ij}^m p_j^{n-m}(j))$ se puede cambiar el orden de la suma y, usando el lema anterior resulta:

$$\begin{aligned}\sum_{n\geq 1} p_i^n(j) = \sum_{n\geq 1}(\sum_{m=1}^{m=n} f_{ij}^m p_j^{n-m}(j)) = \sum_{m\geq 1}(\sum_{n=m}^{n=+\infty} f_{ij}^m p_j^{n-m}(j)) \\ = \sum_{m\geq 1} f_{ij}^m \sum_{n=0}^{n=+\infty} p_j^n(j) = f_{ij}(1 + \sum_{n\geq 1} p_j^n(j)).\end{aligned} \tag{1.3}$$

Procedamos ahora a probar los ítems:

1. Si $\sum_{n\geq 1} p_i^n(i) < +\infty$ podemos usar la observación anterior obteniendo:

$$f_{ii} = \frac{\sum_{n\geq 1} p_i^n(i)}{\sum_{n\geq 1} p_i^n(i) + 1} < 1,$$

con lo cual i es transitorio.

Para la otra implicancia consideremos $\sum_{n\geq 1} p_i^n(i) = +\infty$. Veamos que $f_{ii} = 1$ (i.e, que i es recurrente). Consideremos la sucesión $(a_N)_{N\in\mathbb{N}}$ definida como $a_N := \sum_{n=1}^{n=N} p_i^n(i)$. Así:

$$a_N = \sum_{n=1}^{n=N} p_i^n(i) = \sum_{n=1}^{n=N}(\sum_{m=1}^{m=n} f_{ii}^m p_i^{n-m}(i)) = \sum_{m=1}^{m=N}\sum_{n=m}^{n=N} f_{ii}^m p_i^{n-m}(i) \leq$$
$$\leq \sum_{m=1}^{m=N} f_{ii}^m \sum_{n=0}^{n=N} p_i^n(i) = \sum_{m=1}^{m=N} f_{ii}^m(1+a_N).$$

Despejando lo anterior y utilizando la definición de f_{ii} tenemos entonces:

$$f_{ii} \geq \sum_{m=1}^{m=N} fii^m \geq \frac{a_N}{1+a_N} \xrightarrow{N\to\infty} 1.$$

Recordar que por hipótesis $a_N \to +\infty$ cuando $N \to \infty$. Luego se tiene que $f_{ii} = 1$ y por lo tanto i es recurrente lo cual prueba (1).

2. Sean i, j dos estados que se comunican, i recurrente. Como $i \leftrightarrow j$ existen naturales r y s tales que $p_i^r(j) > 0$ y $p_j^s(i) > 0$. Además, para cada $n \in \mathbb{N}, n \geq 1$ se tiene que $p_j^{n+r+s}(j) \geq p_i^r(j)p_i^n(i)p_j^s(i)$, donde por la recurrencia de i $p_i^n(i)$ es el término principal de una serie divergente. En consecuencia:

$$\sum_{n\geq 1} p_j^n(j) \geq \sum_{n\geq 1} p_j^{n+r+s}(j) \geq p_i^r(j)p_j^s(i)\sum_{n\geq 1} p_i^n(i) = +\infty,$$

lo cual -por el ítem anterior, implica que j es recurrente.

3. Recordemos primero que, del lema 1.5.3 se desprende que $f_{ij} > 0$ si y sólo si $i \to j$. Así, como $i \to j$ tenemos que $\sum_{n\geq 1} f_{ij}^n > 0$ por lo que para algún $n_0 \geq 1$ se cumple $f_{ij}^{n_0} > 0$. Aplicando nuevamente el lema para $n \geq n_0$ se tiene que $p_i^n(j) \geq f_{ij}^{n_0} p_j^{n-n_0}(j)$ con lo cual

$$\sum_{n=n_0}^{n=N} p_i^n(j) \geq f_{ij}^{n_0} \sum_{n=n_0}^{n=N} p_j^{n-n_0}(j) \xrightarrow{N\to\infty} +\infty,$$

donde la divergencia se debe a que j es recurrente. Se concluye que la serie $\sum_n p_i^n(j)$ es divergente lo cual prueba (3).

4. Como j es transitorio $\sum_{n\geq 1} p_j^n(j) < \infty$. Luego

$$\sum_{n\geq 1} p_i^n(j) = \lim_{N\to\infty} \sum_{n=1}^{n=N} p_i^n(j) = \sum_{n=1}^{n=N} (\sum_{m=1}^{m=n} f_{ij}^m p_j^{n-m}(j)) =$$
$$= \sum_{m=1}^{m=N} (\sum_{n=m}^{n=N} f_{ij}^m p_j^{n-m}(j)) \leq \sum_{m=1}^{m=N} f_{ij}^m \sum_{n=0}^{n=N-m} p_j^n(j) =$$
$$= \sum_{m=1}^{m=N} f_{ij}^m (1 + \sum_{n=1}^{n=N-m} p_j^n(j)) = f_{ij}(1 + \sum_{n=1}^{n=N-m} p_j^n(j)) < \infty$$

por ser j transitorio.

Finalmente, $\lim_{n\to\infty} p_i^n(j) = 0$ es una consecuencia directa de la convergencia de la serie anterior, con lo cual quedó demostrado el cuarto y último ítem.

□

Cabe observar que en el teorema recién demostrado se presentó un resultado que permite describir las probabilidades asintóticas de visitar un estado transitorio. Más adelante estudiaremos el comportamiento asintótico para los estados recurrentes.

Ejemplo. Para finalizar esta sección estudiemos la recurrencia en el paseo al azar simple (ejemplo 1.4.4). Para ello recordemos que tenemos una cadena de Markov en $\mathbb{Z}$ cuyas probabilidades de transición son

$$p_i(j) = \begin{cases} p, & \text{si j = i+1,} \\ 1-p, & \text{si j = i-1,} \\ 0, & \text{en otro caso.} \end{cases}$$

Ya vimos que la cadena es irreducible con lo que, por el teorema recién probado, basta estudiar la recurrencia en un sólo estado (elegiremos el 0) y los demás estados tendrán el mismo comportamiento. Además, como la cadena tiene período 2 se tiene que $p_0^{2n+1}(0) = 0, \forall n \in \mathbb{N}$. Calculemos ahora las probabilidades de la forma $p_0^{2n}(0)$, para lo cual hay que observar que una trayectoria que va de 0 a 0 en $2n$ pasos sube exactamente n veces (y baja otras n). Por lo tanto cada trayectoria tiene probabilidad $p^n(1-p)^n$ y hay que contar la cantidad de trayectorias posibles.

Para ello, consideremos que tenemos $2n$ transiciones y hay que elegir n de ellas (que serán, por ejemplo, las ascendentes). Tenemos así C_n^{2n} trayectorias posibles y por lo tanto

$$p_0^{2n}(0) = C_n^{2n} p^n (1-p)^n.$$

Utilizando la *fórmula de Stirling*, que establece que $n! = n^n e^{-n}\sqrt{2\pi n}(1+\alpha_n)$ con $\alpha_n \to 0$ en el número combinatorio tenemos que

$$C_n^{2n} = \frac{2^{2n}(1+\alpha_n)}{\sqrt{\pi n}(1+\beta_n)^2}.$$

Luego sustituyendo se tiene que $p_0^{2n}(0) = \frac{(4p(1-p))^n}{\sqrt{n\pi}}\delta_n$, con $\delta_n = \frac{1+\alpha_n}{(1+\beta_n)^2} \to 1$.

Por otra parte $p(1-p)$ se maximiza cuando $p = \frac{1}{2}$ (paseo al azar simétrico). En tal caso se tiene:

$$p_0^{2n}(0) = \frac{1}{\sqrt{\pi n}}\delta_n,$$

con lo cual la serie $\sum_{n=0}^{n=+\infty} p_0^n(0)$ diverge y por lo tanto 0 (y todos los estados de la cadena) son recurrentes.

En cambio, si $p \neq \frac{1}{2}$ tenemos que $p_0^{2n}(0) = \frac{c}{\sqrt{\pi n}}\delta_n$ con $c = 4p(1-p) < 1$ que es el término principal de una serie convergente, con lo cual los estados son transitorios.

1.6. Comportamiento asintótico de las cadenas de Markov

Estudiaremos ahora el comportamiento de una cadena de Markov tras muchas transiciones. Concretamente nos interesa saber cómo son las probabilidades de transición y la distribución de probabilidad de orden n cuando n es muy grande. Es decir, queremos conocer

$$\lim_{n\to\infty} \mu^n \quad \text{y} \quad \lim_{n\to\infty} p_i^n(j) \ \forall i, j \in E.$$

Los cuales, a priori, podrían no existir. También habrá que discutir entonces las condiciones para que dichos límites existan. Nótese que en la sección anterior se estudió el comportamiento límite de $p_i^n(j)$ cuando j es transitorio, caso en el cual dicho límite es 0.

Para estudiar los estados recurrentes, comenzaremos con una defnición.

Definición 1.6.1. Sea $\{X_k\}_{k\in\mathbb{N}}$ una cadena de Markov con espacio de estados E y matriz de transición P. Si $\lambda = (\lambda_i)_{i\in E}$ es un vector de probabilidades tal que $\lambda P = \lambda$, decimos que λ es una *distribución estacionaria* (o invariante) para la cadena.

Observación.
- La condición de ser estacionaria depende únicamente de la matriz P, con lo cual puede hablarse también de distribución estacionaria para P en lugar de la cadena.
- Si λ es una distribución estacionaria es inmediato que $\lambda P^n = \lambda$, $\forall n \in \mathbb{N}$, lo cual explica el uso del término *invariante*.

En términos del comportamiento asintótico nos interesan las distribuciones estacionarias ya que nos dan información sobre el mismo, como veremos a continuación:

Proposición 1.6.2. *Consideremos una cadena de Markov con espacio de estados E* ***finito****. Si existen $i \in E$ y $\mu = (\mu_j)_{j\in E}$ tales que*

$$p_i^n(j) \xrightarrow[n\to\infty]{} \mu_j \ \forall j \in E,$$

entonces μ es una distribución estacionaria.

Demostración. Primero observemos que, efectivamente μ es una distribución (i.e, sus entradas suman 1):

$$\sum_{j\in E}\mu_j = \sum_{j\in E}\lim_{n\to\infty} p_i^n(j) \overset{\text{E finito}}{=} \lim_{n\to\infty}\sum_{j\in E} p_i^n(j) = 1,$$

donde la última igualdad se debe además a que las matrices P^n son estocásticas.

Veamos ahora que $\mu P = \mu$. Para ello será de utilidad recordar que $p_i^n(j) = \sum_{k\in\mathbb{E}} p_i^{n-1}(k)p_k(j)$ con lo cual

$$\mu_j = \lim_{n\to\infty} p_i^n(j) = \lim_{n\to\infty}\sum_{k\in\mathbb{E}} p_i^{n-1}(k)p_k(j) \overset{\text{E finito}}{=} \sum_{k\in\mathbb{E}}\lim_{n\to\infty} p_i^{n-1}(k)p_k(j) = \sum_{k\in\mathbb{E}}\mu_k p_k(j),$$

dado que por hipótesis $p_i^{n-1}(k) \to \mu_k$. Luego matricialmente tenemos que $\mu = \mu P$ concluyendo la demostración. □

Cabe notar que la finitud del espacio de estados –que se utilizó para intercambiar límite y suma– no es prescindible. Un ejemplo de ello se obtiene si consideramos el paseo al azar simple con $p \neq \frac{1}{2}$. Como se vio anteriormente todos sus estados son transitorios y por lo tanto $p_i^n(j) \xrightarrow[n\to\infty]{} 0\ \forall j \in E$. Sin embargo, el vector nulo no puede ser una distribución.

Obtuvimos pues una primera –y bastante limitada– relación entre distribuciones asintóticas y estacionarias. Sin embargo aún no presentamos ningún resultado que afirme que las distribuciones estacionarias son distribuciones límite (lo cual resultaría de utilidad para el cálculo de estas últimas). A continuación veremos una serie de resultados más generales que nos permitirán enunciar y demostrar tal propiedad.

Definición 1.6.3. Sea $\{X_k\}_{k\in\mathbb{N}}$ una cadena de Markov con espacio de estados E, matriz de transición P y distribución inicial μ.

Dado un estado i se define el *tiempo medio de retorno a i* como

$$\gamma_i := \sum_{n\geq 1} f_{ii}^n = E_i\tau_i,$$

donde E_i refiere a la esperanza de la variable, condicionada a $X_0 = i$.

Cuando γ_i es finito se dice que el estado i es *recurrente positivo*, mientras que si $\gamma_i = \infty$ decimos que i es *recurrente nulo*.

Teorema 1.6.4. *Consideremos una cadena de Markov con espacio de estados E, $i \in E$ un estado recurrente y aperiódico. Se tiene que*

$$\lim_{n\to\infty} p_i^n(i) = \frac{1}{\gamma_i},$$

con γ_i como en la definición anterior si es finito. Si $\gamma_i = \infty$ consideramos $\frac{1}{\gamma_i} = 0$.

Para demostrarlo utilizaremos el siguiente lema, que probaremos al final.

Lema 1.6.5. *Sea $d_f(i) := m.c.d\{n : f_{ii}^n > 0\}$. Se verifica que $d(i) = d_f(i)$. En otras palabras, para calcular el periodo de un estado se puede considerar indistintamente el conjunto de los $p_i^n(i)$ o los f_{ii}^n.*

Demostración del teorema 1.6.4. Dividamos la demostración en varios pasos.

Paso 1

Definamos $a_n = \sum_{m\geq n} f_{ii}^m$ (como i está fijo de antemano omitimos dicho índice en a_n). Así, por ejemplo, tenemos que $f_{ii}^n = a_n - a_{n+1}$ $\forall n \in \mathbb{N}$ y, desarrollando la suma en la definición de γ_i:

$$\gamma_i = \sum_{n=1}^{\infty} n f_{ii}^n = \sum_{n=1}^{\infty}\sum_{m=1}^{m=n} f_{ii}^n = \sum_{m=1}^{\infty}\sum_{n=m}^{\infty} f_{ii}^n \overset{\text{def } a_n}{=} \sum_{m=1}^{\infty} a_m.$$

Asimismo, utilizando el lema 1.5.3 y escribiendo f_{ii}^n en función de a_n se tiene que

$$p_i^n(i) = \sum_{m=1}^{m=n} f_{ii}^m p_i^{n-m}(i) = \sum_{m=1}^{m=n} (a_m - a_{m+1}) p_i^{n-m}(i).$$

Notemos que por la recurrencia de i y la definición de a_n se tiene que $a_1 = 1$. Así despejando y escribiendo la suma anterior de forma más explícita resulta:

$$p_i^n(i)a_1 + p_i^{n-1}(i)a_2 + \ldots p_i^0(i)a_{n+1} = p_i^{n-1}(i)a_1 + p_i^{n-2}(i)a_2 + \ldots p_i^0(i)a_n,$$

donde, si llamamos ϕ_n al miembro izquierdo, la igualdad anterior puede expresarse como $\phi_n = \phi_{n-1}$. Además, tomando $n = 1$ y recordando que $p_i^0(i) = 1$, se verifica que $\phi_1 = p_i^1(i)a_1 + p_i^0(i)a_2 = p_i^0(i)a_1 = 1$ y por inducción resulta $\phi_n = 1$ $\forall n \geq 1$, es decir:

$$p_i^n(i)a_1 + p_i^{n-1}(i)a_2 + \ldots p_i^0(i)a_{n+1} = 1. \tag{1.4}$$

Paso 2

Consideremos $\alpha := \limsup_n p_i^n(i)$ y $\{n_m\}$ una sucesión de índices tal que $p_i^{n_m}(i) \xrightarrow[n\to\infty]{} \alpha$. Así, para todo s tal que $f_{ii}^s > 0$ se tiene:

$$\begin{aligned}
\alpha = \liminf_m p_i^{n_m}(i) &= \liminf_m \{f_{ii}^s p_i^{n_m - s}(i) + \sum_{r=1, r\neq s}^{n_m} f_{ii}^r p_i^{n_m - r}(i)\} \\
&\leq f_{ii}^s \liminf_m p_i^{n_m - s}(i) + \limsup_m \{\sum_{r=1, r\neq s}^{n_m} f_{ii}^r p_i^{n_m - r}(i)\} \\
&\leq f_{ii}^s \liminf_m p_i^{n_m - s}(i) + \sum_{r=1, r\neq s}^{n_m} f_{ii}^r \limsup_m p_i^{n_m - r}(i)\} \\
&\overset{*}{\leq} f_{ii}^s \liminf_m p_i^{n_m - s}(i) + (1 - f_{ii}^s)\alpha.
\end{aligned}$$

donde en para la desigualdad $*$ se usó que $\sum_{k=1}^{\infty} f_{ii}^k = 1$ y que $\limsup_m p_{n_k} \leq \alpha$ para toda subsucesión de índices $\{n_k\}$. Si despejamos la desigualdad obtenemos

$$\begin{aligned}
\alpha \leq f_{ii}^s \liminf_m p_i^{n_m - s}(i) + \alpha - f_{ii}^s \alpha &\iff \\
0 \leq f_{ii}^s (\liminf_m p_i^{n_m - s}(i) - \alpha) &\iff \\
\alpha \leq \liminf_m p_i^{n_m - s}(i).
\end{aligned}$$

Y por su definición, $\alpha = \limsup_m p_i^{n_m - s}(i)$ con lo cual se concluye que:

$$\alpha = \lim_{m\to\infty} p_i^{n_m - s}(i), \tag{1.5}$$

para todo s tal que $f_s > 0$.

Paso 3

Ahora queremos probar que existe $s' \in \mathbb{N}$ tal que $\alpha = \lim_{m\to\infty} p_i^{n_m - s}(i),\ \forall s \geq s'$. Nótese que en la parte anterior probamos esta igualdad para todos los índices s tales que $f_{ii}^s > 0$.

Como i es aperiódico es posible considerar una subfamilia $\{s_1, s_2, \ldots, s_r\}$ de índices coprimos contenida en $\{n : f_{ii}^n > 0\}$. Por lo probado en el paso 2, para cada uno de estos índices s_k se tiene que:

$$\alpha = \lim_{m\to\infty} p_i^{n_m - s_k}(i).$$

Si elegimos $s' = \prod_{k=1}^{k=r} s_k$, tenemos que todo $s \geq s'$ es de la forma $\sum s_k t_k$ con t_k naturales. Probemos entonces que la igualdad (1.5) es válida para los s de la forma $s_k t_k$, con $k \in \{1, \ldots, r\}$. Para ello basta notar que los $\{n_m - s_k\}_{m \geq 1}$ verifican $\alpha = \lim_{m \to \infty} p_i^{n_m - s_k}(i)$, con lo cual puede aplicárseles nuevamente el procedimiento usado en el paso 2 obteniendo $\alpha = \lim_{m \to \infty} p_i^{n_m - 2s_k}(i)$.

Aplicando sucesivas veces el mismo razonamiento se tiene que $\alpha = \lim_{m \to \infty} p_i^{n_m - t_k s_k}(i)\ \forall t_k \in \mathbb{N}$. De igual modo, se extiende a la suma $\sum_{k=1}^{k=r} t_k s_k$ resultando así:

$$\alpha = \lim_{m \to \infty} p_i^{n_m - s}(i) \quad \forall s \geq s',$$

que es lo que se quería probar.

Paso 4

Sea $s \geq s'$. Aplicando la ecuación 1.4 y tomando sólo parte de su miembro izquierdo, se tiene:

$$p_i^{n_m - s'}(i) a_1 + p_i^{n_m - (s'+1)}(i) a_2 + \ldots + p_i^{n_m - (s'+s)}(i) a_{s+1} \leq 1,$$

donde $p_i^{n_m - (s'+k)} \xrightarrow{m} \alpha\ \forall k \in \{0, \ldots, s\}$ por lo visto en pasos anteriores. Tomando límites en la desigualdad anterior se cumple que:

$$\alpha(a_1 + \ldots + a_{s+1}) \leq 1,$$

para todo $s \geq s'$. Así, si i es recurrente nulo, se tiene que $\gamma_i = \sum_{n=1}^{\infty} a_n = \infty$ y además $\alpha \sum_{n=1}^{\infty} a_n \leq 1$ con lo cual $\alpha = 0$.

Esto prueba el teorema en el caso en que i es recurrente nulo, ya que $\alpha = \limsup_m p_i^n(i) = 0$. Para el caso $\gamma_i < \infty$ nos queda la desigualdad

$$\limsup_m p_i^n(i) = \alpha \leq \frac{1}{\gamma_i}. \tag{1.6}$$

Paso 5 Para probar el teorema cuando $\gamma_i < \infty$, definamos $\beta := \liminf_n p_i^n(i)$ y probemos que $\beta \geq \frac{1}{\gamma_i}$. Para ello basta repetir lo hecho en los pasos 2, 3 y 4 con α y obtenemos que $s'' \in \mathbb{N}$ de modo que $\lim p_{n_m - s} = \beta\ \forall s \geq s''$. Tomando $n = n_m - s''$ en la ecuación 1.4 se obtiene:

$$
\begin{aligned}
1 &\leq p_i^{n_m-s''}(i)a_1 + p_i^{n_m-(s''+1)}(i)a_2 + \dots p_i^{0}(i)a_{n_m-s''+1} = \\
&= p_i^{n_m-s''}(i)a_1 + p_i^{n_m-(s''+1)}(i)a_2 + \dots p_i^{n_m-(s+s'')}(i)a_{s+1} + \sum_{k=s+2}^{k=n_m-s''+1} p_i^{n_m-(s''+k-1}(i)a_k \\
&\leq p_i^{n_m-s''}(i)a_1 + p_i^{n_m-(s''+1)}(i)a_2 + \dots p_i^{n_m-(s+s'')}(i)a_{s+1} + \sum_{k=s+2}^{\infty} a_k.
\end{aligned}
$$

Si $m \to \infty$ en la ecuación anterior resulta:

$$1 \leq \beta(a_1 + \dots + a_{s+1}) + \sum_{k=s+2}^{\infty} a_k.$$

Donde, por ser γ_i finito, $\sum_{k=s+2}^{\infty} a_k$ es la cola de una serie convergente, con lo cual al tomar $s \to \infty$ tenemos:

$$\frac{1}{\gamma_i} = \frac{1}{\sum_{k=1}^{\infty} a_k} \leq \beta = \liminf_n p_i^n(i).$$

Finalmente, de esta igualdad y la ecuación (1.6) se tiene que existe $\lim_n p_i^n(i)$ y es $\frac{1}{\gamma_i}$, lo cual concluye la demostración. □

Demostración del lema 1.6.5. Para probar que $d(i) = d_f(i)$ veremos que $d_f(i)$ divide a $d(i)$ y viceversa.

Como $f_{ii}^n \leq p_i^n(i)$ se verifica que $\{n : f_{ii}^n > 0\} \subset \{n : p_i^n(i) > 0\}$ y por lo tanto $d(i)$ divide a $d_f(i)$. Para probar que $d_f(i)$ divide a $d(i)$ consideremos n tal que $p_i^n(i) > 0$ y veamos que $d_f(i)$ divide a n, lo cual es inmediato si $f_{ii}^n > 0$.

Si $f_{ii}^n = 0$, significa que hay una trayectoria de i a i en n pasos que pasa por i en algún instante intermedio, es decir, existen n_1 y n_2 tales que $n_1 + n_2 = n$ y $p_i^{n_1}(i)p_i^{n_2}(i) > 0$. Si $f_{ii}^{n_1}$ y $f_{ii}^{n_2}$ no son ambos positivos se repite el procedimiento hasta encontrar una secuencia de naturales $n_1 \dots, n_m$ que suman n y tales que $f_{ii}^{n_k} > 0 \ \forall k \in \{1, \dots, m\}$. Nótese que, por la definición de f_{ii}^k, dicha secuencia existe.

Como $d_f(i)$ divide a todos los n_k recién definidos, también divide a n y por lo tanto $d_f(i)$ divide a $d(i)$ terminando la demostración.

□

Con el teorema anterior conocemos el comportamiento asintótico de las probabilidades de la forma $p_i^n(i)$ cuando el estado i es recurrente y aperiódico. Para extenderlo a las cadenas veremos primero que ser recurrente positivo/nulo es una propiedad invariante por clase irreducible.

Proposición 1.6.6. *Sean i, j dos estados de una cadena de Markov tales que $i \leftrightarrow j$. i es recurrente positivo (nulo) si y sólo si j es recurrente positivo (nulo).*

Demostración. Consideremos r y s naturales tales que $p_i^r(j) > 0$ y $p_j^s(i) > 0$. Así, dado $n \in \mathbb{N}$ vale:

$$p_j^{s+n+r}(i) \geq p_j^s(i) p_i^n(i) p_i^r(j).$$

Tomando límite cuando $n \to \infty$ en ambos lados se obtiene así que si $\lim_{n\to\infty} p_i^n(i) > 0$ entonces $\lim_{n\to\infty} p_j^n(j) > 0$. Es decir, si i es recurrente positivo, j también lo es. Luego usando el mismo argumento se prueba que si j es recurrente positivo i también lo es.

□

Ahora estamos en condiciones de resumir el comportamiento asintótico en general.

Teorema 1.6.7. *Sea $\{X_k\}_{k\in\mathbb{N}}$ una cadena de Markov irreducible y aperiódica con espacio de estados E, matriz de transición P y distribución inicial μ. Se cumple exactamente una de las tres condiciones:*

1. *La cadena es transitioria y en ese caso para todo par de estados i, j se tiene que $\lim_n p_i^n(j) = \lim_n \mu_j^n = 0$. Además $\sum_{n=1}^{\infty} p_i^n(j) < \infty$.*
2. *La cadena es recurrente nula y para todo par $i, j \in E$ se tiene también que $\lim_n p_i^n(j) = \lim_n \mu_j^n = 0$, pero $\sum_{n=1}^{\infty} p_i^n(j) = \infty$.*
3. *La cadena es recurrente positiva y para todo par $i, j \in E$ se tiene $\lim_n p_i^n(j) = \lim_n \mu_j^n = \dfrac{1}{\gamma_j} > 0$.*

Demostración. Notemos que, por ser la cadena irreducible, tiene sentido decir que la misma es transitoria/recurrente positiva/recurrente nula. Por definición de estas tres condiciones, es inmediato que se cumple una y sólo una de las tres, restando únicamente probar los límites. Veremos ahora el caso en que la cadena es transitoria.

En el teorema 1.5.4 se probó que $p_i^n(j) \xrightarrow{n} 0$ y la convergencia de la serie. Probemos que $\mu_j^n \xrightarrow{n} 0$:

$$\lim_n \mu_j^n = \lim_n \sum_{k\in E} \mu_k p_k^n(j) \overset{(*)}{=} \sum_{k\in E} \mu_k \lim_n p_k^n(j) = 0, \tag{1.7}$$

donde en $(*)$ se utilizó el teorema de convergencia dominada para intercambiar el límite con la suma. Queda así probado el primer ítem.

En el caso en que la cadena es recurrente, usando nuevamente el teorema 1.5.4 tenemos que $\sum_{n=1}^{\infty} p_i^n(j) = \infty$. Luego como por el teorema 1.6.4 se sabe que $p_j^n(j) \xrightarrow{n} 0$ tenemos:

$$p_i^n(j) = \sum_{m=1}^{\infty} f_{ij}^m p_j^{n-m}(j),$$

considerando $p_j^m(j) = 0$ si m es negativo. Tomando límites y usando nuevamente convergencia dominada resulta:

$$\lim_n p_i^n(j) = \lim_n \sum_{m=1}^{\infty} f_{ij}^m p_j^{n-m}(j) = \sum_{m=1}^{\infty} f_{ij}^m \lim_n p_j^{n-m}(j) = 0,$$

ya que $\sum_{m=1}^{\infty} f_{ij}^m = 1$. Para ver que $\mu_j^n \xrightarrow{n} 0$ basta volver a utilizar la ecuación (1.7) y queda demostrado el segundo ítem.

Cuando la cadena es recurrente positiva, como ya vimos que $\lim_n p_i^n(j) = \sum_{m=1}^{\infty} f_{ij}^m \lim_n p_j^{n-m}(j)$ con $\sum_{m=1}^{\infty} f_{ij}^m = 1$ y esta vez $p_j^n(j) \xrightarrow{n} \frac{1}{\gamma_j}$, se tiene:

$$\lim_n p_i^n(j) = \frac{1}{\gamma_j}.$$

Luego se deduce que $\lim_n \mu_j^n = \dfrac{1}{\gamma_j}$ de igual modo que en los casos anteriores.

□

Con este resultado podemos conocer el comportamiento asintótico de una cadena de Markov. Sin embargo en la práctica sería de más utilidad vincular la distribución límite con la estacionaria como se hizo en el caso finito, puesto que la existencia de distribuciones estacionarias puede ser más fácil de estudiar en algunos situaciones. Para ello veamos el siguiente teorema:

Teorema 1.6.8. *Sea $\{X_k\}_{k\in\mathbb{N}}$ una cadena de Markov irreducible y aperiódica con espacio de estados E y matriz de transición $P = (p_i(j))_{i,j\in E}$. La cadena es recurrente positiva si y sólo si tiene una distribución estacionaria $\nu = (\nu_i)_{i\in E}$. En ese caso $\nu_i = \frac{1}{\gamma_i}$ $\forall i \in E$ con lo cual la distribución estacionaria es única y coincide con la distribución límite.*

Demostración. Si la cadena es recurrente positiva, verifiquemos que ν, con $\nu_i := \frac{1}{\gamma_i}$, es una distribución estacionaria. Para ello notemos primero que:

$$\sum_{j\in E}\nu_j \overset{(1.6.7)}{=} \sum_{j\in E}\lim_n p_i^n(j) = \sum_{j\in E}\liminf_n p_i^n(j) \overset{\text{Fatou}}{\leq} \liminf_n \sum_{j\in E} p_i^n(j) = 1. \tag{1.8}$$

De modo similar hallemos las entradas de $\nu \times P$:

$$\sum_{i\in E}\nu_i p_i(j) = \sum_i \lim_n p_k^n(i)p_i(j) \leq \liminf_n \sum_i p_k^n(i)p_i(j) = \liminf_n p_k^{n+1}(j) = \nu_j.$$

Y para verificar que vale la otra desigualdad, es decir, $\sum_{i\in E}\nu_i p_i(j) \geq \nu_j$ supongamos que para algún j' esto no es cierto y entonces sumando en j se tiene:

$$\sum_{j\in E}\nu_j > \sum_{j\in E}\sum_{i\in E}\nu_i p_i(j) = \sum_{i\in E}\nu_i(\sum_{j\in E} p_i(j)) = \sum_{i\in E}\nu_i \text{ (absurdo)},$$

con lo cual se cumple que $\nu = \nu \times P$. Para probar que ν es efectivamente una distribución, por la ecuación (1.8) se tiene que $\sum_{i\in E}\nu_i \leq 1$, y por lo demostrado recién se verifica:

$$\nu_j = \lim_n \sum_{i\in E}\nu_i p_i^n(j) \overset{(*)}{=} \sum_{i\in E}\lim_n \nu_i p_i^n(j) = \nu_j \sum_{i\in E}\nu_i.$$

Notemos que en $(*)$ se utilizó el teorema de convergencia dominada (recordar que $\sum_{i\in E}\nu_i \leq 1$). Obtuvimos así que $\nu_j \geq \nu_j \sum_{i\in E}\nu_i$ con $\nu_j > 0$, lo cual implica que $\sum_{i\in E}\nu_i = 1$.

Para probar la otra implicancia consideremos $\nu = (\nu_i)$ una distribución estacionaria. Como $\lim_n p_i^n(j)$ en general existe y no depende de i, basta verficar que dicho límite y ν_i coinciden. Para ello, como $\nu_j = \sum_{i\in E}\nu_i p_i^n(j)$ y $\sum_{i\in E}\nu_i = 1$ tenemos:

$$\nu_j = \lim_n \sum_{i\in E}\nu_i p_i^n(j) = \sum_{i\in E}\nu_i \lim_n p_i^n(j) = \lim_n p_i^n(j),$$

donde por la irreducibilidad de la cadena, $\lim_n p_i^n(j)$ debe ser nulo o estrictamente positivo para todos los estados j. En este caso no puede ser 0 ya que los ν_j suman 1, con lo que $\lim_n p_i^n(j) > 0$ y en consecuencia la cadena es recurrente positiva y por el teorema 1.6.7 vale $\nu_j = 1/\gamma_j$.

□

Tenemos así una base teórica que nos permite entender el comportamiento de las cadenas de Markov homogéneas. En la siguiente sección ajustaremos algunos conceptos y resultados al caso de las cadenas no homogéneas enfocándonos en un caso de interés: las cadenas de Markov con restricciones.

Capítulo 2

Cadenas de Markov con restricciones

2.1. Matrices y probabilidades de transición

En esta sección trabajaremos con cadenas de Markov en general (no necesariamente homogéneas en el tiempo), con énfasis en las cadenas de Markov con restricciones presentadas por Pachet, Roy y Barbieri en [18]. Para ello tendremos que ajustar algunas definiciones y notaciones. Por ejemplo, pierde sentido la noción de matriz de transición como única matriz asociada a la cadena, y para referirnos a las probabilidades de transición necesitaremos indicar más que los estados involucrados.

Sea $\{X_n\}_{n\in\mathbb{N}}$ es una cadena de Markov con espacio de estados E. Dados $i, j \in E$ notaremos $p_{ij}^{(k)} := P(X_k = j \mid P(X_{k-1} = i)$ con k entero positivo, a la probabilidad de ir de i a j en la k-ésima transición. Así para la k-ésima transición tenemos asociada una matriz de transición $P^{(k)} := (p_{ij}^{(k)})_{i,j}$ y denominamos μ a su distribución inicial.

Si bien estas matrices de transición son en cierto modo similares a la matriz de transición definida para el caso homogéneo, en este caso no podemos escribir las probabilidades de orden superior como potencia de una matriz de transición. Aún así es posible formular las siguientes propiedades básicas en función de las nuevas matrices y probabilidades de transición.

Proposición 2.1.1. *Sean $\{X_k\}_{k\in\mathbb{N}}$ una cadena de Markov con espacio de estados E, matrices de transición $P^{(k)} = (p_{ij}^{(k)})$ y distribución inicial μ, y n, m naturales con $n > m$ .Se cumple:*

1. *Si llamamos $P_{nm} = (p_{ij}(n,m))_{i,j\in E}$ a la matriz cuyas entradas son las probabilidades de transición de i a j en $n-m$ pasos partiendo de i en el instante m* (i.e, $p_{ij}(n,m) = P(X_n = j|X_m = i)$), *entonces se verifica que $P_{nm} = P^{(m+1)}P^{(m+2)}\dots P^{(n)}$.*
2. *Se cumple que $\mu^n = \mu \times P_{n0} = \mu P^{(1)}P^{(2)}\dots P^{(n)}, \forall n \in \mathbb{N}$, siendo $\mu^{(n)}$ la distribución en el instante n.*

3. *Las probabilidades de la forma* $P(X_n = i_n, X_{n-1} = i_{n-1}, \ldots, X_m = i_m)$ *pueden calcularse como* $p^{(n)}_{i_{n-1}i_n} p^{(n-1)}_{i_{n-2}i_{n-1}} \ldots p^{(m+1)}_{i_m i_{m+1}} \mu^{(m)}_{i_m}$.

La demostración es una adaptación directa de la prueba realizada para el caso homogéneo.

En este trabajo será de interés un tipo de cadenas no homogéneas que utilizaremos más adelante. Son las que permiten "adaptar" una cadena homogénea para que toda trayectoria finita pase, por ejemplo, por determinados estados fijos en ciertos instantes dados. Son las denominadas cadenas de Markov con restricciones, que presentaremos en lo que resta del capítulo.

2.2. Restricciones y consistencia por caminos

Para comenzar, es necesario definir que entenderemos por *restricciones*. Informalmente, supongamos que queremos obtener trayectorias finitas $x_0, x_1, \ldots, x_N$ a partir de una cadena de Markov, de forma tal que podamos imponer algunas condiciones para el comportamiento de la cadena en los distintos instantes $0, 1, \ldots, N$. Dichas condiciones podemos clasificarlas como:

- **Restricciones unitarias:** son condiciones que refieren a un único instante $k \in \{0, 1, \ldots N\}$ (siendo $N \in \mathbb{Z}^+$ el largo de la trayectoria a considerar) y consisten en indicar cuáles son los estados que puede tomar la variable X_k.

- **Restricciones binarias:** son condiciones que refieren a dos instantes consecutivos k y $k+1$ ($k \in \mathbb{N}$) e indican cuáles son los pares de estados posibles en X_k y X_{k+1}.

Veremos que es posible considerar este tipo de restricciones sin perder la condición de Markov. Sin embargo no será posible considerar restricciones que involucren a más de 2 estados consecutivos ya que la "pérdida de memoria" de las cadenas de Markov lo impide[1].

Definamos más rigurosamente las restricciones

Definición 2.2.1. Sea $\{X_k\}_{k\in\mathbb{N}}$ una cadena de Markov con espacio de estados E. Para todo $n \in \mathbb{N}$ definimos $U_n \subset E$ *restricciones (unitarias) sobre el instante* n como el conjunto de los estados en los que la variable X_n tiene probabilidad positiva, es decir:

$$U_n := \{i \in E : P(X_n = i) > 0\}.$$

[1]En general, para cadenas de orden N pueden considerarse restricciones que involucren hasta N estados consecutivos.

Asimismo para $n \geq 1$ definimos $B_n \subset E \times E$ conjunto de *restricciones (binarias) sobre la transición de $n-1$ a n* como los pares de estados en los que el vector (X_{n-1}, X_n) tiene probabilidad positiva, es decir:

$$B_n = \{(i,j) \in E \times E : P(X_{n-1} = i, X_n = j) > 0\}.$$

Nos interesará, dada de una cadena de Markov homogénea, modificarla imponiendo algunas restricciones adicionales, lo cual resultará en una cadena no homogénea, cuyas restricciones estarán contenidas en las restricciones de la original. Para ello será necesario tener algunas precauciones, como se ve en el siguiente ejemplo:

Ejemplo 2.2.2. Consideremos $\{X_k\}_{k\in\mathbb{N}}$ un paseo al azar simple en $\mathbb{Z}$, con estado inicial 0 fijo (es decir, $\mu_0 = 1$) y probabilidades de transición $p_i(i+1) = p$ y $p_i(i-1) = 1-p$ para todo $i \in \mathbb{Z}$. Supongamos que queremos generar trayectorias de largo 5 que comiencen y terminen en 0. La cadena original permite ello ya que, por ejemplo, la secuencia $0, 1, 2, 1, 0$ tiene probabilidad positiva. Además, antes de hacer modificaciones a la cadena, sus conjuntos de restricciones son

$$\begin{array}{ll} U_0 = \{0\} & \\ U_1 = \{1,-1\} & B_1 = \{(0,1),(0,-1)\} \\ U_2 = \{-2,0,2\} & B_2 = \{(-1,-2),(-1,0),(1,0),(1,2)\} \\ U_3 = \{-3,-1,1,3\} & B_3 = \{(-2,-3),(-2,-1),(0,-1),(0,1),(2,1),(2,3)\} \\ U_4 = \{-4,-2,0,2,4\} & B_4 = \{(-3,-4),(-3,-2),(-1,-2),(-1,0),(1,0),(1,2),(3,2),(3,4)\} \end{array}$$

Podría pensarse entonces que para imponer que el último estado sea 0 basta con modificar U_4 tomando $\hat{U}_4 = \{0\}$. Sin embargo la familia de restricciones que obtenemos no es consistente, ya que en ese caso podemos obtener la secuencia $(X_0 = 0, X_1 = 1, X_2 = 2, X_3 = 3)$, pero luego el par $(3,0) \notin B_4$ con lo cual no es posible que la trayectoria finalice en 0.

Así, si queremos que el paseo termine en 0, tendremos que imponer restricciones no sólo sobre U_4 sino sobre los demás estados y transiciones. Por ejemplo, $\hat{B}_4$ solo podrá contener aquellas transiciones que terminen en 0. Así, habrá que considerar $\hat{B}_4 = \{(-1,0),(1,0)\}$ y en consecuencia $\hat{U}_3 = \{-1,1\}$. Esto a su vez obliga a ajustar B_3, considerando en su lugar $\hat{B}_3 = \{(-2,-1),(0,-1),(0,1),(2,1)\}$. En resumen, los nuevos conjuntos de restricciones serán

$$\begin{array}{ll} \hat{U}_0 = U_0 = \{0\} & \\ \hat{U}_1 = U_1 = \{1,-1\} & \hat{B}_1 = B_1 = \{(0,1),(0,-1)\} \\ \hat{U}_2 = U_2 = \{-2,0,2\} & \hat{B}_2 = B_2 = \{(-1,-2),(-1,0),(1,0),(1,2)\} \\ \hat{U}_3 = \{-1,1\} & \hat{B}_3 = \{(-2,-1),(0,-1),(0,1),(2,1)\} \\ \hat{U}_4 = \{0\} & \hat{B}_4 = \{(-1,0),(1,0)\} \end{array}$$

Esto nos ayudará a determinar las nuevas matrices de transición, aunque aún nos falta terminar de delimitar el problema.

Observemos que al imponer restricciones es de esperarse que se reduzca el espacio de trayectorias posibles (i.e. que tienen probabilidad positiva). Así, en el ejemplo anterior la secuencia 0,1,2,3,2 tiene probabilidad positiva en el paseo al azar original, pero es de esperarse que tenga probabilidad 0 tras considerar las restricciones. Se definirán pues las nuevas probabilidades de transición de modo que las trayectorias prohibidas por las restricciones tengan probabilidad nula y las demás tendrán la probabilidad condicionada al nuevo conjunto de trayectorias posibles. Es decir, dada una trayectoria s definiremos su nueva probabilidad $\tilde{P}(s)$ del siguiente modo:

$$\tilde{P}(s) = \begin{cases} 0, & \text{si } s \notin S', \\ P(s|s \in S'), & \text{si } s \in S', \end{cases} \tag{2.1}$$

donde S' es el conjunto de las trayectorias que satisfacen las restricciones y P la probabilidad bajo la cadena homogénea. Más adelante determinaremos las probabilidades de transición bajo estas condiciones.

Vimos en el ejemplo que es necesario asegurar cierta consistencia en las restricciones. Para formalizar esto consideremos la siguiente definición:

Definición 2.2.3. Sean $\{X_k\}_{k\in\mathbb{N}}$ una cadena de Markov con espacio de estados E, $N \geq 2$ un número entero, $\{U_n\}_{n\in\{0,\dots,N\}}$ una familia de subconjuntos de E y $\{B_n\}_{n\in\{1,\dots,N\}}$ una familia de subconjuntos de $E\times E$. Diremos que $\{U_n\}$ y $\{B_n\}$ son *consistentes por caminos* como restricciones de la cadena si para todo $n \in \{0,\dots,N-1\}$ se verifica que

$$\forall i \in U_n \ \exists j \in U_{n+1} \ / \ (i,j) \in B_{n+1}.$$

Además, diremos que una secuencia de estados $i_0, i_1, \dots, i_l$ con $l \leq N$ es *consistente* si puede realizarse verificando todas las restricciones, es decir si cumple:

- $i_k \in U_k \ \forall k \in \{0,\dots,l\}$,
- $(i_{k-1}, i_k) \in B_k \ \forall k \in \{1,\dots,l\}$.

En palabras, que las restricciones sean consistentes por caminos asegura que toda trayectoria que comience verificando las restricciones podrá terminar haciéndolo (a diferencia de lo ocurrido inicialmente en el ejemplo anterior). Veamos que esto efectivamente es así:

Proposición 2.2.4. *Sean E un espacio de estados, $N \geq 2$ entero y $\{U_n\}_{n\in0,\dots,N}$ y $\{B_n\}_{n\in1,\dots N}$ familias de restricciones unitarias y binarias respectivamente. Si dichas restricciones son consistentes por caminos para toda trayectoria parcial consistente $i_0, i_1, \dots, i_l$ con $l < N$ (es decir, que puede realizarse verificando las restricciones para los instantes 0 a l) se cumple:*

1. *Existe $i_{l+1} \in U_{l+1}$ tal que $i_0, i_1, \dots, i_l, i_{l+i}$ es consistente.*
2. *La trayectoria $i_0, i_1, \dots, i_l, i_{l+1}$ es consistente si y sólo si i_l, i_{l+1} lo es.*

Demostración.

1. Como la trayectoria $i_0, i_1, ..., i_l$ es consistente sabemos que $i_l \in U_l$ y como las restricciones son consistentes por caminos, para $i_l \in U_l$ tenemos $i_{l+1} \in U_{l+1}$ tal que $(i_l, i_{l+1}) \in B_{l+1}$. Luego la trayectoria $i_0, \dots i_{l+1}$ resulta consistente.

2. Veamos que si i_l, i_{l+1} es consistente, la secuencia entera lo es (la otra implicancia es inmediata). Para ello basta notar que como la secuencia $i_0, i_1, \dots, i_l$ es consistente, y i_l, i_{l+1} también, se tiene que $i_k \in U_k \ \forall k \in \{0, 1, \dots l+1\}$ y $(i_{k-1}, i_k) \in B_k \ \forall k \in \{1, \dots l+1\}$ lo cual concluye la demostración.

□

Con esta proposición podremos generalizar el procedimiento visto en el ejemplo para determinar las nuevas restricciones de modo tal que sean consistentes por caminos, **siempre y cuando en la cadena original exista alguna trayectoria con probabilidad positiva que verifique las condiciones que queremos imponer**. Para ello pondremos el foco en la propagación de las restricciones unitarias, mientras que las restricciones binarias que de ello se desprendan quedarán de hecho impuestas en las matrices de transición.

- **Fijando un estado:** si se quiere imponer una condición de la forma $U_k = \{a\}$, hay que eliminar de U_{k+1} todos los estados a los que no se puede acceder desde a, es decir, los $b \in E \ / P(X_{k+1} = b | X_k = a) = 0$. De modo similar, hay que quitar de U_{k-1} los estados que no pueden ir hacia a, que son los $b \in E$ tales que $P(X_k = a | X_{k-1} = b) = 0$. Una vez removidos estos estados hay que seguir propagando las restricciones que generó la remoción de más estados, de acuerdo a lo siguiente.

- **Removiendo estados:** si se quiere quitar un estado a del conjunto U_k, habra que quitar de U_{k+1} todos los estados a los que sólo se puede acceder desde a, esto es, quitar los $b \in E$ tales que $P(U_{k+1} = b | U_k = c) = 0 \ \forall c \neq a$. De U_{k-1} quitaremos los estados que sólo podían ir hacia a, es decir los $b \in E$ tales que $P(X_k = c | X_{k-1} = b) = 0 \ \forall c \neq a$.

El proceso termina cuando las restricciones son consistentes por caminos, esto es, cuando en los pasos anteriores no hay más nada por hacer. Nótese que si hay al menos una trayectoria con probabilidad no nula en la cadena original que satisface las restricciones, el procedimiento anterior preserva los estados y trayectorias involucrados y por lo tanto el espacio de restricciones resultante no tiene conjuntos vacíos.

2.3. Condición principal y nuevas probabilidades

Dada una cadena de Markov homogénea, y un conjunto de restricciones a aplicar queremos determinar la cadena de Markov resultante de aplicar y propagar las restricciones. Anteriormente

vimos cómo modificar los conjuntos de restricciones para que sean consistentes. En esta sección determinaremos las matrices de transición y distribución inicial de la cadena resultante.

En este contexto, las matrices que consideraremos podrán no ser estocásticas: admitiremos también filas de ceros correspondientes a estados que no pertenecen al conjunto de restricciones unitarias en ese instante. Así, las filas de las matrices de transición deberan o bien sumar 1, o contener úncamente ceros. Llamaremos a estas matrices *cuasi estocásticas*.

Veamos entonces cómo se interpreta una familia de restricciones en términos de las matrices de transición y la distribución inicial. Para ello tenemos inicialmente una cadena homogénea $\{X_k\}_{k\ in\mathbb{N}}$ con distribución inicial μ y matriz de transición P, N un entero positivo y $\{U_n\}_{n\in\{0,\dots,N\}}$, $\{B_n\}_{n\in\{1,\dots,N\}}$ conjuntos de restricciones unitarias y binarias **consistentes por caminos** que queremos imponer.

Consideraremos una familia auxiliar $\{Z^{(n)}\}_{n\in\{0,\dots N\}}$ que se construirá del siguiente modo:

- **Inicialización.** Definimos $Z^{(0)} = \mu$ y $Z^{(n)} = P\ \forall n \in \{1,\dots,N\}$.
- **Remoción de estados.** Llamemos $z_{ij}^{(n)}$ a la entrada i,j de la matriz $Z^{(n)}$ Para cada $j \in E$ removido de U_n, se establece $z_{ij}^{(n)} = 0\ \forall i \in E$ (es decir, se lleva la j-ésima columna de la matriz a 0).
- **Remoción de transiciones.** Las transiciones prohibidas imponen ceros en las matrices del siguiente modo: $\forall i,j \in E,\ n \in \{1,\dots,N\}$ tales que $(i,j) \notin B_n$ se establece $z_{ij}^{(n)} = 0$.

Veamos cómo queda la familia $Z^{(n)}$ en el caso del paseo al azar con estado inicial y final 0.

Ejemplo. Recordemos que nuestro paseo tiene 4 transiciones (es decir $N = 4$) y las entradas de la matriz de transición P de la cadena homogénea son $p_i(j) = p\mathbb{1}_{\{j=i+1\}} + (1-p)\mathbb{1}_{\{j=i-1\}}$. Utilizando el procedimiento anterior con las restricciones $\{\hat{U}_n\}$ y $\{\hat{B}_n\}$ tenemos:

$$z_i^{(0)} = \begin{cases} 1, & \text{si } i = 0, \\ 0, & \text{en otro caso.} \end{cases}$$

Y las matrices $Z^{(n)} = (z_{ij}^{(n)})_{i,j\in E},\ n \geq 1$ donde $z_{ij}^{(n)}$ es:

- Para $n = 1$, las restricciones imponen $z_{ij}^{(n)} = 0$ para $j \notin \{-1, 1\}$. Así, la matriz resultante

tiene la forma

$$\begin{pmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \iddots \\ \dots & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ \dots & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ \dots & 0 & 0 & 1-p & 0 & p & 0 & \dots \\ \dots & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ \dots & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ \iddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

donde la fila azul y la columna amarilla corresponden al estado 0.

- Para $n = 2$ las restricciones generan la siguiente matriz:

$$\begin{pmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \iddots \\ \dots & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ \dots & 0 & 1-p & 0 & p & 0 & 0 & \dots \\ \dots & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ \dots & 0 & 0 & 0 & 1-p & 0 & p & \dots \\ \dots & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ \iddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

- Si $n = 3$ tenemos:

$$\begin{pmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \iddots \\ \dots & 0 & 0 & p & 0 & 0 & 0 & \dots \\ \dots & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ \dots & 0 & 0 & 1-p & 0 & p & 0 & \dots \\ \dots & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ \dots & 0 & 0 & 0 & 0 & 1-p & 0 & \dots \\ \iddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

- Y para $n = 4$:

$$\begin{pmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \iddots \\ \dots & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ \dots & 0 & 0 & 0 & p & 0 & 0 & \dots \\ \dots & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ \dots & 0 & 0 & 0 & 1-p & 0 & 0 & \dots \\ \dots & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ \iddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Cabe observar que las matrices $Z^{(1)}$ y $Z^{(2)}$ son cuasi estocásticas mientras que $Z^{(3)}$ y $Z^{(4)}$ no, con lo cual habrá que determinar un criterio para renormalizar las matrices obtenidas.

Recordemos que, como se mencionó anteriormente, si llamamos S al conjunto de trayectorias de largo N posibles para la cadena homogénea original al imponer restricciones sobre la misma queda determinado un subconjunto $S' \subset S$ que contiene únicamente a las trayectorias que verifican las restricciones. Para cada trayectoria $i = i_0 i_1 \dots i_N \in S'$ su nueva probabilidad será $\tilde{P}(i) = P(i|i \in S')$ (siendo $\tilde{P}(i) = 0$ si $i \notin S'$).

Observación 2.3.1. Este modo de definir las probabilidades no es equivalente a renormalizar las matrices $Z^{(n)}$ por filas. En otras palabras, si se verifica la condición (2.1) incluso las matrices $Z^{(n)}$ que sean estocásticas pueden verse modificadas.

Ejemplo. Retomemos el paseo al azar (con restricción sobre el último estado) de los ejemplos anteriores. Veamos que renormalizar las matrices individualmente no equivale a renormalizar considerando las trayectorias completas como en (2.1).

Al renormalizar individualmente las matrices lo que hacemos es crear $\tilde{Z}^{(n)}$ cuyas entradas son

$$\tilde{z}_{ij}^n = \frac{z_{ij}^{(n)}}{\sum_{j \in E} z_{ij}^{(n)}} \ \forall n \in \mathbb{N}, \ \forall i, j \in E.$$

Consideremos por ejemplo la trayectoria $i = 0, 1, 2, 1, 0$. Si consideramos las $\tilde{Z}^{(n)}$ como matrices de transición, y distribución inicial $Z^{(0)}$ tenemos que la probabilidad de la trayectoria es

$$\tilde{z}_0^{(0)} \tilde{z}_{01}^{(1)} \tilde{z}_{12}^{(2)} \tilde{z}_{21}^{(3)} \tilde{z}_{10}^{(4)} = (1)(p)(p)(1)(1) = p^2,$$

ya que $\tilde{z}_0^{(0)} = \mu_0 = 1$, las matrices $Z^{(1)}$ y $Z^{(2)}$ son cuasi estocásticas y las entradas involucradas en $Z^{(3)}$ y $Z^{(4)}$ son las únicas no nulas de sus filas. Ahora consideremos la cadena con matrices de transición $\{P^{(n)}\}$ y distribución inicial $\tilde{\mu}$ que verifique la condición (2.1) (aún no conocemos dichas matrices, pero asumamos de momento que existe tal cadena). Como la trayectoria i es consistente, su probabilidad es

$$\tilde{P}(i) = P(i|i \in S') = \frac{P(i)}{P(S')}, \tag{2.2}$$

siendo P la probabilidad en la cadena homogénea y S' el conjunto de las trayectorias posibles que terminan en 0. Ası, se tiene que $P(i) = \mu_0 p_0(1) p_1(2) p_2(1) p_1(0) = p^2(1-p)^2$ y además

$$P(S') = \sum_{s \in S'} P(s).$$

Es decir, $P(S')$ es la suma de las probabilidades de todas las trayectorias s que terminan en 0. Como el largo de las mismas es constante, se tiene que toda trayectoria $s \in S'$ deberá subir exactamente 2 veces y bajar otras 2, con lo cual $P(s) = p^2(1-p)^2 \ \forall s \in S'$ y en consecuencia para hallar $P(S')$ basta conocer la cantidad de trayectorias que contiene. Como para determinar

una trayectoria basta elegir cuáles 2 de las 4 transiciones serán las ascendentes, se tienen C_2^4 trayectorias en S'. Por lo tanto

$$P(S') = C_2^4 p^2 (1-p)^2.$$

Y sustituyendo en la ecuación (2.2):

$$\tilde{P}(i) = \frac{p^2(1-p)}{C_2^4 p^2 (1-p)^2} = \frac{1}{C_2^4}.$$

Que no coincide con la probabilidad hallada al normalizar cada matriz por separado. En particular, en este caso las probabilidades no dependen del p original, ya que como todas las trayectorias deben subir y bajar la misma cantidad de veces tiene sentido considerarlas equiprobables. Así, este último enfoque resulta más adecuado probabilísticamente, ya que aquellas trayectorias que tenían igual probabilidad en la cadena homogénea tendrán igual probabilidad bajo el modelo con restricciones.

La equiprobabilidad de las trayectorias nos permitirá hallar explícitamente las matrices de transición en este y todos los paseos al azar con último estado fijo como veremos a continuación.

2.4. Matrices de transición de las cadenas con restricciones

Nuestro objetivo es determinar las matrices de transición resultantes de aplicar restricciones a una cadena de Markov homogénea. Comencemos haciendo el cálculo para un caso de interés, aunque bastante particular

Ejemplo 2.4.1 (Paseo al azar con restricciones). Generalicemos un poco lo visto en los ejemplos anteriores considerando un paseo al azar con ciertos estados fijos. Considerando $\{X_n\}_{n\in\mathbb{N}}$ un paseo al azar simple en $\mathbb{Z}$ con probabilidad $p \in (0,1)$ de transitar del estado i al $i+1$ y estado inicial i_0 fijo. Consideramos trayectorias de N transiciones y queremos fijar restricciones unitarias de la forma $U_{n_k} = \{i_k\}$ con $k \in \{1,\ldots,K\}, K \leq N$ e $i_1 < i_2 < \ldots < i_K \in E$ estados.

Cabe notar que una vez que la trayectoria alcanza uno de los estados prefijados, por tratarse de una cadena de Markov la probabilidad de transición al siguiente estado no depende de los estados anteriores, con lo cual el problema puede reducirse al caso donde se restringe únicamente el último estado (y concatenando varias trayectorias de esas). Así, tenemos un estado inicial s_i y un estado final s_f fijos (con $|s_f - s_i| \leq N$) y queremos generar trayectorias de largo N entre ellos aplicando restricciones al paseo al azar.

Observemos que esto no siempre es posible: por ejemplo, el paseo al azar no permite trayectorias con 2 transiciones que comiencen en 1 y terminen en 0. En general, para que la trayectoria sea viable N deberá tener la misma paridad que $|s_f - s_i|$, ya que la cadena original tiene período 2.

Supongamos entonces que N y $|s_f - s_i|$ tienen la misma paridad. Hallemos las matrices de transición $\{P^{(n)}\}_{n\in 1,\dots,N}$ y la nueva distribución inicial.

Afirmación. Si $P^{(n)} = (p_{ij}^{(n)})_{i,j}$ se tiene que $\forall i, j \in \mathbb{Z},\ \forall n \in \{1, \dots, N\}$.

$$p_{ij}^{(n)} = \begin{cases} \frac{N-n-i+s_f+1}{2(N-n+1)}, & \text{si } j = i+1, \\ \frac{N-n+i-s_f+1}{2(N-n+1)}, & \text{si } j = i-1, \\ 0, & \text{en otro caso.} \end{cases} \tag{2.3}$$

Nótese que nuevamente las probabilidades no dependen del p elegido ni del estado inicial s_i. Lo único relevante es el estado final s_f, y la cantidad de pasos que faltan para terminar la trayectoria $(N - i - 1)$. Verifiquemos entonces dicha ecuación:

Demostración de la afirmación. Observemos primero que en el paseo al azar simple todas las trayectorias que van de s_i a s_f en N pasos son equiprobables: cada trayectoria deberá subir exactamente $k_1 := \frac{N-s_i+s_f}{2}$ veces y bajar las otras $k_2 := \frac{N+s_i-s_f}{2}$ con lo que cada una de ellas tiene probabilidad $p^{k_1}(1-p)^{k_2}$ en el paseo sin restricciones.

Al aplicar las restricciones, las nuevas probabilidades de las trayectorias son proporcionales a las anteriores con lo cual las trayectorias en el paseo al azar con restricciones también son equiprobables. Así, hallar probabilidades de transición será un problema de conteo donde la probabilidad de ir de i a $i+1$ en la n-ésima transición puede verse como

$$p_{ii+1}^{(n)} = \frac{\text{Cantidad de trayectorias que verifican } (X_{n-1} = i, X_n = i+1)}{\text{Cantidad de trayectorias que verifican } (X_{n-1} = i)}.$$

O, equivalentemente

$$p_{ii+1}^{(n)} = \frac{\#\{(s_{n+1}, \dots, s_{N-1}) \in E : P(X_n = i+1, X_{n+1} = s_{n+1}, \dots, X_{N-1} = s_{N-1}, X_N = s_f) > 0\}}{\#\{(s_n, \dots, s_{N-1}) \in E : P(X_{n-1} = i, X_n = s_n, X_{n+1} = s_{n+1}, \dots, X_{N-1} = s_{N-1}, X_N = s_f) > 0\}}. \tag{2.4}$$

Donde el numerador es la cantidad de trayectorias que van desde $i+1$ a s_f en $N-n$ pasos. Como vimos al comenzar la demostración, estas trayectorias suben exactamente $\frac{N-n-(i+1)+s_f}{2}$ veces con lo que el numerador resulta

$$C^{N-n}_{\frac{N-n-(i+1)+s_f}{2}} = \frac{(N-n)!}{\left(\frac{N-n-(i+1)+s_f}{2}\right)!\left(\frac{N-n+(i+1)-s_f}{2}\right)!}.$$

Razonando de modo similar tenemos que el denominador es

$$C^{N-n+1}_{\frac{N-n+1-i+s_f}{2}} = \frac{(N-n+1)!}{\left(\frac{N-n+1-i+s_f}{2}\right)!\left(\frac{N-n+1+i-s_f}{2}\right)!}.$$

Y sustituyendo en la igualdad (2.4) tenemos

$$p^{(n)}_{ii+1} = \frac{\cancel{(N-n)!}}{\cancel{(\frac{N-n-(i+1)+s_f}{2})!}\cancel{(\frac{N-n+(i+1)-s_f}{2})!}} \frac{(\frac{N-n+1-i+s_f}{2})!\cancel{(\frac{N-n+1+i-s_f}{2})!}}{(N-n+1)!} = \frac{N-n-i+s_f+1}{2(N-n+1)}$$

Luego como $p^{(n)}_{ij} = 0$ si $j \notin \{i-1, i+1\}$ se tiene que

$$p^{(n)}_{ii-1} = 1 - p^{(n)}_{ii+1} = \frac{N-n+i-s_f+1}{2(N-n+1)}$$

Lo cual prueba la afirmación

□

En particular, se aprecia que las probabilidades de transición son simétricas con respecto a $i = s_f$, ya que $p^{(n)}_{s_f+i}(s_f+i+1) = p^{(n)}_{s_f-i}(s_f-i-1)$.

Conocer las probabilidades de transición será de utilidad práctica para simular trayectorias aleatorias con restricciones bajo dichas probabilidades. Sin embargo no siempre será posible aplicar la estrategia utilizada en este caso para calcularlas ya que una condición que fue relevante para el cálculo fue la equiprobabilidad de las trayectorias, cosa que no necesariamente ocurre en un contexto más general.

Por otra parte, el ejemplo visto en esta sección presenta una limitación: si N y $|s_f - s_i|$ tienen distinta paridad, o si la distancia entre los estados es mayor a N, no es posible generar una trayectoria que cumpla las restricciones. Pero en la práctica queremos crear trayectorias que comuniquen s_i con s_f en N pasos sin que la paridad sea una limitación. Una opción "simple" para esto es modificar el paseo al azar, de modo que en cada transición también se pueda permanecer en un mismo estado. Sin embargo eso hace que no todas las trayectorias consistentes suban, bajen y permanezcan igual cantidad de veces, dificultando el cálculo de las probabilidades de transición.

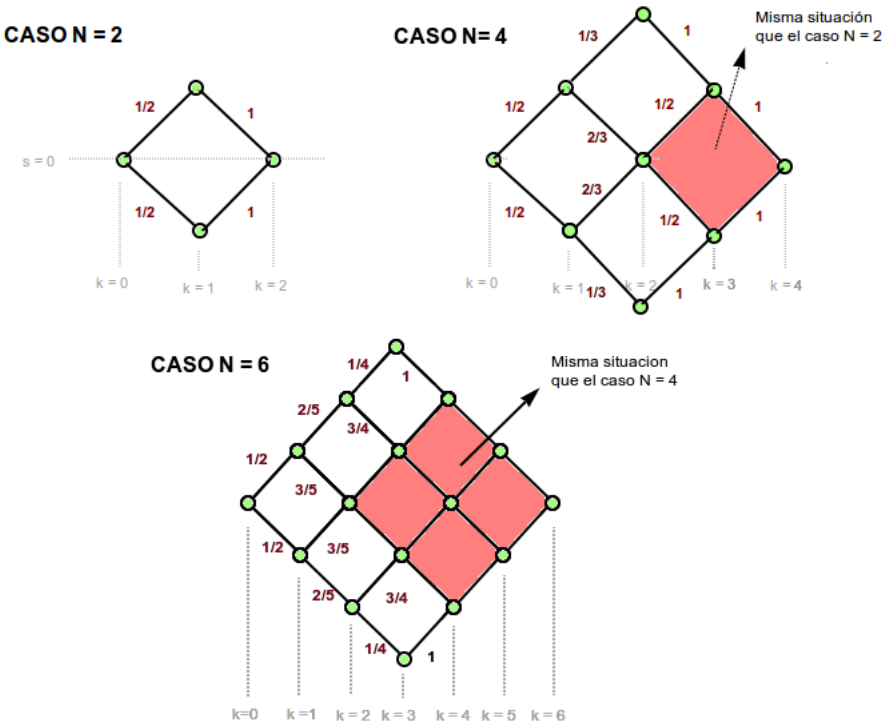


Figura 2.1: Diagramas de los paseos al azar de 0 a 0 en 2, 4 y 6 pasos con sus probabilidades de transición

Veamos pues una expresión genérica para las probabilidades de transición de una cadena de Markov con restricciones

Teorema 2.4.2. *Consideremos $\{X_k\}_{k\in\mathbb{N}}$ una cadena de Markov homogénea en E finito con distribución inicial μ y matriz de transición P, $N \geq 2$ un entero, $\{U_n\}_{n\in\{0,\dots,N\}}$,$\{B_n\}_{n\in\{1,\dots,N\}}$ familias de restricciones –unitarias y binarias respectivamente– consistentes por caminos, y $\{Z^{(n)}\}_{n\in\{1,\dots,N\}}$ matrices construidas según el procedimiento descrito en la sección 2.3. Definiremos las matrices de transición $\{\tilde{P}^{(n)}\}_{n\in\{1,\dots,N\}}$ con $\tilde{P}^{(n)} = (\tilde{p}_{ij}^{(n)})_{i,j}$ de modo que*

$$\tilde{p}_{ij}^{(N)} = \frac{z_{ij}^{(N)}}{\alpha_i^{(N)}}, \qquad \text{siendo } \alpha_i^{(N)} = \sum_{k\in E} z_{ik}^{(N)},$$
$$\tilde{p}_{ij}^{(n)} = \frac{\alpha_j^{(n+1)} z_{ij}^{(n)}}{\alpha_i^{(n)}}, \quad \text{siendo } \alpha_i^{(n)} = \sum_{k\in E} \alpha_k^{(n+1)} z_{ik}^{(n)} \ \forall n \in \{1,\dots,N-1\}.$$

Y la distribución inicial $\tilde{\mu}$ de modo que

$$\tilde{\mu}_i = \frac{\alpha_i^{(1)} z_i^{(0)}}{\alpha^{(0)}}, \quad \text{con } \alpha^{(0)} = \sum_{k\in E} \alpha_k^{(1)} z_k^{(0)}.$$

En caso de que $\alpha_i^{(n)} = 0$ *las expresiones de los* $p_{ij}^{(n)}$ *son de la forma* $\frac{0}{0}$ *e impondremos* $p_{ij}^{(n)} = 0$.

Entonces la cadena (no homogénea) de trayectorias finitas, con matrices de transición $\{\tilde{P}^{(n)}\}_{n\in\{1,\dots,N\}}$ *y distribución inicial* $\tilde{\mu}$ *definidas como antes verifica la condición* (2.1).

En palabras, el procedimiento para hallar las matrices de transición consiste en normalizar individualmente la matriz correspondiente a la última transición, para luego propagar la normalización hacia atrás de modo que cumpla la propiedad buscada, como se verá en la demostración.

Demostración. Verifiquemos que las matrices $\{\tilde{P}^{(n)}\}$ son cuasi estocásticas, y que las entradas de μ efectivamente suman 1.

Si $n = N$ tenemos:

$$\sum_{j\in E} p_{ij}^{(N)} = \sum_{j\in E} \frac{z_{ij}^{(N)}}{\alpha_i^{(N)}} = \frac{1}{\alpha_i^{(N)}} \sum_{j\in E} z_{ij}^{(N)} = 1,$$

cuando $\alpha_i^{(N)} \neq 0$. En caso contrario $p_{ij}^{(N)} = 0\ \forall j \in E$ y tenemos una fila de ceros. De modo similar para $n \in \{1, \dots, N-1\}$, si $\alpha_i^{(n)} \neq 0$:

$$\sum_{j\in E} p_{ij}^{(n)} = \sum_{j\in E} \frac{\alpha_j^{(n+1)} z_{ij}^{(n)}}{\alpha_i^{(n)}} \overset{\text{def } \alpha}{=} \sum_{j\in E} \frac{\alpha_j^{(n+1)} z_{ij}^{(n)}}{\sum_{k\in E} \alpha_k^{n+1} z_{ik}^{(n)}} = 1,$$

mientras que si $\alpha_i^{(n)} = 0$ tenemos nuevamente una fila de ceros. De modo similar se verifica que las entradas de $\tilde{\mu}$ suman 1.

Veamos ahora que se cumple la condición (2.1), para lo cual consideramos una trayectoria $s = s_0, s_1, \dots, s_N$ que verifica las restriciones (en caso contrario es inmediato que $\tilde{P}(s) = 0$). Tenemos entonces:

$$\begin{aligned}
\tilde{P}(s) &= \tilde{\mu}_{s_0} p_{s_0 s_1}^{(1)} p_{s_1 s_2}^{(2)} \dots p_{s_{N-1} s_N}^{(N)} \\
&= \frac{\cancel{\alpha_{s_0}^{(1)}} z_{s_0}^{(0)}}{\alpha^{(0)}} \frac{\cancel{\alpha_{s_1}^{(2)}} z_{s_0 s_1}^{(1)}}{\cancel{\alpha_{s_0}^{(1)}}} \dots \frac{\cancel{\alpha_{s_{N-1}}^{(N)}} z_{s_{N-2} s_{N-1}}^{(N-1)}}{\cancel{\alpha_{s_{N-2}}^{(N-1)}}} \frac{z_{s_{N-1} s_N}^{(N)}}{\cancel{\alpha_{s_{N-1}}^{(N)}}} \\
&= \frac{1}{\alpha^{(0)}} z_{s_0}^{(0)} z_{s_0 s_1}^{(1)} \dots z_{s_{N-2} s_{N-1}}^{(N-1)} z_{s_{N-1} s_N}^{(N)},
\end{aligned}$$

donde el producto resultante $z_{s_0}^{(0)} z_{s_0 s_1}^{(1)} \dots z_{s_{N-2} s_{N-1}}^{(N-1)} z_{s_{N-1} s_N}^{(N)}$ es no nulo ya que s verifica las restricciones. Más aún, por construcción, se tiene que si $z_{ij}^{(n)} \neq 0$, $z_{ij}^{(n)} = p_i(j)$ y $z_i^{(0)} = \mu_i$ con lo

cual

$$\begin{aligned}\tilde{P}(s) &= \frac{1}{\alpha^{(0)}} z^{(0)}_{s_0} z^{(1)}_{s_0 s_1} \dots z^{(N)}_{s_{N-2} s_{N-1}} z^{(N)}_{s_{N-1} s_N} \\ &= \frac{1}{\alpha^{(0)}} \mu_{s_0} p_{s_0}(s_1) \dots p_{s_{N-1}}(s_N) \\ &= \frac{1}{\alpha^{(0)}} P(s),\end{aligned}$$

siendo P la probabilidad bajo la cadena homogénea. Es decir, toda trayectoria s que cumpla las restricciones verifica $\tilde{P}(s) = cP(s)$ donde $c = \frac{1}{\alpha^{(0)}}$ es una constante que no depende de s. Luego si S' es el conjunto de trayectorias que verifican las restricciones se tiene que $\tilde{P}(S') = 1$. Por lo tanto $P(S') = \alpha^{(0)}$ y

$$\tilde{P}(s) = \frac{P(s)}{P(S')} = P(s|s \in S') \ \forall s \in S'.$$

Lo cual verifica la condición (2.1) y concluye la demostración.

□

Tenemos así un algoritmo que nos permite, dada una cadena de Markov homogénea y ciertas restricciones, determinar todos los parámetros de la cadena inducida por las mismas. En el siguiente capítulo trataremos la implementación del mismo, así como la simulación de cadenas de Markov en general.

Capítulo 3

Estadística y simulación de cadenas de Markov

Este capítulo abordará dos aspectos de interés para las posteriores aplicaciones. Estos son:

- Estimar parámetros en cadenas de Markov homogéneas: dado un conjunto de trayectorias, determinar la cadena "más adecuada" para las mismas (por ejemplo, la de máxima verosimilitud).
- Dada una cadena de Markov (homogénea o no) simular trayectorias que respeten sus probabilidades.

3.1. Estadística en cadenas de Markov homogéneas

Para el problema de estimación recordaremos algunas definiciones y resultados de interés, y plantearemos estimadores para los parámetros de cadenas de Markov homogéneas.

3.1.1. Nociones previas de estimación

Comencemos recordando algunos conceptos básicos

Definición 3.1.1. Sea X una variable aleatoria con función de distribución F_X. Diremos que $X_1, \ldots, X_n$ es una *muestra aleatoria simple* (M.A.S) de X si las variables son independientes e idénticamente distribuídas con distribución F_X.

Definición 3.1.2. Sean $X_1, \ldots, X_n$ M.A.S de $X \sim F_X(x|\theta)$ (léase, la función de distribución depende de un parámetro $\theta \in \mathbb{R}^k$) y $\hat{\theta} : \mathbb{R}^n \to \mathbb{R}^k$ función medible que no depende de θ. Llamaremos *estimador* de θ a la variable/vector aleatorio $\theta_n := \hat{\theta}(X_1, \ldots, X_n)$.

Si se tiene que $\hat{\theta_n} \xrightarrow[n\to+\infty]{c.s} \theta$ diremos que el estimador es *fuertemente consistente*, mientras que si $\hat{\theta_n} \xrightarrow[n\to+\infty]{P} \theta$ el estimador es *débilmente consistente*.

En la práctica, dado un parámetro nos interesará encontrar estimadores (si es posible fuertemente) consistentes para el mismo. Veremos algunos resultados que nos serán de utilidad para ello.

Teorema 3.1.3 (Ley fuerte de los grandes números)**.** *Sea $\{X_k\}_{k\in\mathbb{N}}$ una sucesión de variables aleatorias independientes e igualmente distribuidas tales que $E(X_1) = \mu < \infty$. Se cumple:*

$$\frac{1}{n}\sum_{k=1}^{k=n} X_k \xrightarrow{c.s} \mu, \qquad \text{cuando } n \to \infty.$$

Además notaremos $\overline{X}_n := \frac{1}{n}\sum_{k=1}^{k=n} X_k$.

Es decir, el promedio de una M.A.S de X es un estimador fuertemente consistente de EX, si dicha esperanza es finita. Veamos cómo aplicando el teorema a las variables $\{X^k\}$ podremos estimar parámetros que dependan de los momentos poblacionales de X.

Ejemplo 3.1.4 (Método de los momentos). Sea $X_1, \ldots, X_n$ M.A.S de $X \sim F_X(x|\theta)$ con $\theta = (\theta_1, \ldots, \theta_k) \in \mathbb{R}^k$ y X tal que $E|X|^k < \infty$. Consideremos el siguiente sistema de ecuaciones

$$\begin{cases} EX &= \overline{X_n} \\ EX^2 &= \overline{X_n^2} \\ \vdots & \vdots \quad \vdots \\ EX^k &= \overline{X_n^k} \end{cases}$$

Donde los EX^n se expresan en función de las coordenadas de θ resultando así un sistema de k ecuaciones y otras tantas incógnitas. Llamemos $f : \Theta \to f(\Theta) \subset \mathbb{R}^k$ a la función tal que $(E(X), E(X^2), \ldots, E(X^k)) = f(\theta_1, \ldots, \theta_k)$. Si f es invertible (i.e. inyectiva), f^{-1} es contínua y $(E(X), E(X^2), \ldots, E(X^k)) \in f(\Theta)$ se tiene que el sistema tiene una única solución $\hat{\theta} = (\hat{\theta_1}, \ldots, \hat{\theta_k})$ que es un estimador fuertemente consistente de θ. Esto se debe a la continuidad de f^{-1} y a que, por la ley fuerte de los grandes números, $\overline{X^k}_n \to EX^k$. Una descripción más detallada del método puede encontrarse en [2]

Otro método de estimación consiste en hallar los máximos de la función de verosimilitud

Ejemplo 3.1.5 (Método de máxima verosimilitud). Sea $X_1 \dots X_n$ una M.A.S de una variable X discreta (o absolutamente continua) con función de probabilidad (densidad) p, $\theta \in \mathbb{R}^k$ un parámetro y $\Theta \subset \mathbb{R}^k$ el conjunto de los posibles θ. Llamemos $p(x|\theta)$ la función de probabilidad (densidad) asumiendo que θ es el parámetro. Definiremos la *función de verosimilitud* $L : \Theta \times \mathbb{R}^n \to \mathbb{R}$ como:

$$L(\theta, (x_1, \dots x_n)) = \prod_{i=1}^{i=n} p(x_i|\theta).$$

Esta función nos dice qué tan probable es la secuencia $(x_1, \dots, x_n)$ cuando el parámetro es θ. En particular, si alguno de los argumentos x_k tiene probabilidad nula bajo θ la función de verosimilitud vale 0.

El método de máxima verosimilitud propone como estimador de θ al $\hat{\theta} \in \Theta$ que maximiza la función de verosimilitud para la M.A.S dada. Es decir

$$\hat{\theta} := \hat{\theta}(X_1, \dots, X_n) = \arg\max_{\theta \in \Theta} L(\theta, (X_1, \dots, X_n)).$$

Si bien en este caso la demostración no es inmediata, se puede probar bajo ciertas condiciones el estimador de máxima verosimilitud es fuertemente consistente: si tenemos una M.A.S $X_1, \dots, X_n$ con probabilidad/densidad p_{θ_0}, el espacio Θ es compacto, la función $g_x(\theta) := p(x|\theta)$ es continua para todo $x \in \mathbb{R}$, el mapa $\theta \longmapsto p_\theta$ es inyectivo y existe una función $K : \mathbb{R} \to \mathbb{R}$ tal que $E_{\theta_0}|K(X)| < \infty$ (siendo E_{θ_0} la esperanza condicionada a que $\theta = \theta_0$) y $log_{p_\theta}(x) - log_{\theta_0}(x) \leq K(x)$ $\forall x \in \mathbb{R}^+, \theta \in \Theta$, se tiene la consistencia del estimador máximo verosímil (que converge casi seguramente a θ_0). Una prueba de este resultado así como un estudio un poco más detallado de los estimadores por máxima verosimilitud puede verse en [10].

Utilizando cualquiera de estos métodos podemos estimar, por ejemplo, las probabilidades de aparición de cierto valor en una secuencia de observaciones independientes. Para ello consideremos una variable discreta X que toma valores en un conjunto E y $X_1, \dots, X_n$ una M.A.S de X. Dado $e \in E$ queremos estimar $p := P(X = e)$.

Definiendo variables $Y_k := \mathbb{1}_{\{X_k = e\}}$ $\forall k \in \{1, \dots, n\}$ tenemos que $\{Y_k\}_{k \in \{1,\dots,n\}}$ es una M.A.S de una variable $Y \sim Ber(p)$. Luego por la ley fuerte de los grandes números $EY = p$ y por lo tanto $\overline{Y}_n$ es el estimador por momentos de p (además puede probarse fácilmente que, en este caso, el estimador por máxima verosimilitud de p también es $\overline{Y}_n$). Como $\sum Y_k$ es la cantidad de veces que $X_k = e$, tenemos

$$\hat{p} = \frac{\#\{k : X_k = e\}}{n}. \tag{3.1}$$

Este ejemplo particular nos será de utilidad para estimar las entradas de una matriz de

transición. Notemos sin embargo que una secuencias $X_1, \ldots, X_n$ donde $\{X_k\}$ es una cadena de Markov no es una M.A.S ya que las variables no son independientes y por lo tanto no podremos utilizar los métodos de estimación vistos directamente sobre las variables X_k. Veamos cómo estimar entonces los parámetros buscados.

3.1.2. Estimación en cadenas de Markov

Supongamos que observamos un conjunto finito $S = \{s^n\}_{n\in\{1,\ldots,N\}}$ de trayectorias finitas e independientes entre sí de una cadena de Markov homogénea con espacio de estados E finito, cuya matriz de transición P y distribución inicial μ desconocemos y queremos estimar.

Las entradas de μ son las probabilidades puntuales de X_0. Si escribimos las trayectorias en S como $s^n = s_0^n s_1^n \ldots s_{k_n}^n$, podemos construir una M.A.S de X_0 considerando la primera observación de cada trayectoria, es decir, el conjunto $\{s_0^n : n \in \{1, \ldots, N\}\}$. Luego para cada $i \in E$ podemos aplicar la ecuación (3.1) obteniendo:

$$\hat{\mu}_i = \frac{1}{N} \sum_{n=1}^{n=N} \mathbb{1}_{\{s_0^n = i\}},$$

siendo N la cantidad de elementos en S.

Por lo visto anteriormente, sabemos que los $\hat{\mu}_i$ son estimadores consistentes de μ_i y en consecuencia $\hat{\mu} := (\hat{\mu}_i)_{i\in E}$ es un estimador consistente de μ. Veamos que además tiene sentido considerar $\hat{\mu}$ como distribución inicial: por construcción sabemos que $\hat{\mu}_i \in [0, 1]$ $\forall i \in E$, y sumando las entradas se tiene:

$$\sum_{i\in E} \hat{\mu}_i = \frac{1}{N} \sum_{i\in E} \sum_{s^n \in S} \mathbb{1}_{\{s_0^n = i\}} = \frac{1}{N} \sum_{i\in E} \#\{n : s_0^n = i\} = 1.$$

Encontremos ahora estimadores para las entradas $p_i(j)$ de la matriz de transición. Para ello fijemos un estado i y para cada trayectoria se crea una secuencia auxiliar que sólo contiene a aquellas observaciones precedidas de i. Es decir, para cada $s^n = s_0^n, s_1^n, \ldots, s_{k_n}^n$ se considera $s'^n = \{s_k^n\}_{\{k\in\{1,\ldots,k_n\} \,:\, s_{k-1}=i\}}$.

Notemos que, por tratarse de una cadena de Markov, estas nuevas secuencias son muestras independientes de una variable Y con valores en E y tal que $P(Y = j) = p_i(j)$ $\forall j \in E$. Estimando nuevamente de acuerdo a la ecuación (3.1) resulta:

$$\hat{p}_i(j) = \begin{cases} \frac{1}{N_i} \displaystyle\sum_{n=1}^{n=N} \sum_{s_k \in s'^n} \mathbb{1}_{\{s_k = j\}}, & \text{si } N_i \neq 0, \\ 0, & \text{si } N_i = 0, \end{cases}$$

con $N_i = \sum_{n\in\{1,\dots,N\}} \#\{s'^n\}$ la cantidad total de observaciones precedidas por i. En otras palabras, para estimar $p_i(j)$ se consideran todas las transiciones en S que comienzan en i, y se calcula la proporción de éstas que resuelve en j. Una vez más tenemos que los $\hat{p}_i(j)$ así definidos son estimadores fuertemente consistentes de $p_i(j)$ y tiene sentido definir $\hat{P} = (\hat{p}_i(j))_{i,j}$. Para verificar que la matriz es cuasi estocástica consideremos i tal que $N_i \neq 0$ y sumando las entradas de la i-ésima fila tenemos:

$$\sum_{j\in E} \hat{p}_i(j) = \frac{1}{N_i} \sum_{j\in E} \sum_{n\in\{1,\dots,N\}} \sum_{s_k\in s'^n} \mathbb{1}_{\{s_k=j\}} = \frac{1}{N_i} \sum_{j\in E} \#\{\text{transiciones de } i \text{ a } j\} = 1.$$

Veamos el procedimiento aplicado a un pequeño ejemplo.

Ejemplo 3.1.6. Supongamos que queremos estimar los parámetros de una cadena de Markov homogénea con espacio de estados $E = \{1, 2, 3, 4\}$. Se observan las siguientes trayectorias:

$$\begin{array}{ccc} s^1 = 3,1,1,2,4,1,2 & s^3 = 2,4,4,3,1 & s^5 = 4,3,4,4,1,2,1,3 \\ s^2 = 4,1,3,3,2,1,4,3,2 & s^4 = 4,2,3,3,2,2,1 & s^6 = 3,2,4,1,2,1. \end{array}$$

Obsérvese que las secuencias no tienen por qué ser de igual largo.

Para estimar μ utilizamos la primer observación de cada secuencia, es decir tenemos como muestra $3, 4, 2, 4, 4, 3$ y por lo tanto $\hat{\mu} = (0, \frac{1}{6}, \frac{1}{3}, \frac{1}{2})$. Ahora fijemos $i = 1$. Si consideramos sólo aquellas observaciones precedidas por el estado 1 tenemos:

$$\begin{array}{ccccc} s'^1 = s^1_2, s^1_3, s^1_6 = 1,2,2 & & s'^3 = \emptyset & & s'^5 = s^5_5, s^5_7 = 2,3 \\ s'^2 = s^2_2, s^2_6 = 3,4 & & s'^4 = \emptyset & & s'^6 = s^6_4 = 2. \end{array}$$

Luego contando la cantidad de observaciones tenemos $N_1 = 8$ y si consideramos $j = 1$, $\hat{p}_i(j) = \frac{1}{8}$ ya que el estado 1 sólo se observa una vez en los s'. Procediendo de igual modo con los demás $j \in E$ resulta:

$$\hat{p}_1(1) = \tfrac{1}{8} \quad \hat{p}_1(2) = \tfrac{1}{2} \quad \hat{p}_1(3) = \tfrac{1}{4} \quad \hat{p}_1(4) = \tfrac{1}{8}.$$

Luego se repite el procedimiento para $i = 2, 3, 4$, obteniendo la siguiente matriz:

$$\hat{P} = \begin{pmatrix} 1/8 & 1/2 & 1/4 & 1/8 \\ 4/9 & 1/9 & 1/9 & 3/9 \\ 2/9 & 4/9 & 2/9 & 1/9 \\ 2/5 & 1/10 & 3/10 & 1/5 \end{pmatrix}.$$

Con lo cual se tienen estimadas la distribución inicial y matriz de transición de la cadena.

Cabe notar que, como en toda estimación, es deseable disponer de una cantidad grande de observaciones. Dado que estamos trabajando con cadenas homogéneas, esto se puede lograr o bien incrementando el número de secuencias o bien incrementando la longitud de las mismas. En el caso de las cadenas no homogéneas, el número de parámetros a estimar crece en tanto se tiene una matriz por cada transición, lo cual hace que se requiera observar más secuencias para obtener una estimación razonable. Así, en nuestras aplicaciones estimaremos cadenas homogéneas sobre las cuales posteriormente se aplicarán restricciones (obteniendo cadenas no homogéneas pero sin estimarlas de forma directa).

3.2. Simulación de cadenas de Markov

3.2.1. Simulación de cadenas homogéneas

Una vez resuelto el problema de estimar los parámetros de una cadena de Markov a partir de ciertas trayectorias, nos interesará generar nuevas trayectorias bajo la cadena estimada. Para ello veremos cómo simular una cadena de Markov dada, con espacio de estados E finito, matriz de transición P y distribución inicial μ.

Asumiremos -pese a que no es estrictamente cierto- que disponemos de un generador de números aleatorios e independientes, con distribución uniforme en $[0, 1]$, de modo que el único componente aleatorio que podrá utilizar nuestro algoritmo son variables i.i.d uniformes. Comencemos definiendo dos funciones que nos ayudarán en la construcción:

- La *función de iniciación* nos permitirá determinar el primer estado de nuestras trayectorias. Depende de la distribución inicial de la cadena.
- La *función de actualización* nos permitirá, dado un estado, determinar el estado siguiente. Depende de la matriz de transición.

Queremos construir una función de iniciación $\psi : [0, 1] \to E$, de modo que si U_0 es una variable uniforme en $[0, 1]$ se cumple:

$$P(\psi(U_0) = i) = \mu_i, \tag{3.2}$$

ya que de ese modo tiene sentido considerar $X_0 = \psi(U_0)$. Esto motiva la siguiente definición

Definición 3.2.1. Sea $\{X_k\}$ una cadena de Markov con espacio de estados E finito, matriz de transición P y distribución inicial μ. Llamaremos *función de iniciación* de la cadena a una función $\psi : [0, 1] \to E$ que verifica:

1. ψ es constante a trozos.

2. Para todo estado i se cumple que la medida total de la región en la que $\psi(x) = i$ es μ_i. Es decir que

$$\int_0^1 \mathbb{1}_{\{\psi(x)=i\}} dx = \mu_i \ , \ \forall i \in E.$$

Observación. ψ así definida verifica la ecuación (3.2)

Para probarlo basta calcular $P(\psi(U_0) = i)$ bajo estas condiciones. Considerando $U_0 \sim \mathcal{U}[0,1]$ e $i \in E$ se tiene:

$$P(\psi(U_0) = i) = P(U_0 \in \psi^{-1}(i)) = \int_0^1 \mathbb{1}_{\{x\in\psi^{-1}(i)\}} dx =$$
$$= \int_0^1 \mathbb{1}_{\{\psi(x)=i\}} dx \overset{2}{=} \mu_i,$$

donde la condición 1 nos asegura que ψ es integrable en $\psi^{-1}(i)$.

Habiendo determinado las condiciones que queremos que cumpla la función de iniciación, veamos una implementación concreta de la misma.

Ejemplo 3.2.2. Consideremos una cadena de Markov con espacio de estados $E = \{1, \dots, k\}$ finito, distribución inicial μ y matriz de transición P. Para todo $j \in \{1, \dots, k\}$ definamos $A_j = [\sum_{i=1}^{i=j-1} \mu_i, \sum_{i=1}^{i=j} \mu_i]$ intervalos. La función $\psi : [0,1] \to E$ tal que

$$\psi(x) = \sum_{j\in E} j \mathbb{1}_{\{x\in A_j\}} =$$
$$= \begin{cases} 1, & \text{si } x \in [0, \mu_1], \\ 2, & \text{si } x \in [\mu_1, \mu_1 + \mu_2], \\ \vdots & \vdots \\ k, & \text{si } x \in [\sum_{i=1}^{i=k-1} \mu_i, 1], \end{cases}$$

es una función de iniciación para la cadena.

Nota: recordar que los estados no tienen por qué ser necesariamente de la forma $\{1, \dots, k\}$, pero por ser E biyectivo a $\{1, \dots, k\}$ no se pierde generalidad. Mantenemos así la convención del capítulo anterior de referir a los estados por su numeración.

Para verificar la afirmación notemos que de la definición de ψ se desprende de inmediato que es constante a trozos, con lo cual resta verificar la segunda condición. Dado $i \in E$, tenemos que

$\{x : \psi(x) = i\} = A_i$ es un intervalo y por lo tanto:

$$\int_0^1 \mathbb{1}_{\{\psi(x)=i\}} dx = \int_0^1 \mathbb{1}_{A_i} dx = \sum_{j=1}^{j=i} \mu_j - \sum_{j=1}^{j=i-1} \mu_j = \mu_i.$$

Una vez definida la función de iniciación, estamos en condiciones de simular X_0, la primera variable de nuestras trayectorias. Para ello basta tomar $X_0 = \psi(U_0)$ con $U_0 \sim \mathcal{U}[0,1]$. La función de actualización nos permitirá generar el resto de la trayectoria.

Definición 3.2.3. Sea $\{X_k\}$ una cadena de Markov con espacio de estados E finito, matriz de transición P y distribución inicial μ. Llamaremos *función de actualización* de la cadena a una función $\Phi : E \times [0,1] \to E$ que verifica:

1. Para todo $i \in E$ el mapa $x \longmapsto \Phi(i,x)$ es constante a trozos.
2. Dados $i, j \in E$, el conjunto $\{x : \Phi(i,x) = j\}$ tiene medida $p_i(j)$, es decir:

$$\int_0^1 \mathbb{1}_{\{\Phi(i,x)=j\}} dx = p_i(j) \ , \ \forall i, j \in E.$$

Notemos que, si se sabe que $X_n = i$, usando la función de actualización podemos establecer $X_{n+1} := \Phi(X_n, U)$ siendo U una variable aleatoria uniforme en $[0,1]$ e independiente de las variables X anteriores. En efecto, dado $j \in E$ se tiene:

$$P(X_{n+1} = j | X_n = i) = P(\Phi(i,U) = j | X_n = i) = \int_0^1 \mathbb{1}_{\{x:(i,x)\in\Phi^{-1}(j)\}} dx = \int_0^1 \mathbb{1}_{\{\Phi(i,x)=j\}} dx = p_i(j),$$

con lo cual X_{n+1} tiene el comportamiento esperado. Notemos que, al igual que ocurre con la función de iniciación, no es difícil construir explícitamente una función de actualización.

Ejemplo 3.2.4. Consideremos una cadena de Markov $\{X_n\}$ como en la definición 3.2.3 con $E = \{1, \dots k\}$. Para $i, j \in E$ definimos $B_{ij} = [\sum_{h=1}^{h=j-1} p_i(h), \sum_{h=1}^{h=j} p_i(h)]$. Entonces la función $\Phi : E \times [0,1] \to E$ tal que para todo $i \in E$ se verifica:

$$\begin{aligned}\Phi(i,x) &= \sum_{j\in E} j \mathbb{1}_{\{x\in B_{ij}\}} \\ &= \begin{cases} 1, & \text{si } x \in [0, p_i(1)], \\ 2, & \text{si } x \in [p_i(1), p_i(1) + p_i(2)], \\ \vdots & \vdots \\ k, & \text{si } x \in [\sum_{h=1}^{h=k-1} p_i(h), 1], \end{cases}\end{aligned}$$

es una función de actualización para la cadena.

Una vez más es inmediato por construcción que $x \longmapsto \Phi(i,x)$ es constante a trozos. Para la segunda condición se considera $i,j \in E$ y entonces:

$$\int_0^1 \mathbb{1}_{\{\Phi(i,x)=j\}} dx = \int_0^1 \mathbb{1}_{B_{ij}} dx = \sum_{h=1}^{h=j} p_i(h) - \sum_{h=1}^{h=j-1} p_i(h) = p_i(j).$$

Con estas funciones podemos construir trayectorias de largo arbitrario iterando la función de actualización. Resumiendo lo visto anteriormente, si dada una cadena de Markov como las anteriores y $N \geq 1$ se quieren simular secuencias $X_0, \ldots, X_N$, se generan $U_1, \ldots, U_N$ variables i.i.d con distribución $\mathcal{U} \sim [0,1]$ y se define:

- $X_0 = \psi(U_0)$.
- $X_n = \Phi(X_{n-1}, U_n)$ para $n \in \{1, \ldots N\}$.

3.2.2. Adaptación al caso no homogéneo

Supongamos ahora que queremos simular trayectorias de una cadena de Markov no homogénea. Para ello nuevamente consideraremos funciones de iniciación y de actualización, sólo que esta vez la función de actualización depende -además- del tiempo.

Generalicemos entonces la función de actualización.

Definición 3.2.5. Se considera una cadena de Markov $\{X_k\}_{k\in\mathbb{N}}$ no homogénea, con matrices de transición $\{P^{(n)}\}_{n\geq 1}$, espacio de estados E finito y distribución inicial μ. Diremos que $\Phi : E \times [0,1] \times \mathbb{Z}^+ \to E$ es una *función de actualización* de la cadena si verifica:

1. Para todo $i \in E$ y $n \geq 1$ el mapa $x \longmapsto \Phi(i,x,n)$ es constante a trozos.
2. Dados $i,j \in E$ y $n \geq 1$, el conjunto $\{x : \Phi(i,x,n) = j\}$ tiene medida $p_i^{(n)}(j)$, es decir:

$$\int_0^1 \mathbb{1}_{\{\Phi(i,x,n)=j\}} dx = p_{ij}^{(n)} \ , \ \forall i,j \in E, \ \forall n \geq 1.$$

Nótese que si la cadena es homogénea (es decir si $P^{(n)} = P \ \forall n \geq 1$), esta definición es equivalente a la anterior. Luego para $X_n = i$ y tomando una variable uniforme U_{n+1} se define $X_{n+1} = \Phi(i, U_{n+1}, n+1)$. La verificación de $P(X_n = j | X_{n-1} = i) = p_{ij}^{(n)}$ es análoga a la del caso homogéneo.

También se puede adaptar la construcción explícita de una función de actualización: basta tomar $\Phi : E \times [0,1] \times \mathbb{Z}^+ \to E$ tal que para $i \in E$ y $n \geq 1$

$$\Phi(i,x,n) = \begin{cases} 1, & \text{si } x \in [0, p_{i1}^{(n)}], \\ 2, & \text{si } x \in [p_{i1}^{(n)}, p_{i1}^{(n)} + p_{i2}(n)], \\ \vdots & \vdots \\ k, & \text{si } x \in [\sum_{h=1}^{h=k-1} p_{ih}^{(n)}, 1]. \end{cases}$$

Nuevamente la demostración de que Φ es una función de actualización es similar al caso homogéneo.

Observación 3.2.6. El método antes descrito permite -en teoría- simular cadenas de Markov finitas en general. Sin embargo en la práctica resulta poco eficiente cuando el espacio de estados E es grande (problema que no tendremos en las aplicaciones que veremos en el próximo capítulo). Por otra parte, algoritmos de Montecarlo basados en el comportamiento asintótico de las cadenas de Markov (MCMC, por sus siglas en inglés) pueden utilizarse para simular eficientemente variables discretas con espacio de estados finito pero grande. Detalles sobre la implementación y convergencia de los mismos pueden encontrarse en [8].

Capítulo 4

Aplicación a la generación de música

4.1. Motivación general

En este capítulo aplicaremos lo estudiado hasta el momento en la generación de música aleatoria, tentativamente, tonal. Más aún, intentaremos que nuestro modelo "aprenda" a partir de obras dadas.

Nuestra pregunta principal es ¿hasta qué punto es posible lograr crear aleatoriamente piezas musicales que respondan a un cierto estilo? Más allá de la falta de una definición clara de "estilo", cabe preguntarse cómo identificar automáticamente los rasgos que son característicos del mismo y aprender de ellos para generar nuevas "obras". Elegimos como estrategia para generar nuestras piezas usar cadenas de Markov con restricciones de modo similar al propuesto por Pachet, Roy y Barbieri en [18]. Las mismas serán creadas tomando como base una cadena de Markov homogénea cuyos parámetros se estimarán a partir de un *corpus* adecuado.

Un objetivo más simple puede ser simplemente generar aleatoriamente melodías musicalmente "coherentes" pero evitando el problema de la extracción de características de un *corpus*. En este caso consideraremos una melodía preexistente y generaremos aleatoriamente "variaciones" de la misma.

En todos los casos la correcta elección de las restricciones tiene un papel clave, siendo importante que las mismas respondan a criterios musicales. En nuestro caso elegiremos las restricciones manual y automáticamente, siendo el caso manual en el que aprovecharemos más la información que la teoría musical nos aporta. La implementación automática de restricciones, deseable para obras de mayor extensión, será musicalmente más elemental y las posibles estrategias para mejorarlas constituyen un amplio tema de estudio que no abordaremos en este trabajo.

4.2. Nociones previas y consideraciones técnicas

4.2.1. Acordes, notas y clases de octava

Consideremos un modelo –bastante simplificado– en el cual un sonido está determinado por dos parámetros:

- **Altura:** depende la frecuencia del sonido percibido y se asocia con la condición de *grave* o *agudo* del mismo.
- **Duración:** es la longitud del sonido en el tiempo.

Supondremos que los valores que estos parámetros toman son discretos. Se le llama *nota* a un par (altura, duración). Dado que nuestro trabajo modela sólo las alturas, la palabra "nota" aparecerá también en referencia a alturas. En general las duraciones se asumirán conocidas.

En este contexto, llamaremos melodía a una secuencia de notas (alturas) y supondremos que se comportan como una cadena de Markov es decir, que cada nota depende –a lo sumo– de la anterior. Para delimitar el espacio de estados y determinar los parámetros a utilizar necesitaremos presentar algunos conceptos más.

Comencemos por las notas. Como se mencionó antes, asumiremos que el conjunto de alturas posibles es discreto, y se identifica cada una de ellas con (al menos) un nombre del siguiente modo:

En teoría se tienen infinitas notas -el teclado puede extenderse arbitrariamente a ambos lados aunque en la práctica dicha cantidad está acotada ya que, por ejemplo, podemos excluir los sonidos cuya frecuencia no es audible para el oído humano. Por otra parte, separamos el conjunto de las notas en *octavas* (cada una de las cuales tiene 12 sonidos distintos). En distintas octavas se encuentran notas con igual nombre con lo cual si se quiere distinguirlas hay que especificar la octava. Así "do" hace referencia a varias notas pero "do4" refiere a uno específico (el cual es conocido como *do central*).

Para formalizar esto podemos definir una relación $\sim$ entre notas, donde dos notas están relacionadas si y sólo si tienen igual nombre (sin considerar la octava). Así se tiene, por ejemplo re3 $\sim$ re4.

Observación 4.2.1. La relación $\sim$ antes definida es una relación de equivalencia. Llamaremos *clase de octava* a sus clases de equivalencia.

La división en octavas no es arbitraria. Las notas que se encuentran en la misma clase de octava son altamente consonantes, lo cual hará que en algunas situaciones sea más adecuado considerar la clase de octava sin distinguir representantes.

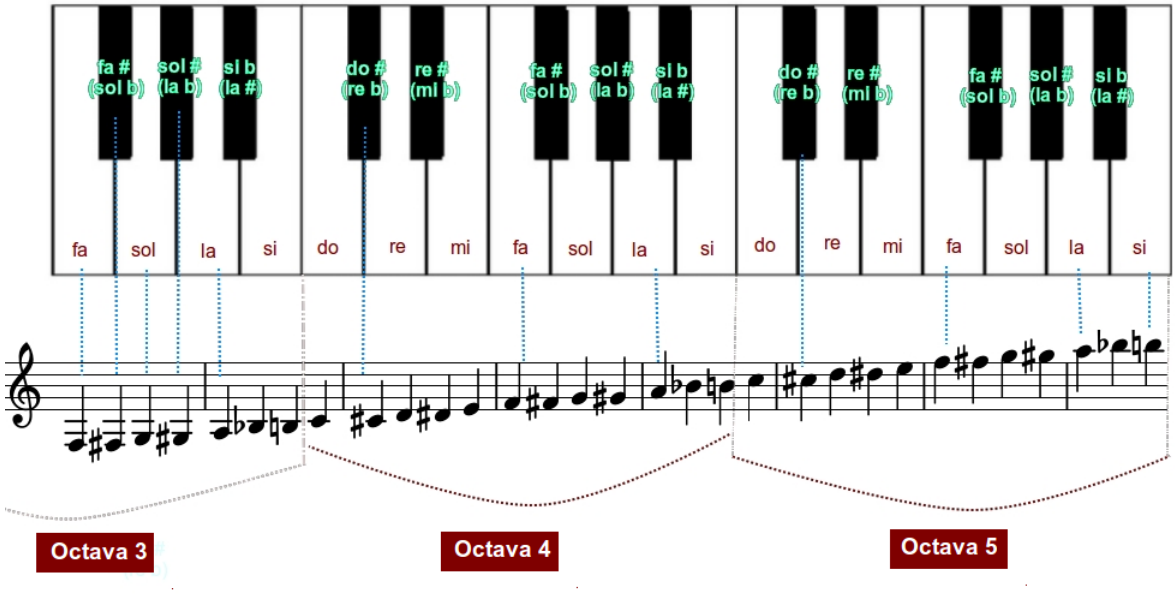


Figura 4.1: Notas según su nombre, ubicación en un teclado y su representación en el pentagrama.

En el tipo de música con el que trabajaremos (conocida como *música tonal*) las notas están fuertemente jerarquizadas, distribuyendo sus roles en torno a una nota principal denominada tónica. Para que nuestros ejemplos respeten esta jerarquía o bien elegiremos manualmente el subconjunto de notas a utilizar, o bien utilizaremos todas las notas disponibles, estimando sus probabilidades de aparición a partir del corpus.

Otros conceptos importantes son los de *armonía* y *acorde*. Un acorde suele definirse como un conjunto de notas tocadas simultáneamente, sin embargo, incluso en las melodías en las que no hay superposición de voces suele haber siempre acordes subyacentes, cada uno de los cuales tiene una función y se combinan de acuerdo a ciertas reglas. Armonía refiere justamente a esa combinación de acordes, así como a la disciplina que trata cómo combinarlos.

No nos detendremos en la clasificación de acordes ni en los criterios para determinar un acorde que no siempre es explícito. A los efectos de comprender el trabajo basta entender el acorde como conjunto de (habitualmente 3 o 4) notas, por lo general identificadas por su clase de octava. Mientras rige un acorde, las notas de uso prioritario en la melodía son precisamente las notas que lo integran y el uso de las demás notas queda supeditado a éstas (es decir, se usan como ornamentos o puentes entre notas del acorde).

4.2.2. Software utilizado

Para la simulación de las cadenas de Markov y la implementación de las restricciones se trabajó con el lenguaje R [6], mientras que para la conversión de las trayectorias a melodías y la manipulación del corpus se utilizó el lenguaje python con la biblioteca *music21* [13].

Cabe notar que dicha biblioteca dispone de un *corpus* del cual tomaremos obras para el nuestro. Así, disponemos de dichas obras (o cualquier otra que se esté dispuesto a ingresar manualmente) en un formato propio de music21 (objetos de tipo *Stream* o *Score*) donde las notas con sus parámetros, por ejemplo, son objetos cuyos atributos pueden ser modificados. Se puede también iterar a lo largo de las notas de una obra, o de las obras de un *corpus*, entre otros aspectos útiles para este trabajo. *music21* nos permite también exportar sus objetos como audio (MIDI) o partitura (con el apoyo de un editor de partituras compatible) a fin de obtener una salida legible. En este caso se utilizó el software *Lilypond* [11] para generar las partituras.

El código utilizado, así como algunos ejemplos adicionales, pueden encontrarse en el anexo.

4.3. Variaciones sobre Arroz con leche

Como el título lo indica, en esta primera aplicación se considera una melodía ya existente ("Arroz con leche"), lo suficientemente breve como para poder trabajar manualmente con ella. Pese a su brevedad y a la ausencia de polifonía, la pieza permite explorar algunos problemas centrales en nuestro modelo (como por ejemplo la elección de las restricciones y de los estados a considerar). A partir de esta melodía generaremos otras nuevas que preservan ciertos elementos de la original:

- La estructura rítmica, es decir, la cantidad y duraciones de las notas originales.
- Algunas notas –convenientemente elegidas– de la melodía.
- La estructura armónica, es decir, los acordes subyacentes.

De este modo se pretende obtener melodías que puedan percibirse como "variaciones" del tema del cual provienen. Para ello se redefinirán las alturas utilizando una cadena de Markov con restriciones unitarias sobre las notas que queremos preservar o, equivalentemente, varias cadenas de Markov concatenadas con una única restricción al final. Para preservar la estructura armónica consideraremos como espacio de estados únicamente notas del acorde subyacente (lo cual será eficaz pero algo restrictivo)[1].

[1]Un ejemplo basado en un enfoque más libre, sin consideraciones armónicas, puede verse en el anexo

Finalmente, para generar los trayectos aleatorios se proponen inicialmente paseos al azar con restricciones unitarias como los vistos en el ejemplo 2.4.1 del siguiente modo: consideremos primero nuestra melodía inicial, "Arroz con leche" y sus acordes subyacentes como se aprecia en la figura.

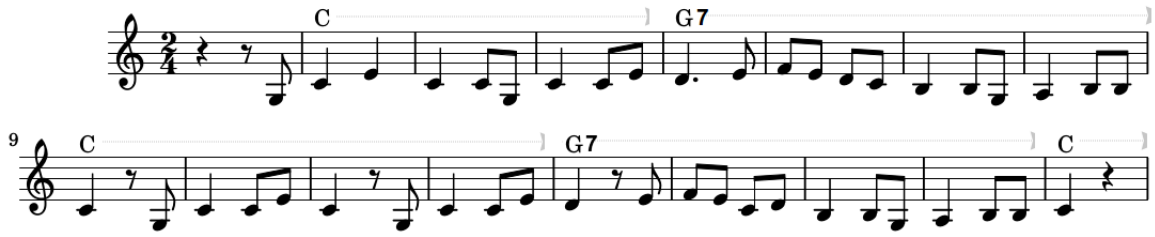


Figura 4.2: Partitura de la melodía de Arroz con leche, con sus acordes cifrados arriba.

Aquí C refiere al acorde de do Mayor (cuyas notas son do-mi-sol) y G7 a sol séptima (con notas sol-si-re-fa). Observando los acordes, la melodía nos queda separada en cinco regiones. Para generar las nuevas melodías conservaremos la primer nota de la pieza y la primera de cada región, generando las demás mediante paseos al azar con restricciones en $\mathbb{Z}$ que comienzan y terminan en 0. Asociamos enteros con notas de la siguiente forma:

- En las regiones delimitadas por el acorde C, consideramos como 0 la nota do4 y se sube o baja siempre a la nota más cercana que pertenezca a dicho acorde.
- En las regiones delimitadas por el acorde G7, consideramos como 0 la nota re4 y se sube o baja siempre a la nota más cercana que pertenezca al acorde.

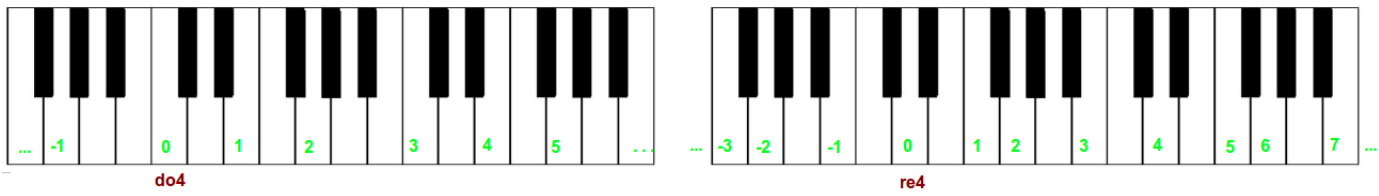


Figura 4.3: Esquema de las notas asociadas a cada entero según el acorde. En el teclado izquierdo: acorde C, en el derecho: acorde G7

Observemos que según las restricciones consideradas queremos cuatro paseos al azar que vayan del estado 0 al 0 en 8, 12, 10 y 12 transiciones respectivamente. En todas las regiones es posible realizar tales trayectorias, pero podría no ser así. En general podría haber problemas de paridad entre el largo de los tramos y la distancia entre estados como se vio en el ejemplo 2.4.1. Para evitar este problema modificaremos el paseo al azar permitiendo transiciones de cada estado a si mismo. Para implementarlo se utiliza el algoritmo visto en el teorema 2.4.2. Las probabilidades $p_i(j)$ de transición de cadena homogénea a restringir se eligen como

$$p_i(i+1) = \frac{18}{43}, \quad p_i(i) = \frac{8}{43}, \quad p_i(i-1) = \frac{17}{43},$$

y las demás 0. Dichos números se estimaron a partir de la melodía original, contando la cantidad de transiciones hacia un estado superior, igual o inferior respectivamente. Veamos la partitura de una de las variaciones generadas.

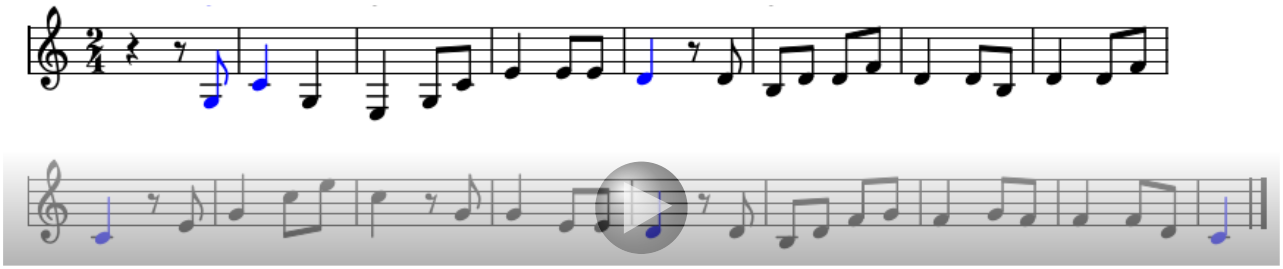


Figura 4.4: Una de las variaciones simuladas. En azul las notas que se preservaron de la melodía original. El audio se encuentra en *Ejemplos/Arroz con leche/arroz4_5.mp3* del repositorio anexo

Nótese que la elección de las notas asociadas a cada estado se hace de forma manual, al igual que la determinación de los acordes y la región asociada a éstos. Una vez determinado esto se pueden generar automáticamente tantas variaciones como se quiera, sin embargo no es una estrategia viable cuando las piezas con las que se trabaja son muy extensas. En el siguiente ejemplo se resolverán esos aspectos de forma automática.

4.4. Los corales de J.S. Bach

Como se mencionó al principio del capítulo, uno de los objetivos es intentar "aprender" ciertos rasgos de un estilo dado y, a diferencia de lo realizado en el ejemplo anterior, no se quiere ajustar manualmente las restricciones o el espacio de estados. Para ello utilizaremos "todas" las notas, confiando en que la estimación a partir del corpus le dará a cada una de ellas las probabilidades adecuadas.

Para realizar adecuadamente dicha estimación necesitamos que nuestro corpus -además de estar contenido en un mismo "estilo"- sea lo suficientemente grande y a su vez esté disponible en un formato que nos permita operar con sus notas sin realizar la transcripción manual de las piezas. Esto, entre otras razones, motiva la elección de los corales de J.S. Bach.

4.4.1. Descripción del corpus

Johann Sebastian Bach (1685-1750) fue un compositor alemán que, entre otras cosas, realizó una destacada armonización de melodías litúrgicas luteranas. Las obras resultantes son corales

de cuatro voces (soprano, contralto, tenor y bajo) en los que se puede reconocer un riguroso tratamiento en la conducción de las voces. Si bien no se conoce con exactitud el número exacto de armonizaciones creadas por Bach [2], se estiman entre 216 y 420 corales. Una edición póstuma de 371 corales realizada por Carl Philipp Emanuel Bach (hijo de J.S. Bach) llevó a que éstos sean considerados como un posible total.

Por otra parte, el antes mencionado *corpus* de *music21* contiene 347 de estos corales, de modo que podremos manipularlos con facilidad.

Elegiremos para nuestro *corpus* un subconjunto de 163 corales compuestos fundamentalmente en modo mayor[3] y se transportan todos a una misma tonalidad. El transporte permite que cada nota/acorde (por ejemplo, cada do) tenga en los distintos corales la misma jerarquía y función de modo que la estimación sobre todo el corpus tenga sentido.

Los corales pueden verse como cuatro líneas melódicas superpuestas de modo tal que determinan los acordes que rigen la obra. Dada la importancia del aspecto armónico y la escasa importancia de la elección de octava a la hora de determinar un acorde, realizaremos el análisis de los corales observando sólo la clase de octava.

En resumen, las principales caractéristicas del corpus a tener en cuenta son:

- Es una selección de 163 corales mayores de J.S. Bach, transportados a do Mayor.
- Los mismos cuentan con cuatro voces: soprano, contralto, tenor y bajo.
- El formato en que vienen dados es como objetos –denominados *stream*– de *music21*.

A partir de este corpus intentaremos generar nuevos corales mayores modelándolos como trayectorias de una cadena de Markov con restricciones.

4.4.2. Estimación e implementación de las restricciones

Para generar los nuevos corales estimaremos a partir del corpus los parámetros de una cadena de Markov homogénea utilizando lo visto en la sección 3.1. Tomando un coral del corpus como referencia, conservaremos su estructura métrica y fijaremos ciertas notas de referencia (aunque eventualmente transportadas a otro tono) dando lugar a cadenas con restricciones que nos permitirán generar "variaciones" del coral elegido.

[2]El número varía según la fuente consultada ya que algunas obras son incluídas o excluídas de la categoría por ser totalmente inéditas, estar repetidas o por existir dudas sobre la autenticidad de las ediciones disponibles

[3]Para realizar la detección de modo de forma automática se utiliza la función que Music21 designa a tales efectos. Sin embargo la misma no detecta posibles cambios de tonalidad a lo largo de la obra, problema que discutiremos brevemente más adelante.

Para esta primera etapa se considerará únicamente la clase de octava de las notas. Las principales razones para ello son:

- Armónicamente es adecuado no distinguir notas que difieren en octavas. Tampoco las transiciones que involucran notas equivalentes.

- Se reduce el espacio de estados, haciendo que dispongamos de una cantidad de datos estadísticamente más significativa.

Con respecto a lo último cabe recordar que hay en total 12 clases de octava. Por lo tanto nuestro espacio de estados tiene 12 elementos. Por otra parte, supondremos que cada voz se rige por una cadena de Markov (posiblemente) diferente. De este modo tendremos que estimar los parámetros de cuatro cadenas de Markov (una para cada voz). Llamaremos P_S, P_A, P_T y P_B a las matrices de transición asociadas a la voz soprano, contralto, tenor y bajo respectivamente (en la cadena homogénea). Hablaremos luego de la distribución inicial.

Así, para hallar $\hat{P}_S$, se considera la secuencia de estados correspondiente a las notas de la voz soprano para cada uno de los corales. Se obtienen en total 163 secuencias de las cuales se estima $\hat{P}_S$ usando el procedimiento visto en 3.1.2. De modo similar se hallan $\hat{P}_A$, $\hat{P}_T$ y $\hat{P}_B$.

Estas matrices son auxiliares y corresponden a cadenas de Markov homogéneas. Serán siempre las mismas independientemente de qué coral elijamos como "esqueleto" para la simulación. Para aplicar las restricciones obsérvese que musicalmente tiene sentido pensar en cierta homogeneidad en el tiempo, pero considerando los *compases*. Un compás puede pensarse como una división de la melodía en bloques de igual duración temporal. Como no todas las notas tienen igual duración, el número de notas por compás es variable. Dentro del compás, cada *tiempo* (unidad temporal que subdivide al compás) tiene su jerarquía, siendo generalmente el primer tiempo el más fuerte.

Por esta razón nuestras restricciones serán la primer nota de cada compás (que no necesariamente está en el primer tiempo pero suele estarlo) y la última de todo el coral (para dar sensación de final). Es decir que esas serán las notas que preservaremos del coral elegido. En particular eso quiere decir que la distribución inicial luego de aplicar las restricciones será determinista, y por esa razón no nos preocuparemos en estimarla.

Tras la estimación y previo redondeo obtenemos las siguientes matrices

$$\hat{P}_S = \begin{pmatrix}
0{,}254 & 0 & 0{,}348 & 0 & 0{,}065 & 0{,}009 & 0 & 0{,}095 & 0 & 0{,}029 & 0{,}006 & 0{,}194 \\
0 & 0{,}0625 & 0{,}8750 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}0625 \\
0{,}340 & 0{,}008 & 0{,}144 & 0 & 0{,}431 & 0{,}026 & 0 & 0{,}038 & 0 & 0{,}009 & 0{,}001 & 0{,}003 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0{,}067 & 0 & 0{,}392 & 0 & 0{,}130 & 0{,}315 & 0{,}016 & 0{,}052 & 0 & 0{,}027 & 0 & 0{,}001 \\
0{,}018 & 0 & 0{,}038 & 0 & 0{,}537 & 0{,}062 & 0 & 0{,}333 & 0 & 0{,}012 & 0 & 0 \\
0 & 0 & 0 & 0 & 0{,}296 & 0 & 0 & 0{,}694 & 0{,}010 & 0 & 0 & 0 \\
0{,}107 & 0 & 0{,}036 & 0 & 0{,}079 & 0{,}225 & 0{,}042 & 0{,}257 & 0{,}001 & 0{,}230 & 0{,}003 & 0{,}020 \\
0 & 0 & 0 & 0 & 0{,}091 & 0 & 0 & 0 & 0{,}182 & 0{,}727 & 0 & 0 \\
0{,}045 & 0{,}001 & 0{,}021 & 0 & 0{,}038 & 0{,}012 & 0{,}001 & 0{,}408 & 0{,}007 & 0{,}190 & 0{,}027 & 0{,}250 \\
0{,}232 & 0 & 0{,}070 & 0 & 0 & 0 & 0 & 0{,}093 & 0 & 0{,}605 & 0 & 0 \\
0{,}482 & 0 & 0{,}013 & 0 & 0{,}011 & 0 & 0 & 0{,}050 & 0 & 0{,}361 & 0 & 0{,}083
\end{pmatrix}$$

$$\hat{P}_A \begin{pmatrix}
0{,}354 & 0{,}003 & 0{,}180 & 0 & 0{,}018 & 0{,}017 & 0{,}001 & 0{,}060 & 0{,}001 & 0{,}021 & 0{,}021 & 0{,}325 \\
0{,}035 & 0{,}026 & 0{,}6 & 0{,}009 & 0{,}061 & 0 & 0 & 0 & 0 & 0{,}087 & 0 & 0{,}182 \\
0{,}319 & 0{,}049 & 0{,}223 & 0{,}004 & 0{,}230 & 0{,}011 & 0{,}031 & 0{,}068 & 0 & 0{,}033 & 0{,}001 & 0{,}031 \\
0 & 0 & 0{,}231 & 0 & 0{,}538 & 0{,}231 & 0 & 0 & 0 & 0 & 0 & 0 \\
0{,}036 & 0{,}005 & 0{,}303 & 0{,}005 & 0{,}196 & 0{,}248 & 0{,}076 & 0{,}056 & 0{,}012 & 0{,}039 & 0 & 0{,}024 \\
0{,}040 & 0 & 0{,}029 & 0{,}001 & 0{,}437 & 0{,}142 & 0{,}004 & 0{,}318 & 0 & 0{,}029 & 0 & 0 \\
0{,}003 & 0 & 0{,}129 & 0 & 0{,}204 & 0{,}014 & 0{,}040 & 0{,}553 & 0{,}011 & 0{,}037 & 0 & 0{,}009 \\
0{,}061 & 0{,}000 & 0{,}041 & 0 & 0{,}027 & 0{,}111 & 0{,}083 & 0{,}388 & 0{,}005 & 0{,}215 & 0{,}003 & 0{,}066 \\
0 & 0 & 0{,}008 & 0 & 0{,}114 & 0 & 0{,}049 & 0{,}065 & 0{,}008 & 0{,}675 & 0{,}016 & 0{,}065 \\
0{,}036 & 0{,}008 & 0{,}028 & 0 & 0{,}026 & 0{,}012 & 0{,}002 & 0{,}326 & 0{,}056 & 0{,}174 & 0{,}058 & 0{,}274 \\
0{,}263 & 0{,}006 & 0{,}012 & 0 & 0{,}006 & 0{,}041 & 0 & 0{,}029 & 0 & 0{,}491 & 0{,}129 & 0{,}023 \\
0{,}459 & 0{,}011 & 0{,}032 & 0 & 0{,}021 & 0{,}001 & 0{,}002 & 0{,}101 & 0{,}002 & 0{,}279 & 0{,}001 & 0{,}091
\end{pmatrix}$$

$$\hat{P}_T = \begin{pmatrix}
0{,}267 & 0{,}004 & 0{,}258 & 0{,}001 & 0{,}032 & 0{,}029 & 0{,}001 & 0{,}079 & 0 & 0{,}024 & 0{,}029 & 0{,}276 \\
0{,}119 & 0{,}024 & 0{,}631 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}071 & 0 & 0{,}155 \\
0{,}307 & 0{,}029 & 0{,}179 & 0{,}012 & 0{,}273 & 0{,}028 & 0{,}012 & 0{,}081 & 0{,}000 & 0{,}055 & 0 & 0{,}024 \\
0 & 0 & 0{,}364 & 0{,}030 & 0{,}212 & 0{,}333 & 0 & 0{,}061 & 0 & 0 & 0 & 0 \\
0{,}039 & 0{,}002 & 0{,}300 & 0{,}002 & 0{,}193 & 0{,}275 & 0{,}027 & 0{,}070 & 0{,}002 & 0{,}052 & 0{,}001 & 0{,}037 \\
0{,}046 & 0{,}001 & 0{,}032 & 0{,}004 & 0{,}420 & 0{,}074 & 0{,}003 & 0{,}375 & 0 & 0{,}037 & 0{,}002 & 0{,}006 \\
0{,}005 & 0{,}005 & 0{,}083 & 0{,}005 & 0{,}244 & 0{,}010 & 0{,}015 & 0{,}546 & 0{,}024 & 0{,}005 & 0 & 0{,}058 \\
0{,}046 & 0{,}001 & 0{,}065 & 0 & 0{,}048 & 0{,}205 & 0{,}040 & 0{,}378 & 0{,}001 & 0{,}182 & 0{,}005 & 0{,}029 \\
0{,}045 & 0 & 0{,}091 & 0 & 0{,}114 & 0 & 0{,}159 & 0{,}091 & 0{,}023 & 0{,}454 & 0 & 0{,}023 \\
0{,}037 & 0{,}002 & 0{,}060 & 0 & 0{,}064 & 0{,}034 & 0{,}002 & 0{,}367 & 0{,}025 & 0{,}167 & 0{,}067 & 0{,}175 \\
0{,}339 & 0 & 0 & 0{,}012 & 0{,}006 & 0{,}018 & 0 & 0{,}042 & 0 & 0{,}494 & 0{,}059 & 0{,}030 \\
0{,}460 & 0{,}013 & 0{,}033 & 0 & 0{,}092 & 0{,}021 & 0{,}021 & 0{,}072 & 0{,}001 & 0{,}200 & 0{,}004 & 0{,}083
\end{pmatrix}$$

$$\hat{P}_B = \begin{pmatrix} 0{,}175 & 0{,}008 & 0{,}168 & 0{,}001 & 0{,}050 & 0{,}083 & 0{,}008 & 0{,}234 & 0{,}004 & 0{,}035 & 0{,}018 & 0{,}216 \\ 0{,}127 & 0{,}009 & 0{,}686 & 0{,}034 & 0 & 0{,}009 & 0 & 0{,}008 & 0 & 0{,}025 & 0 & 0{,}102 \\ 0{,}273 & 0{,}042 & 0{,}054 & 0{,}007 & 0{,}321 & 0{,}041 & 0{,}006 & 0{,}155 & 0{,}008 & 0{,}079 & 0{,}003 & 0{,}011 \\ 0 & 0{,}075 & 0{,}225 & 0 & 0{,}5 & 0{,}05 & 0 & 0{,}05 & 0 & 0{,}05 & 0{,}025 & 0{,}025 \\ 0{,}071 & 0{,}003 & 0{,}309 & 0{,}011 & 0{,}030 & 0{,}339 & 0{,}044 & 0{,}041 & 0{,}009 & 0{,}128 & 0 & 0{,}015 \\ 0{,}141 & 0{,}001 & 0{,}031 & 0{,}001 & 0{,}365 & 0{,}038 & 0{,}010 & 0{,}359 & 0{,}001 & 0{,}038 & 0{,}011 & 0{,}004 \\ 0{,}027 & 0 & 0{,}039 & 0 & 0{,}240 & 0{,}081 & 0 & 0{,}523 & 0{,}058 & 0{,}019 & 0 & 0{,}012 \\ 0{,}235 & 0{,}004 & 0{,}108 & 0{,}001 & 0{,}033 & 0{,}219 & 0{,}059 & 0{,}106 & 0{,}011 & 0{,}200 & 0{,}001 & 0{,}023 \\ 0{,}038 & 0 & 0{,}032 & 0 & 0{,}057 & 0{,}025 & 0{,}139 & 0{,}171 & 0{,}006 & 0{,}526 & 0 & 0{,}006 \\ 0{,}058 & 0{,}005 & 0{,}101 & 0{,}003 & 0{,}116 & 0{,}022 & 0{,}002 & 0{,}276 & 0{,}055 & 0{,}072 & 0{,}041 & 0{,}249 \\ 0{,}295 & 0{,}014 & 0 & 0 & 0{,}007 & 0{,}034 & 0{,}007 & 0{,}048 & 0{,}007 & 0{,}521 & 0{,}055 & 0{,}014 \\ 0{,}520 & 0{,}010 & 0{,}012 & 0 & 0{,}032 & 0{,}008 & 0 & 0{,}054 & 0{,}001 & 0{,}347 & 0{,}014 & 0{,}002 \end{pmatrix}$$

donde la i-ésima fila/columna corresponde al estado i, y asociamos cada número con una nota de la octava. Más específicamente se asocia el estado 1 a la nota do, el 2 a do# y así sucesivamente hasta el 12, que se corresponde con la nota si.

Obsérvese que las columnas correspondientes a las notas naturales (es decir *do*, *re*, *mi*, *fa*, *sol*, *la* y *si*) son las que acumulan más probabilidad. Esto se debe a que son las notas que aparecen con más frecuencia en los corales, como se puede apreciar en el siguiente histograma:

Figura 4.5: Frecuencia relativa de las 12 clases de octava en los corales mayores de Bach.

Una vez estimadas las cuatro matrices auxiliares, aplicando las restricciones (fijar las notas elegidas) se obtienen otras tantas familias de matrices de transición que son las que se utilizarán para generar las nuevas trayectorias.

4.4.3. Generación de nuevos corales

Una vez que se tienen las matrices de transición que se utilizará para las nuevas trayectorias, éstas pueden generarse con el procedimiento visto en la sección 3.2. Sin embargo la salida obtenida serán secuencias de números entre 1 y 12 los cuales se corresponden con notas vistas como clase de octava. Este hecho, que fue conveniente para la estimación, no lo es tanto a la hora de escribir la

nueva melodía ya que para ello no basta con indicar la clase de octava sino que hay que explicitar el representante.

Por otra parte una vez determinadas las notas de las nuevas melodías, basta reemplazarlas en el coral que se eligió previamente como referencia (el mismo que se utilizó para determinar las restricciones). El único asunto a resolver es, por lo tanto, la adecuada elección de las octavas para cada nota.

Cabe recordar que se está generando música coral, y por lo tanto cada voz tiene un rango de alturas posibles (denominado *registro*). Dichos rangos están en el siguiente entorno:

- De *do4* a *do6* para la soprano.
- De *mi3* a *re5* para la contralto.
- De *si2* a *sol4* para el tenor.
- De *mi2* a *do4* para el bajo.

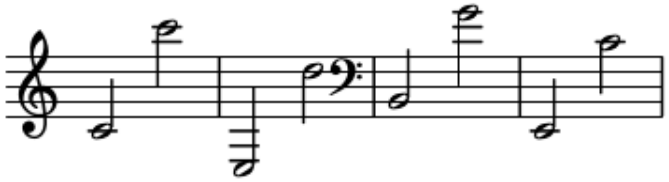


Figura 4.6: Los registros utilizados para soprano, contralto, tenor y bajo respectivamente.

Para elegir la octava de cada nota utilizaremos el siguiente criterio: se impone la octava de la primera nota (octavas 5,4,3 y 3 para soprano, contralto, tenor y bajo respectivamente), mientras que para las demás se elige la octava que, sin salirse del registro, minimice la distancia respecto a la altura de la nota anterior.

Esta estrategia, aunque bastante adecuada, provoca algunos comportamientos no propios de los verdaderos corales de Bach, por ejemplo porque permite que las voces se crucen (es decir, que por momentos una voz grave quede por encima de una voz aguda). En la figura 4.7 podemos ver un fragmento de partitura de un ejemplo de coral simulado.

Nótese que pese a que permanece escrita la armadura de clave del coral original, el coral simulado está en do Mayor. No es relevante mantener la tonalidad del coral ya que luego es posible transportarlo a cualquier otro tono mayor.

En particular, las notas que se espera encontrar en do Mayor son *do, re, mi, fa, sol, la, si* y *do* (es decir, únicamente las correspondientes a teclas blancas del piano). Notas ajenas a la tonalidad (denominadas *notas cromáticas*) pueden aparecer eventualmente en los corales con un propósito concreto que suele ser:

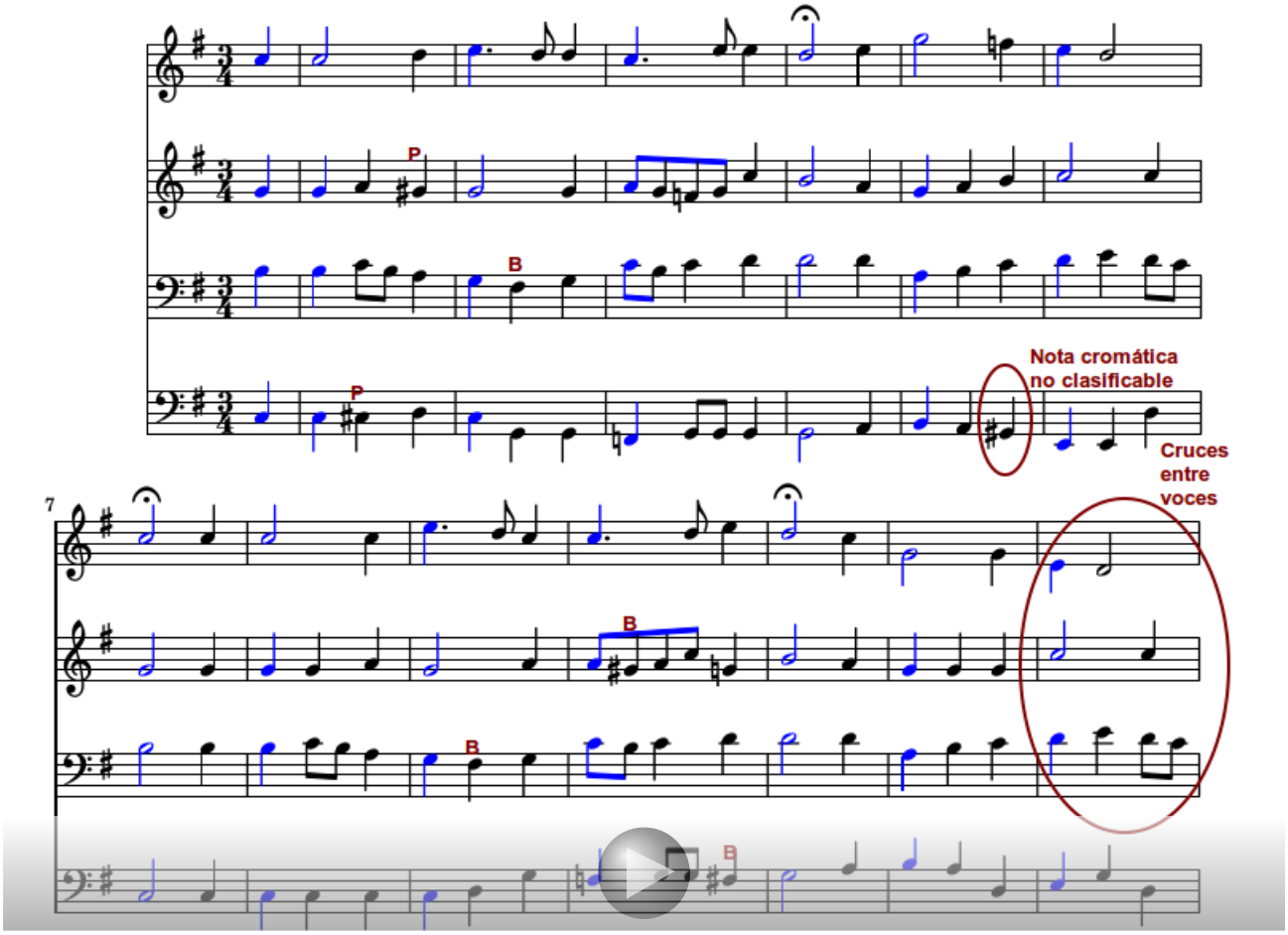


Figura 4.7: Fragmento de coral simulado en base al bwv269 de Bach. En azul las notas fijadas por las restricciones. Se señalan y clasifican las notas cromáticas y se indican el sector donde hay cruces entre las voces. La partitura completa y el audio pueden encontrarse en la carpeta *Ejemplos/Corales* del anexo bajo el nombre *ejem3_5_mayor*.

- Notas ornamentales: ofician como adorno y suelen estar subordinadas a notas de la tonalidad y su uso debe hacerse siguiendo varias normas. Dentro de las notas ornamentales se destacan
  - *Notas de paso*: funcionan como "puente" entre notas de la tonalidad.
  - *Bordaduras cromáticas*: se encuentran a un semitono (en general por debajo) de una nota de la tonalidad, la cual precede y sucede a la bordadura. Suelen ocurrir en tiempos débiles.
  - *Apoyaturas cromáticas*: se encuentran también a un semitono (por debajo) de una nota de la tonalidad, y son sucedidas por dicha nota. Ocurren en tiempos fuertes.
- Modulaciones (o dominantes secundarias): en este caso se tiene un cambio de tonalidad, generalmente provisorio y, en el caso de las denominadas dominantes secundarias, muy

breve.

En nuestro caso nos interesan las notas ornamentales, mientras que las modulaciones introducen un nuevo problema: no tenemos herramientas para determinar automáticamente cuándo ocurrió una modulación, de modo que cuando decimos que un coral está en cierta tonalidad debemos decir que está *predominantemente* en dicha tonalidad. Así, las modulaciones en el *corpus* introducen notas y transiciones ajenas a la tonalidad que no se comportan como ornamentaciones. En la figura 4.7 se indican con P, B y Ap las notas de paso, bordaduras y apoyaturas respectivamente, y se señala explícitamente la única nota cromática del fragmento que no puede clasificarse como ornamental.

4.5. Independencia entre voces

Un aspecto cuestionable del enfoque anteriormente utilizado para generar los corales de la seccíon anterior es que las voces se generan de forma independiente. Músicalmente no hay razones para suponer esto (de hecho la ausencia de cruces entre las voces indica que las mismas son dependientes en los corales originales), sin embargo la alternativa más inmediata es suponer que se tiene una sola cadena de Markov en vez de 4, cuyo espacio de estados es un subconjunto finito de $\mathbb{R}^4$ (es decir, las clases de octava correspondientes a cada una de las voces).

El primer problema que esto presenta es que las cuatro voces no siempre atacan en simultáneo[4], sin embargo para resolverlo basta considerar en cada instante dado la nota que está sonando independientemente de si atacó en ese momento o no. En caso de tener un silencio en alguna voz, se puede suponer que es una prolongación de la nota anterior. Describiremos este proceso con más detalle más adelante.

El segundo problema radica en el tamaño del espacio de estados, el cual tendría 12^4 elementos. Aún considerando un corpus relativamente grande como el nuestro, en principio no tendríamos una cantidad total de transiciones razonablemente grande como para que la estimación tenga sentido mientras en el enfoque original estimábamos cuatro matrices 12×12 (un total de 576 entradas), en este caso estimaríamos una matriz $12^4 \times 12^4$ (un total de más de 400 millones de entradas)[5].

Cabe preguntarse entonces qué tan inadecuado es asumir que las voces se mueven independientemente en los corales de Bach. Musicalmente sabemos que algunas reglas -como el no cruce entre voces- indican dependencia entre voces y que son relevantes las nociones de *movimiento contrario* y *movimiento paralelo* entre pares de voces, en referencia a la relación entre sus movimientos ascendentes y descendentes.

[4]Por ataque entendemos el instante en que la nota empieza a sonar.

[5]Notar que, sin embargo, si una proporción significativa de esas entradas resultan nulas la estimación podría ser viable con lo cual la opción de considerar 4-uplas podría ser explorada.

Estudiaremos así la correlación de las subidas y bajadas entre pares de voces en los corales de Bach y realizaremos una prueba de hipótesis para ver si pueden considerarse independientes, para algunos corales específicos.

4.5.1. Cálculo de las correlaciones

Para determinar la correlación entre las subidas y bajadas de las voces en un coral construiremos primero cuatro secuencias auxiliares l^k , $k \in \{1, 2, 3, 4\}$ de modo que $l^k = l_1^k l_2^k \dots l_{n-1}^k$ (con n aún sin definir) y

$$l_i^k = \begin{cases} 1, & \text{si la altura de la } i\text{-ésima nota es menor a la de la } (i{+}1)\text{-ésima,} \\ -1, & \text{si la altura de la } i\text{-ésima nota es mayor a la de la } (i{+}1)\text{-ésima,} \\ 0, & \text{si ambas alturas son iguales.} \end{cases}$$

De este modo, si se tiene una melodía con n notas, ésta puede asociarse a una secuencia de $n-1$ valores 1, 0 y -1.

Sin embargo, como se mencionó al principio de la sección, las voces en los corales no atacan siempre en simultáneo lo cual hace que las secuencias que se obtienen de cada una de sus voces no correspondan a transiciones alineadas en el tiempo e incluso no tengan necesariamente la misma cantidad de valores. Para resolver esto, consideraremos una melodía auxiliar donde las voces se "refinan" de acuerdo al menor divisor común de las duraciones encontrada en la obra, como se ilustra en la figura 4.8. Lo que se hace es reescribir cada nota repetida tantas veces como sea necesario para que la duración total de las repeticiones sea equivalente a la original. De este modo los ataques siempre quedan sincronizados y tiene sentido construir las secuencias numéricas antes definidas.

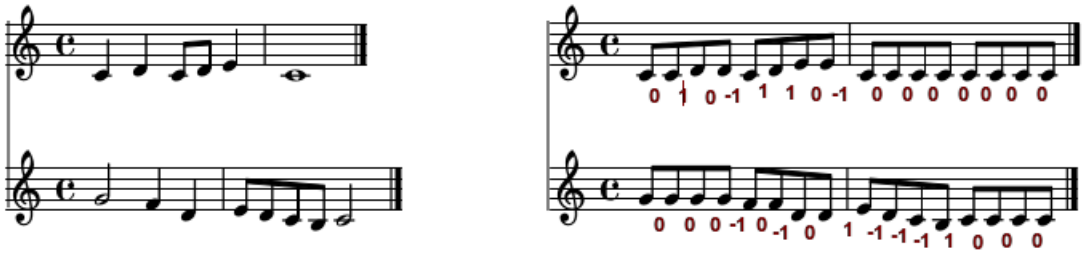


Figura 4.8: A la izquierda, ejemplo de un sistema con dos melodías breves. A la derecha las voces refinadas por su menor duración (corcheas) y las secuencias de números correspondientes.

Una vez obtenidas las secuencias numéricas queremos determinar la correlación entre ellas. Para ello utilizaremos el coeficiente de correlación de Pearson.

Definición 4.5.1. Sean $x = x_1, x_2, \dots, x_n$ e $y = y_1, y_2, \dots, y_n$, con $x_i, y_i \in \mathbb{R}\ \forall i \in \{1, \dots, n\}$ secuencias de datos. Definiremos su *coeficiente de correlación* r_{xy} como

$$r_{xy} = \frac{1}{n-1} \sum_{i=1}^{i=n} \left(\frac{x_i - \overline{x}}{s_x} \right) \left(\frac{y_i - \overline{y}}{s_y} \right)$$

donde $\overline{x}$ y $\overline{y}$ son los promedios de las secuencias y $s_x := \sqrt{\frac{1}{n-1} \sum_{i=1}^{i=n} (x_i - \overline{x})^2}$ es el desvío estándar muestral.

Obsérvese que si consideramos el producto escalar usual en $\mathbb{R}^n$, se tiene que $r_{xy} = \frac{<x-\overline{x}><y-\overline{y}>}{\|x-\overline{x}\|\|y-\overline{y}\|}$, que es por definición el coseno del ángulo entre los vectores x e y. Esto implica que $r_{xy} \in [-1, 1]$. Donde valores con módulo cercano a 1 indican una correlación lineal fuerte entre x e y, al contrario de los valores cercanos a 0. Secuencias independientes deberían arrojar coeficientes cercanos a 0, pero no vale el recíproco (correlación nula no implica independencia).

Por lo tanto un posible test de independencia entre x e y consiste en calcular r_{xy} y rechazar la independencia si su módulo supera cierto umbral. Sin embargo en nuestro caso tenemos 4 secuencias a comparar, por lo que habrá que utilizar un test que nos permita verificar independencia para más de dos secuencias.

Un primer acercamiento consiste en estudiar las correlaciones entre todos los pares de voces (i.e, todos los pares de secuencias numéricas construidos como en la figura 4.8), obteniendo así una matriz de correlaciones R de la forma

$$R := \begin{pmatrix} 1 & r_{12} & r_{13} & r_{14} \\ r_{12} & 1 & r_{23} & r_{24} \\ r_{13} & r_{23} & 1 & r_{34} \\ r_{14} & r_{24} & r_{34} & 1 \end{pmatrix}.$$

donde r_{ij} refiere a los coeficientes de correlación entre las voces 1 a 4 (soprano, contralto, tenor y bajo respectivamente). En particular se tiene para todo $i, j \in \{1, 2, 3, 4\}$ que $r_{ij} = r_{ji}$ con lo que la matriz es simétrica. Además $r_{ii} = 1 \; \forall i \in \{1, 2, 3, 4\}$.

En total son 6 coeficientes no triviales los necesarios para conocer la matriz. Comparemos los coeficientes de correlación de cada uno de los corales de Bach con los obtenidos a partir de un corpus simulado, es decir, 163 corales simulados, cada uno de ellos simulado como variación de un coral de Bach distinto.

En la figura 4.9 se aprecia que en los corales simulados la mediana de los coeficientes de correlación es cercana a 0 como era de esperarse. En el caso de los corales de Bach las medianas varían significativamente según las voces consideradas siendo sus valores 0,237, 0,007, −0,131, 0,208, −0,118 y −0,088 para los r_{12}, r_{13}, r_{14}, r_{23}, r_{24} y r_{34} respectivamente.

A primera vista esto ofrecería evidencia de que en los verdaderos corales sí existe correlación entre voces, aunque se requeriría de un criterio más riguroso para concluír. Por otra no sabemos qué ocurre coral a coral ¿Es posible que existan algunos corales en los que las voces no estén

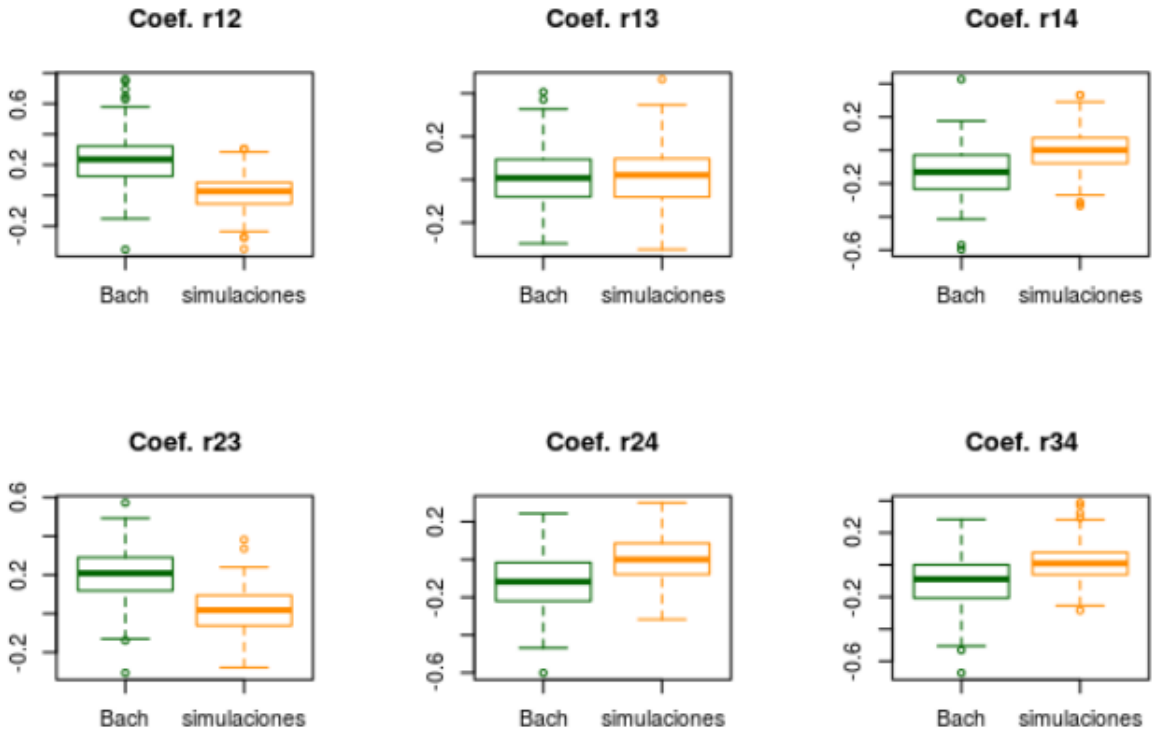


Figura 4.9: Diagramas de caja de las correlaciones de los 163 corales considerados (caja izquierda) y sus 163 versiones simuladas (derecha).

correlacionadas? Seleccionemos algunos corales y estudiemos un poco más en detalle la correlación entre sus voces.

4.5.2. Diseño del test y construcción del estadístico

Dado un coral de Bach, y sus secuencias de 1, 0 y -1 construidas como en la figura 4.8 nos proponemos determinar si las voces pueden considerarse independientes en términos de las subidas, bajadas y permanencias. Consideremos el test

$$\begin{cases} H_0: & \text{las secuencias son independientes} \\ H_1: & \text{No } H_0 \end{cases}$$

Cabe recordar que para que las cuatro secuencias sean independientes no basta con la independencia dos a dos, pero sí es una condición necesaria. Así, un criterio posible es descartar la independencia cuando, en total, el módulo de las correlaciones es grande. Definimos el estadístico

$$D = (r_{12})^2 + (r_{13})^2 + (r_{14})^2 + (r_{23})^2 + (r_{24})^2 + (r_{34})^2,$$

y rechazamos la hipótesis nula cuando D es mayor a un cierto umbral c_α. Como desconocemos la distribución F del estadístico bajo H_0 no podemos determinar umbrales exactos, pero sí aproximados. Para ello se generan observaciones independientes de D bajo H_0 y se calcula su distribución empírica.

Para generar observaciones de D bajo H_0 basta simular variaciones del coral elegido, y calcular D para cada una de ellas. Luego si queremos realizar el test al nivel α basta tomar el menor c_α de modo que

$$F_n(c_\alpha) \geq 1 - \alpha,$$

donde F_n es la distribución empírica. Así, $P_{H_0}(D > c_\alpha) \approx 1 - F_n(c_\alpha) \leq \alpha$.

A continuación veremos el resultado de aplicar el test sobre algunos corales.

Debido a que resulta computacionalmente costoso generar muchas observaciones del estadístico (en particular, generar y exportar muchas secuencias con restricciones), elegiremos unos pocos corales para realizar el test. Hallando primero los D para todos los corales de Bach vemos que tienen el siguiente comportamiento:

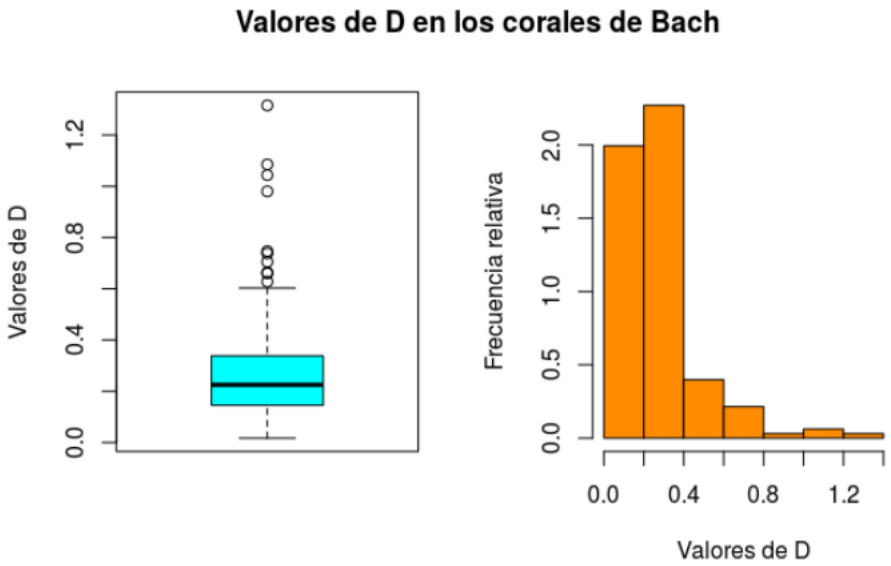


Figura 4.10: Diagrama de cajas e histograma para los 163 valores de D calculados a partir de los corales de Bach.

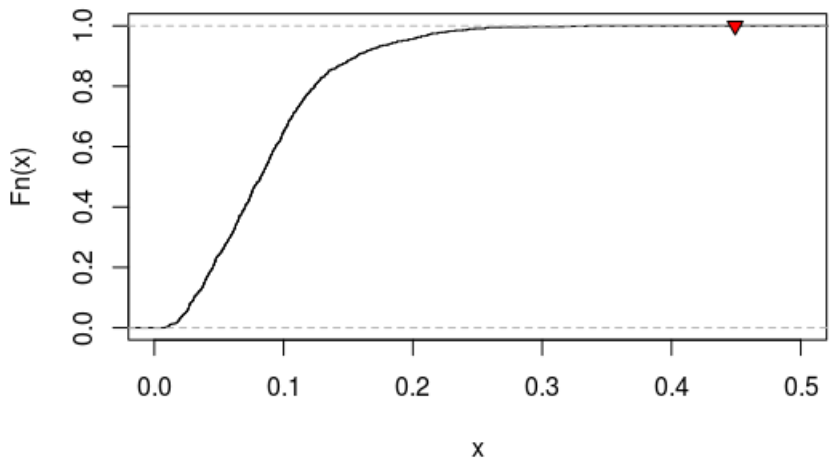
Además, se tiene que $\overline{D} = 0{,}273$ y los cuartiles son $q_1 = 0{,}1454$, $q_2 = 0{,}225$ y $q_3 = 0{,}338$.

Consideraremos pues un coral cuyo estadístico D pueda considerarse "grande" (mayor a q_3), otro no tan grande (entre q_2 y q_3) y uno "pequeño" (menor a q_1). Específicamente utilizaremos los corales:

- bwv133.6, con $D = 0{,}449$,
- bwv281, con $D = 0{,}320$,
- bwv325, con $D = 0{,}119$.

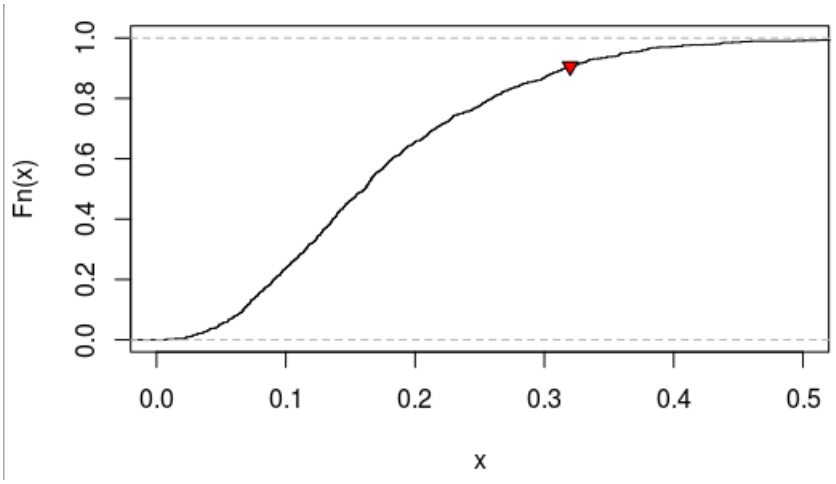
En todos los casos consideramos $n = 1000$, es decir, simulamos 1000 variaciones de cada uno para calcular la distribución empírica, obteniendo los siguientes resultados.

Para el primer caso (con $D = 0{,}449$):



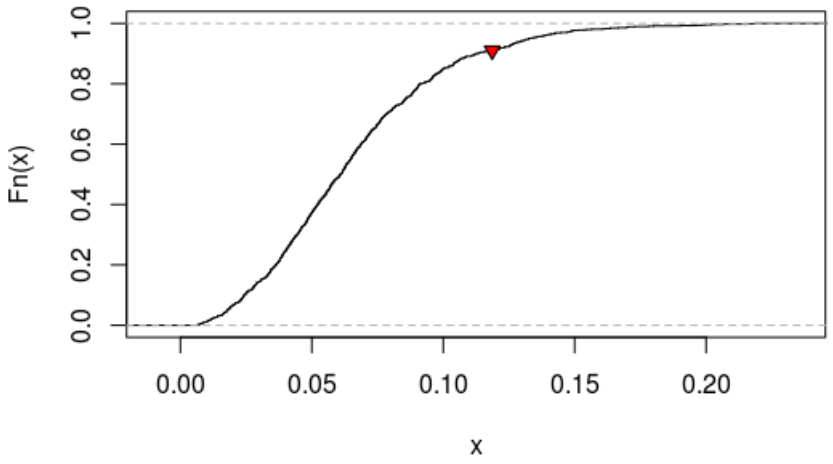
donde el triángulo rojo representa el punto $(D, F_n(D)) = (0{,}449, 1)$, con lo cual el p-valor para el test resulta aproximadamente 0. Resulta razonable ya que el estadístico del coral es grande. Veamos que ocurre en los otros dos casos.

Para el segundo coral, donde $D = 0{,}320$ tenemos:



con $Fn(D) = 0{,}906$. Entonces el p-valor será $\alpha^* \approx 0{,}094$.

Finalmente, en el tercer caso con $D = 0{,}119$ se obtuvo:



y $F_n(D) = 0{,}91$, con lo que $\alpha^* \approx 0{,}09$

Nótese que en los tres casos se rechaza H_0 si tomamos $\alpha = 0{,}1$, aunque para $\alpha = 0{,}05$ sólo se rechazaría en el primer caso. Sin embargo, con certeza del 90 % se podría decir que las subidas y bajadas de los corales de Bach considerados no son independientes

Capítulo 5

Conclusiones y nuevos problemas

A lo largo del trabajo se presentó una posible estrategia para la simulación de música a partir de un corpus dado. En tal caso logramos ver cómo el modelo preserva ciertos rasgos de la música tonal: la presencia de notas fuera de la tonalidad es escasa, así como también escasean las transiciones poco convencionales para el estilo (incluso aquellas que involucran notas dentro de la tonalidad). Un ejemplo de esto último puede observarse en las matrices de transición estimadas en los corales de Bach donde la transición de *fa* a *si* -intervalo conocido como tritono, evitado por su condición disonante- aparece con probabilidad pequeña o incluso nula en cualquiera de las 4 voces. Asimismo la aplicación de las restricciones al principio de cada compás constituye un primer acercamiento a formas automáticas de fijar las restricciones preservando información relevante (notar que el primer tiempo de cada compás es el más fuerte). En el caso de Arroz con Leche, la elección de restricciones y estados basadas en los acordes subyacentes dio como resultado melodías que presentan una reminiscencia a la original aunque resultan, en contraposición, demasiado simples en tanto utilizan únicamente notas del acorde.

Por otra parte, en el caso de los corales de Bach observamos una de las falencias (posiblemente la principal) del modelo: la falsa suposición de independencia entre voces. Como se mencionó previamente, resulta poco viable suponer que las variables son 4-uplas (o para música polifónica en general, n-uplas) debido al tamaño del espacio de estados resultantes, aunque es de esperarse que en la práctica muchas de esas n-uplas tengan probabilidad despreciable o nula: tratándose de música tonal, son muchas las combinaciones de notas que en la práctica casi no se utilizan. Puede ser un posible camino a explorar para la generación de música polifónica.

Otro posible enfoque para modelar los corales de Bach es pensar la soprano como melodía y las demás voces como armonía[1]. Cabe recordar que la mayoría de los corales son, precisamente, armonizaciones de melodías preexistentes con lo cual este enfoque tiene musicalmente más sentido que verlos como cuatro melodías transcurriendo en paralelo. En ese caso la voz melódica (soprano) puede modelarse con un enfoque similar al utilizado en el caso de Arroz con Leche:

[1]Un análisis estadístico de la armonía de los corales puede encontrarse en [20]

las restricciones son las encargadas de fijar la armonía, y entre restricción y restricción se tienen estados y transiciones posibles diferentes. Este enfoque introduce varias dificultades nuevas: para empezar, el cambio de espacio de estados y probabilidades de transición que en Arroz con Leche se hizo de forma manual deberá automatizarse, lo que posiblemente requiera del uso de cadenas de Markov moduladas u ocultas. Por otra parte, será necesario obtener un método que permita identificar automáticamente la ubicación de las restricciones mediante la identificación de los acordes subyacentes (algo que en el otro ejemplo, nuevamente, se determinó de forma manual) o bien, disponer de dicha información dada de antemano. Una vez resuelto esto, aún habrá que encontrar un modo de simular las demás voces, de modo no independiente.

Otro problema a considerar fue la presencia de modulaciones en el corpus y el ruido que éstas introdujeron en la estimación. Una posible -y relativamente sencilla- alternativa para afrontarlo consiste en prohibir aquellas transiciones que son "impropias" de la tonalidad considerada, entendiendo por impropias:

- Transiciones que por criterios de teoría musical no deberían aparecer en la tonalidad, o bien,
- transiciones cuya probabilidad estimada es muy pequeña (menor a cierto umbral), en el entendido de que son anómalas.

En ambos casos dicha prohibición puede implementarse utilizando restricciones binarias. A diferencia de las restricciones unitarias que elegimos para los corales, en este caso no tenemos certeza de poder aplicar tales restricciones de manera consistente (es decir, que el conjunto de corales originales que satisfacen las restricciones sea no vacío). Hay, pues, que determinar las restricciones con más cuidado. Cabe observar también que la implementación de estas nuevas restricciones conllevaría una pérdida de riqueza musical, ya que entre las transiciones que prohibiremos por considerar anómalas se encuentran también las que ornamentan la melodía más allá de las reglas rígidas impuestas por la tonalidad.

Por otra parte, otra suposición cuestionable de nuestro modelo fue la independencia entre alturas y duraciones de las notas. Se puede investigar qué tan correcta es esta suposición realizando tests sobre los corales. Asimismo, queda pendiente el diseño de un modelo que permita generar aleatoriamente las duraciones.

Algunos de los aspectos discutidos en esta sección quedan planteados como temas de estudio a futuro, donde se intentará mejorar los puntos débiles del trabajo y seguir explorando una línea de investigación no muy explorada a nivel local como es la aplicación de modelos matemáticos en música.

Anexo

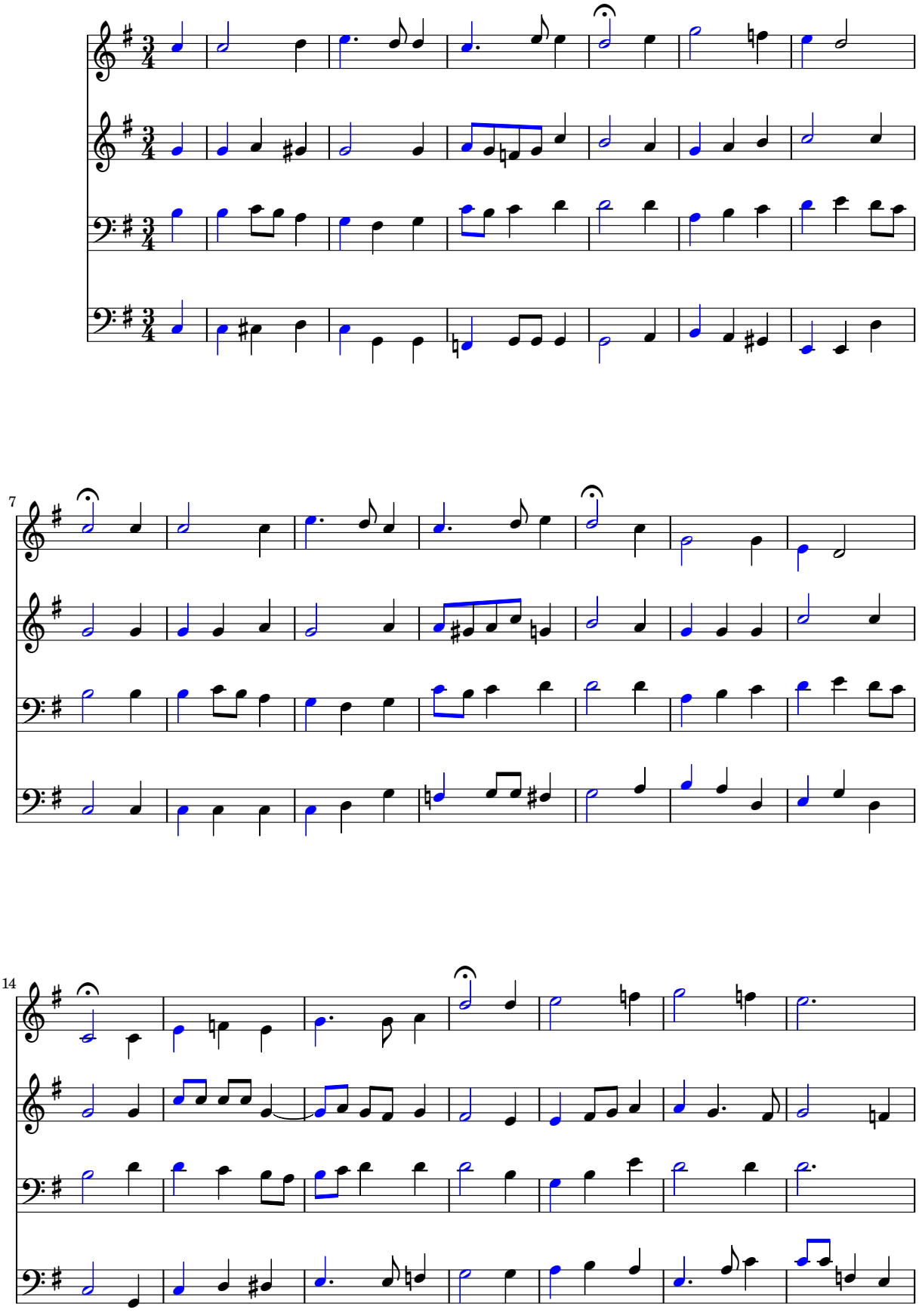
A continuación se pone a disposición algunas de las obras obtenidas por simulación así como un repositorio con más ejemplos (con archivo MIDI incluído) y el código utilizado para las simulaciones.

El material mencionado puede encontrarse en `www.cmat.edu.uy/~vrumbo/markovymusica`

Se trata de un archivo .zip conteniendo:

- El código utilizado para generar las partituras y scripts que facilitan al usuario la generación de nuevos ejemplos.
- El código utilizado para realizar el análisis estadístico de los corales (de Bach y simulados).
- Partituras (en formato pdf) y audios (en formato MIDI) con más ejemplos de Arroz con Leche y corales de Bach.
- Un archivo de texto (LEEME.txt) con descripción del contenido e instrucciones de ejecución.

Además se anexan partituras de algunos de los ejemplos. En todos los casos las restricciones se indican en azul.


7
14

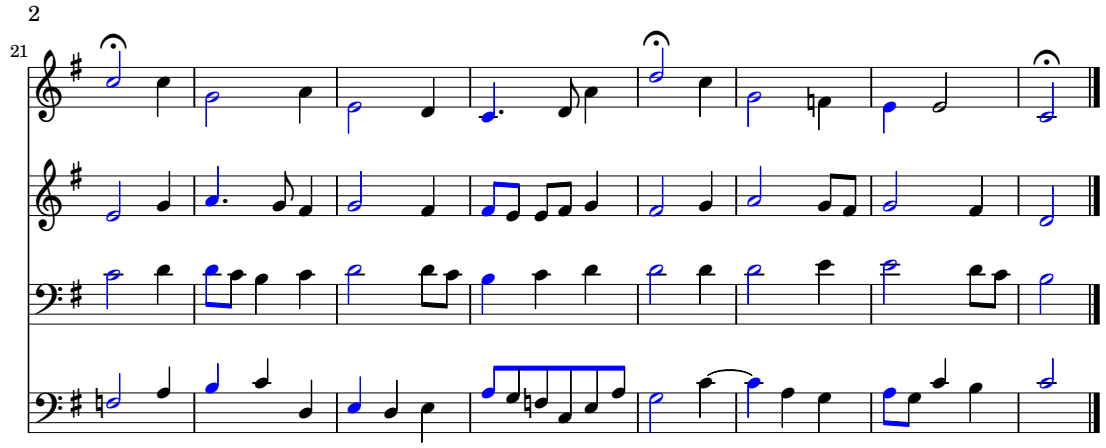


Figura A: Partitura completa del coral simulado en base al bwv269 de Bach ilustrado en la figura 4.7.

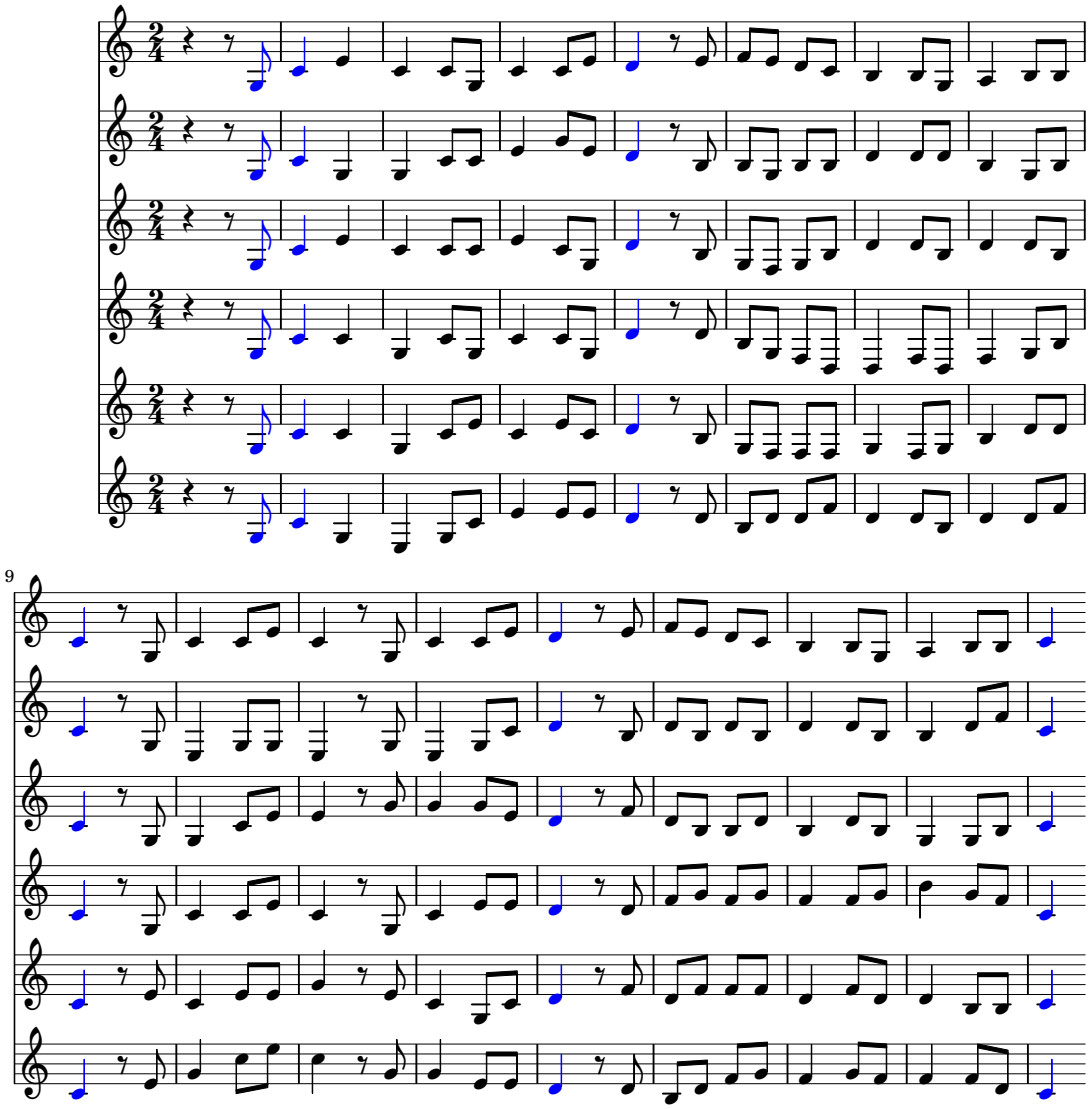


Figura B: Partituras de 5 variantes de “Arroz con Leche” simuladas con bajo el modelo planteado en la sección 4.3. La partitura superior es la melodía original.

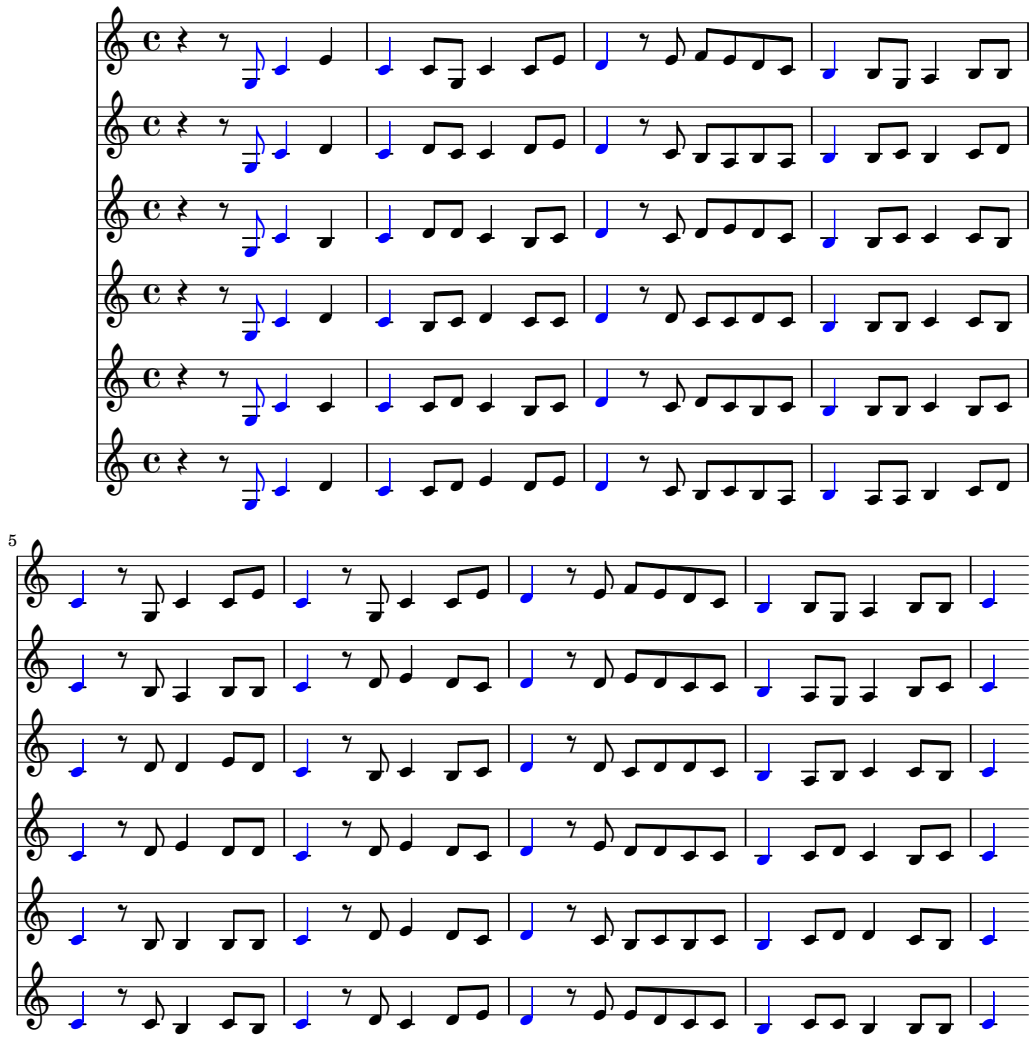


Figura B: Melodía original + otras 5 variantes de “Arroz con Leche”. Este ejemplo -implementado antes que el anterior- impone más restricciones pero no utiliza la información de los acordes para elegir las notas entre una y otra sino que el paseo se realiza dentro de toda la escala. El resultado son melodías en las que parece más difícil identificar la obra de la cual provienen, en tanto se perdió la estructura armónica. Esta situación motivó el enfoque adoptado en la sección 4.3.

Bibliografía

[1] Ames, Charles: *The Markov Process as a Compositional Model: a Survey and Tutorial*. Leonardo, 22(2):175–187, 1989.

[2] Borovkov, A A: *Estadística Matemática*. Editorial Mir Moscú, 1988.

[3] Chen, Ruofeng, Weibin Shen, Ajay Srinivasamurthy y Parag Chordia: *Chord Recognition Using Duration Explicit Hidden Markov Models*. 13th International Society for Music Information Retrieval Conference, 2012.

[4] Cope, David: *Emily Howell*. `http://artsites.ucsc.edu/faculty/cope/Emily-howell.htm`. Visitada el 30/1/2017.

[5] Cope, David: *Experiments in Musical Intelligence*. A-R Editions, Inc, 1996.

[6] *The Comprehensive R Archive Network*. `https://cran.r-project.org/`. Visitada el 30/1/2017.

[7] Dahn, Luke: *So how many Bach four part chorales are there?* `http://bach-chorales.com/HowManyChorales.htm`. Visitada el 30/1/2017.

[8] Häggström, Olle: *Finite Markov Chains and algorithmic applications*. Cambridge University Press, 2002.

[9] Johnson, Daniel: *Composing Music With Recurrent Neural Networks*. `http://www.hexahedria.com/2015/08/03/composing-music-with-recurrent-neural-networks/`. Visitada el 30/1/2017.

[10] Lebanon, Guy: *Consistency of the Maximum Likelihood Estimator*. `http://www.cc.gatech.edu/~lebanon/notes/mleConsistency.pdf`. Visitada el 14/2/2017.

[11] *Lilypond - Music notation for everyone*. `http://lilypond.org/`.

[12] Maurer, IV, John A.: *A Brief History of Algorithmic Composition*. `https://ccrma.stanford.edu/~blackrse/algorithm.htm"`, 1999. Visitada el 30/1/2017.

[13] *music21: A toolkit for computer-aided musicology*. `http://web.mit.edu/music21/`. Visitada el 30/1/2017.

[14] Moore, David S: *The Basic Practice of Statistics*. W.H. Freeman & Co. New York, NY, USA, 1999.

[15] Nierhaus, Gerhard: *Algorithmic Composition - Paradigms of Automated Music Generation*. SpringerWienNewYork, 2009.

[16] Norris, J R: *Markov Chains*. Cambridge University Press, 1997.

[17] Oron, Aryeh: *Chorale Melodies used in Bach's Vocal Works 371 Four-Part Chorales*. `http://www.bach-cantatas.com/CM/IndexCM-Breitkopf.htm`. Visitada el 30/1/2017.

[18] Pachet, François, Pierre Roy y Gabriele Barbieri: *Finite-Length Markov Processes with Constraints*. Proceedings of the Twenty-Second International Joint Conference on Artificial Itelligence, páginas 638–642, 2011.

[19] Petrov, Valentin y Ernesto Mordecki: *Teoría de la Probabilidad*. DIRAC, second edición, 2008.

[20] Rohrmeier, Martin y Ian Cross: *Statistical Properties of Tonal Harmony in Bach's Chorales*. Proceeding of the International Conference of Music Perception and Cognition, 2008.

[21] Wikipedia: *Musikalisches Würfelspiel*. `https://en.wikipedia.org/wiki/Musikalisches_W\%C3\%BCrfelspiel`. Visitada el 30/1/2017.