



UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO

---

---

FACULTAD DE CIENCIAS

MODELACIÓN, CLASIFICACIÓN Y SIMULACIÓN  
DE OBRAS MUSICALES

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIA

PRESENTA:

DANIELA PORTILLO DEL VALLE

DIRECTOR DE TESIS:

DR. SERGIO IVÁN LÓPEZ ORTEGA



Ciudad Universitaria, Cd. Mx., 2019



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

## Hoja de datos del jurado

### 1. Datos del alumno

Portillo  
del Valle  
Daniela  
56 79 46 04  
Universidad Nacional Autónoma de México  
Facultad de Ciencias  
Actuaría  
312049399

### 2. Datos del tutor

Dr.  
Sergio Iván  
López  
Ortega

### 3. Datos del sinodal 1

Dr.  
Pablo  
Padilla  
Longoria

### 4. Datos del sinodal 2

Dra.  
Guillermina  
Eslava  
Gómez

### 5. Datos del sinodal 3

Dra.  
Lizbeth  
Naranjo  
Albarrán

### 6. Datos del sinodal 4

Dr.  
Carlos Erwin  
Rodríguez  
Hernández-Vela

### 7. Datos del trabajo escrito

Modelación, Clasificación y Simulación de Obras Musicales.

219 p.  
2019

*A mis padres, los luceros de mi vida,*

*fruto de su esfuerzo y amor incondicionales.*



*“La Música es un ejercicio aritmético del alma de carácter inconsciente.”*

– G. W. Leibniz, *carta a Goldbach (17 de abril de 1712)*

*“Tal vez sea la Música la matemática del sentimiento y la matemática la Música de la razón.”*

– J. J. Sylvester, *On Newton’s Rule for the Discovery of Imaginary Roots*

*“El matemático puro, como el músico, es creador libre de su mundo de belleza ordenada.”*

– B. Russell, *A History of Western Philosophy*



# ÍNDICE GENERAL

<b>Resumen</b>	<b>1</b>
<b>Introducción y objetivos</b>	<b>3</b>
<b>1 En búsqueda de la representación matemática de una pieza musical</b>	<b>7</b>
1.1 Historia de la Música	8
1.2 Principios musicales	10
1.3 Elementos de la Música	13
1.4 Teoría de la armonía	15
1.5 Consonancia y disonancia	19
1.6 Teoría Musical de Conjuntos	22
1.7 Herramientas musicológicas propuestas	30
<b>2 Cadenas de Markov</b>	<b>39</b>
2.1 Conceptos básicos	40
2.2 Teorema de convergencia	43
2.3 Estimación por Máxima Verosimilitud	45
<b>3 Campos Aleatorios</b>	<b>47</b>
3.1 Campos Aleatorios de Markov	49
3.2 Campos Aleatorios de Gibbs	52
3.3 Estimación paramétrica	57

<b>4</b>	<b>Aprendizaje Estadístico</b>	<b>63</b>
4.1	Regresión Multinomial	64
4.2	Regresión Ordinal	67
4.3	Métodos de Regularización	68
4.4	Análisis de Discriminante	69
4.5	$k$ -Vecinos Más Cercanos.	71
4.6	Árboles de Clasificación	72
4.7	Análisis de Componentes Principales	76
4.8	Análisis de Conglomerados	79
4.9	Validación de Modelos y Remuestreo	81
<b>5</b>	<b>Modelación matemática de obras musicales</b>	<b>83</b>
5.1	Caracterización mediante Cadenas de Markov	87
5.2	Caracterización mediante Campos Aleatorios de Gibbs y Modelos de Markov	91
5.3	Análisis exploratorio de las estimaciones	102
<b>6</b>	<b>Clasificación de obras musicales</b>	<b>125</b>
6.1	Clasificación por corriente musical	126
6.2	Clasificación por corrientes históricas con armonía tonal	151
6.3	Clasificación por autor	158
<b>7</b>	<b>Simulación de obras musicales</b>	<b>169</b>
	<b>Conclusiones</b>	<b>175</b>
	<i>Anexo I. Clasificación de Allen Forte</i>	<b>181</b>
	<i>Anexo II. Aplicación de la modelación propuesta sobre una obra musical</i>	<b>185</b>
	<i>Anexo III. Observaciones atípicas</i>	<b>195</b>
	<i>Anexo IV. Análisis de Componentes Principales adicional</i>	<b>197</b>
	<b>Bibliografía</b>	<b>201</b>

## ÍNDICE DE FIGURAS

### En búsqueda de la representación matemática de una pieza musical

1.1	Ejemplos de pentagramas.	10
1.3	Onda de sonido de la nota $La_0$ .	10
1.4	Octava en el piano con nombres de las notas.	11
1.5	Notas naturales de la cuarta octava.	11
1.6	Duraciones canónicas de las notas.	12
1.7	Duraciones canónicas de los silencios.	13
1.8	Intervalos básicos entre dos notas y su notación.	16
1.9	Ejemplo de inversiones del acorde <i>Do Mayor (C)</i> .	16
1.10	Triadas y séptimas comunes de <i>Do (C)</i> .	17
1.11	Acordes y grados de la tonalidad <i>Do mayor</i> .	18
1.12	Transposiciones del conjunto tonal asociado a <i>FaM7</i> .	27
1.13	Medidas de consonancia y disonancia planteadas aplicadas a <i>Polonaise Op. 89 en Do mayor</i> de Beethoven.	36

### Cadenas de Markov

2.1	Dinámica de ganancia o pérdida del jugador: espacio de estados, sus clases de comunicación y probabilidades de transición.	43
-----	--	----

### Campos Aleatorios

3.1	Ejemplo de configuración de la red neuronal artificial modelada mediante Campos Aleatorios.	48
-----	---	----

3.2	Sistema de vecindades tipo $\alpha$ sobre un conjunto de sitios bidimensional.	49
-----	--	----

### **Modelación matemática de obras musicales**

5.1	Primeros cuatro compases de <i>Polonaise Op. 89 en Do Mayor</i> de Beethoven y su transformación para fines de la modelación.	86
5.2	Primeras tres configuraciones observadas del Campo Aleatorio asociado a <i>Polonaise Op. 89 en Do Mayor</i> de Beethoven.	93
5.3	Potenciales de activación (medidas de consonancia y disonancia) asociados a una sucesión de tres acordes respecto la consonancia del último acorde.	95
5.4	Energía del campo respecto las variaciones en los parámetros $\theta_{\mathcal{C}}$ y $\theta_{\mathcal{D}}$ para algunas progresiones de acordes.	99
5.5	Estimaciones de las probabilidades límite de las Cadenas de Markov asociadas a las duraciones de las intervenciones de las obras musicales en la base de datos.	103
5.6	Estimaciones de las probabilidades límite de las Cadenas de Markov asociadas a las raíces de las intervenciones de las obras musicales en la base de datos.	104
5.7	Estimaciones de las probabilidades límite de las Cadenas de Markov asociadas a la Clase Forte de las intervenciones de las obras musicales en la base de datos.	106
5.9	Dendrogramas resultantes del Análisis de Conglomerados jerárquico de las probabilidades límite estimadas de las Cadenas de Markov asociadas a las duraciones, notas fundamentales y Clase Forte.	109
5.10	Histogramas de los valores estimados $\hat{\theta}_{\mathcal{C}}$ y $\hat{\theta}_{\mathcal{D}}$ y gráfico de caja y violín de las estimaciones de la función de log-Verosimilitud de las obras musicales de la base de datos.	111
5.11	Gráficos de dispersión de los parámetros estimados $\hat{\theta}_{\mathcal{C}}$ y $\hat{\theta}_{\mathcal{D}}$ , individualmente y por corriente histórica a la que pertenece y por su autor.	112
5.12	Porcentaje de varianza explicada y su acumulación respecto el número de Componentes Principales.	113
5.13	<i>Biplots</i> para los primeros dos Componentes Principales por estimaciones: parámetros $\theta_{\mathcal{C}}$ y $\theta_{\mathcal{D}}$ , evaluación de la función de log-Verosimilitud y probabilidades límite asociadas a duraciones, notas fundamentales y Clases Forte.	114
5.14	Estimaciones de las probabilidades límite correspondientes a las Cadenas de Markov asociadas a la duración, raíz y Clase Forte de las intervenciones de una obra musical por corriente histórica.	116

5.15	Proyección de las observaciones sobre los tres primeros Componentes Principales dos a dos por corriente histórica a la que pertenecen.	119
5.16	Estimaciones de las probabilidades límite correspondientes a las Cadenas de Markov asociadas a la duración, raíz y Clase Forte de las intervenciones de una obra musical por autor.	120
5.17	Proyección de las observaciones sobre los tres primeros Componentes Principales dos a dos por su autor.	123

### **Clasificación de obras musicales**

6.1	Tasas de error de clasificación por corriente musical, globales y por grupo, validadas y aparente global obtenidas mediante Regresión Multinomial respecto al número de Componentes Principales.	128
6.2	Coefficientes de regresión estimados y sus intervalos de confianza por <i>Cross-Validation</i> del modelo de Regresión Multinomial para la clasificación por corriente musical.	130
6.3	Proyección de las probabilidades estimadas por corriente musical obtenidas mediante Regresión Multinomial validadas sobre el primer PC, la evaluación de la función de log-Verosimilitud estimada por medidas de Gibbs y los factores de consonancia y disonancia.	132
6.4	Tasas de error de clasificación por corriente musical, globales y por grupo, validadas y aparente global obtenidas mediante Regresión Multinomial Regularizada respecto al número de Componentes Principales considerados para el ajuste del modelo.	134
6.5	Tasas de error de clasificación por corriente musical, globales y por grupo, validadas y aparente global obtenidas mediante Regresión Ordinal respecto al número de Componentes Principales considerados para el ajuste del modelo.	136
6.6	Funciones de densidad y densidad acumulada validadas de la variable latente asociada a la variable respuesta del modelo de Regresión Ordinal para clasificación por corriente histórica.	137
6.7	Proyección sobre el primer Componente Principal de las probabilidades estimadas por corriente musical resultantes del modelo de Regresión Ordinal.	138
6.8	Árboles de clasificación por corriente musical.	140
6.9	Criterio de costo-complejidad, aparente y validado, respecto el tamaño de los Árboles de Clasificación por corriente musical.	141

6.10 Tasas de error de clasificación por corriente musical, globales y por grupo, validadas y aparente global obtenidas mediante <i>Bagging</i> , <i>Random Forests</i> y <i>Boosting</i> , respecto al número de Componentes Principales considerados para el ajuste del modelo.	142
6.11 Tasas de error de clasificación por corriente musical, globales y por grupo, validadas obtenidas mediante <i>k</i> -Vecinos más Cercanos, respecto al número de vecinos.	144
6.12 Tasas de error de clasificación por corriente musical, globales y por grupo, validadas de los modelos de LDA y QDA respecto al número de Componentes Principales.	145
6.13 Tasas de clasificación errónea, validadas y aparentes, asociadas a los modelos de Regresión Multinomial <i>Elastic Net</i> ajustados sobre los Componentes Principales para clasificación por corriente histórica restringida a Música con armonía tonal.	156
6.14 Coeficientes de regresión estimados y sus intervalos de confianza por <i>Cross-Validation</i> del modelo de Regresión Multinomial <i>Elastic Net</i> para la clasificación por corriente histórica restringida a Música con armonía tonal.	157
6.15 Coeficientes de regresión estimados y sus intervalos de confianza por <i>Cross-Validation</i> del modelo de Regresión Multinomial para la clasificación por autor.	163
6.16 Tasas de error de clasificación validadas, global y por grupo, respecto al número de variables a tomar en cuenta para el ajuste de Árboles de Clasificación en cada iteración del modelo de <i>Random Forests</i> para clasificación por autor.	166
6.17 Importancia relativa de las variables originales en el modelo de <i>Random Forests</i> para clasificación por autor.	167

### **Simulación de obras musicales**

7.1 Primeros dos compases de la simulación considerando melodía y acompañamiento independientes sobre <i>Maple Leaf Rag</i> de Scott Joplin.	172
--	-----

### **Anexo IV. Análisis de Componentes Principales adicional**

A.1 Matriz de correlación de las estimaciones de las probabilidades límite, factores de consonancia y disonancia y evaluación de la función de log-Verosimilitud en términos de la medida de Gibbs.	198
A.2 Proyección de las observaciones a los primeros dos Componentes Principales obtenidos sin reescalar las variables.	200

## ÍNDICE DE CUADROS

### **En búsqueda de la representación matemática de una pieza musical**

- |     |  |    |
|-----|--|----|
| 1.1 | Permutaciones circulares del conjunto tonal [0 4 5 9].   | 24 |
| 1.2 | Categorización de las entradas de los vectores de clase de intervalo de acuerdo con la consonancia contrapuntística. | 30 |

### **Modelación matemática de obras musicales**

- |     |   |     |
|-----|---|-----|
| 5.1 | Cantidad de obras musicales y autores en la base de datos a utilizar por corriente histórica.   | 84  |
| 5.2 | Cadenas de Markov a considerar para la primera modelación, sus espacios de estados y su denotación.   | 88  |
| 5.3 | Estimaciones de $\theta_{\mathcal{C}}$ , $\theta_{\mathcal{D}}$ y de la evaluación de la función de log-Verosimilitud para las cinco obras musicales con mayores diferencias entre las valuaciones de su Verosimilitud estimadas al variar el orden del modelo. | 113 |
| 5.4 | Estimaciones de los factores de consonancia y disonancia por corriente histórica.   | 118 |
| 5.5 | Estimaciones de los factores de consonancia y disonancia por autor.   | 122 |

### **Clasificación de obras musicales**

- |     |   |     |
|-----|---|-----|
| 6.1 | Piezas musicales por clase en las bases de datos para la clasificación por corriente histórica.   | 127 |
| 6.2 | Devianza, AIC, BIC y tasa de clasificación errónea global, aparentes y validados, asociados a los modelos de Regresión Multinomial para clasificación de obras musicales por corriente histórica. | 129 |

6.3	Tasas de clasificación errónea globales, aparentes y validadas, asociadas a los modelos de Regresión Multinomial Regularizada ajustados para clasificación de obras musicales por corriente histórica.	135
6.4	Devianza, AIC, BIC y tasa de clasificación errónea global, aparentes y validados, asociados a los modelos de Regresión Ordinal para clasificación de obras musicales por corriente histórica.	136
6.5	Estimaciones validadas, error estándar, intervalos de confianza, sesgo, estadístico del contraste de Wald y <i>p-value</i> de los puntos de corte obtenidos mediante Regresión Ordinal para clasificación por corriente histórica.	137
6.6	Diferencia de logaritmos de las probabilidades acumuladas estimadas obtenidas mediante Regresión Multinomial para la verificación del supuesto de momios proporcionales del modelo ordinal para la clasificación por corrientes musicales .	139
6.7	Tasas de clasificación errónea globales, aparentes y validadas, obtenidas mediante Árboles de Clasificación para corrientes musicales.	143
6.8	Tasas de clasificación errónea globales y por grupo validadas obtenidas mediante <i>k</i> -Vecinos más Cercanos para corrientes musicales.	144
6.9	Tasas de clasificación errónea globales, aparentes y validadas, obtenidas mediante LDA y QDA para corrientes musicales.	146
6.10	Matriz de confusión asociada a la clasificación por corriente histórica y tasas de clasificación errónea aparentes por grupo mediante el clasificador Bayesiano sobre los factores de consonancia y disonancia.	147
6.11	Tasas de error de clasificación globales, aparentes y validadas, correspondientes a clasificación por corriente musical histórica.	148
6.12	Tasas de error de clasificación por grupo validadas correspondientes a clasificación por corriente musical histórica.	150
6.13	Tasas de error de clasificación globales validadas para clasificación por corriente musical histórica restringida a Música con armonía tonal.	152
6.14	Tasas de error de clasificación globales, aparentes y validadas, correspondientes a clasificación por corriente musical histórica restringida a Música con armonía tonal.	153
6.15	Tasas de error de clasificación por grupo validadas correspondientes a clasificación por corriente histórica restringida a Música con armonía tonal por todos los métodos presentados.	154
6.16	Cantidad de piezas musicales por clase en las bases de datos para clasificación por autor.	158

6.17 Tasas de error de clasificación globales validadas para clasificación por autor.	159
6.18 Tasas de error de clasificación globales, aparentes y validadas, correspondientes a clasificación por autor.	160
6.19 Tasas de error de clasificación por grupo validadas por <i>Cross-Validation</i> correspondientes a clasificación por autor.	161
6.20 AIC, BIC, Análisis de Devianza y tasa de error de clasificación global, aparentes y validados, del modelo de Regresión Multinomial para la clasificación por autor.	162
6.21 Matriz de confusión y tasas de clasificación errónea, aparentes y validadas, por grupo del modelo de Regresión Multinomial para la clasificación por autor.	165
6.22 Matriz de confusión y tasas de clasificación errónea, aparentes y validadas, por grupo del modelo de <i>Random Forests</i> para la clasificación por autor.	167

### **Simulación de obras musicales**

7.1 Ejemplos de simulación musical llevados a cabo: su obra musical base y el nombre del archivo para su consulta.	173
--	-----

### **Conclusiones**

C.1 Tasas de clasificación acertada globales, aparentes y validadas, de los mejores modelos para la clasificación por corriente musical histórica, por corriente restringida a Música con armonía tonal y por autor.	179
--	-----

### **Anexo I. Clasificación de Allen Forte**

A.1 Clases Forte y sus conjuntos tonales en forma básica asociados observados en la base de datos.	181
--	-----

### **Anexo II. Aplicación de la modelación propuesta sobre una obra musical**

A.2 Transformación de las intervenciones en los primeros dos compases de <i>Polonaise Op. 89 en Do Mayor</i> de Beethoven y las Cadenas de Markov asociadas a su duración, raíz y Clase Forte.	186
A.3 Matriz de transición de las Cadenas de Markov asociadas a las duraciones y notas fundamentales de las intervenciones de <i>Polonaise Op. 89 en Do Mayor</i> de Beethoven.	187
A.4 Probabilidades límite de las Cadenas de Markov asociadas a las duraciones y notas fundamentales de las intervenciones de <i>Polonaise Op. 89 en Do Mayor</i> de Beethoven.	188
A.5 Transformación de las intervenciones en los primeros dos compases de <i>Polonaise Op. 89 en Do Mayor</i> de Beethoven y sus vectores de clase de intervalo reducidos.	190

**Anexo III. Observaciones atípicas**

- A.6 Observaciones atípicas en la base de datos con base en el último Componente Principal correspondiente al análisis por corriente musical histórica. 195
- A.7 Observaciones atípicas en la base de datos con base en el último Componente Principal correspondiente al análisis por autor. 196

**Anexo IV. Análisis de Componentes Principales adicional**

- A.8 Desviación estándar y varianza explicada por los Componentes Principales obtenidos sin reescalar las variables. 199
- A.9 *Loadings* asociados a los primeros dos Componentes Principales obtenidos sin reescalar las variables. 200

El objetivo del presente es la investigación de la estructura de obras musicales académicas desde un enfoque matemático con el fin de brindar evidencia cuantitativa que permita la realización de análisis musicológico objetivo. En específico, se proponen vías para la modelación de una pieza musical a través de un enfoque probabilístico y estadístico congruentes con la Teoría Musical de la Escuela Occidental. Como objetivos secundarios se fijaron la realización de simulación musical, así como clasificación de obras respecto a su autor y a la corriente musical histórica a la que pertenecen. El trabajo se divide en dos partes: planteamiento teórico y su aplicación –llevada a cabo mediante una base de datos de obras musicales–.

De acuerdo con la Musicología, una obra musical es determinada intrínsecamente por su armonía, de manera que se modelaron las tendencias en la composición de una pieza mediante tal concepto. Lo anterior, con el fin de cuantificarla en medida de que, de acuerdo con la Teoría Musical, a ella tienden intencionalmente las expresiones sucesivas del autor durante la realización de cada pieza. Dicho lo cual, la investigación sobre el contexto musical teórico llevó a conceptualizar a las obras musicales como sistemas dinámicos no deterministas que están completamente caracterizados por la asimetría de sus progresiones armónicas, la variabilidad entre su consonancia y disonancia, conceptos mediante los cuales el autor busca causar las sensaciones que les dan esencia.

En primer lugar, se caracterizó a las obras musicales mediante propiedades de sus intervenciones: notas fundamentales, Clases Forte y duraciones. Para tal planteamiento, es posible estimar por Máxima Verosimilitud las matrices de transición asociadas a los procesos de cada obra musical en la base de datos. Asimismo, tras obtener las matrices de transición de las cadenas, se procedió

a estimar sus probabilidades límite individualmente, así como su extensión a autores y corrientes históricas. Con ello, se cuantifica el tempo armónico de las obras musicales en medida de que tal conceptualización refleja el uso de las progresiones armónicas a largo plazo, a través de las probabilidades límite asociadas a cada cadena.

Ante la búsqueda de un modelo generalizado que, por un lado, relaje la propiedad Markoviana de las Cadenas de Markov y, por otro, describa los cambios entre consonancia y disonancia en una pieza musical de manera independiente a su escala, se recurrió a la teoría de Campos Aleatorios. Primero, se asumió que una pieza musical es un Modelo de Markov de orden  $k - 1$ . Posteriormente, conceptualizando a la consonancia y disonancia como la energía resultante de las interacciones de las intervenciones consecutivas de una obra musical, se modelaron como Campos Aleatorios dinámicos de Markov y de Gibbs. Para ello, se propusieron herramientas musicológicas originales que permitieran describir una pieza musical cuantitativamente, congruentes con la Teoría Musical: medidas de consonancia y disonancia. De manera que, a través de tales conceptos, fue posible caracterizar una obra musical mediante su medida de Gibbs en forma paramétrica –cuyos parámetros que pueden ser interpretados como factores de incurrencia en la consonancia y disonancia, así como ser estimados por Máxima Verosimilitud por obra, autor y corriente musical–.

Tras llevar a cabo las estimaciones relativas a la modelación por las dos vías propuestas, se construyó con ellas una base de datos mediante la cual se clasificó a las obras correspondientes por autor y por corriente musical histórica. Para tal finalidad, se hicieron dos Análisis de Componentes Principales sobre las observaciones originales: primero, con base en la matriz de covarianzas y, luego, con la matriz de correlación. Posteriormente, fueron aplicados los siguientes métodos de Aprendizaje Estadístico sobre los tres conjuntos de datos con miras hacia la minimización de las tasas de error de clasificación globales estimadas por Validación Cruzada: Regresión Multinomial, Regresión Ordinal, Árboles de Clasificación,  $k$ -Vecinos más Cercanos y Análisis de Discriminante.

Finalmente, se llevó a cabo simulación musical mediante Cadenas de Markov asociadas a la melodía, acompañamiento y sus duraciones a lo largo de una pieza musical base en tiempo discreto. Cabe destacar que las simulaciones generadas parecerían respetar la armonía de la obra en la que están basadas; así como otras de sus cualidades, como lo son el signo de compás, escalas, progresiones y usos melódicos, por mencionar algunos.

## INTRODUCCIÓN Y OBJETIVOS

Dos facultades humanas son la sensibilidad y la razón, las cuales emanan de su cosmovisión, ideología y, por lo tanto, su filosofía. Ahora bien, el objeto de la Ciencia<sup>1</sup> es la producción de conocimientos lógicos; mientras que el Arte<sup>2</sup> genera conocimientos sensitivos. Nótese así que el Arte subyace en la primera de las dos facultades mencionadas y la Ciencia hace función por medio de la segunda.

Pese a que socialmente se ha separado de manera fundamental a estas dos disciplinas, su concepción ontológica muestra que se encuentran íntimamente relacionadas. Evidencia histórica clara se encuentra en el Renacimiento, con el florecimiento post-medieval de las Ciencias y las Artes a la par. *Existe un paralelo histórico entre la Música Europea y los intentos sucesivos de explicar el mundo mediante la razón*<sup>3</sup> [35, p. 1].

Galileo Galilei<sup>4</sup>, Albert Einstein<sup>5</sup>, James Stewart<sup>6</sup> y Manjul Bhargava<sup>7</sup>, entre otros, compararon su pasión tanto hacia las Matemáticas, como hacia la Música. Además, cabe destacar que

---

<sup>1</sup>Del latín *scientia* (conocimiento). Se define como el conjunto de conocimientos obtenidos mediante observación y razonamiento, sistemáticamente estructurados de los que se deducen principios o leyes generales con capacidad predictiva y comprobables experimentalmente [28].

<sup>2</sup>Del latín *ars, artis* y del griego *τεχνή* (técnica). Puede definirse como la manifestación de la actividad humana mediante la cual se interpreta lo real o se plasma lo imaginado con recursos plásticos, lingüísticos o sonoros [28].

<sup>3</sup>“There exists a historical parallel between European music and the successive attempts to explain the world by reason”; [35, p. 1].

<sup>4</sup>Intérprete del laúd y el órgano [3].

<sup>5</sup>Pianista y violinista [10].

<sup>6</sup>Matemático y violinista [7].

<sup>7</sup>Ganador de la medalla Fields en 2014 e intérprete de tabla [15].

personajes como Pitágoras<sup>8</sup>, René Descartes<sup>9</sup>, Leonhard Euler<sup>10</sup> y Gottfried Leibniz<sup>11</sup> mostraron un gran interés por la Música e hicieron experimentaciones y aportaciones a la Teoría Musical desde el enfoque matemático. Adicionalmente, ha existido inspiración en el sentido inverso, como en Johann Sebastian Bach<sup>12</sup>, Béla Bartók, Claude Debussy<sup>13</sup> e incluso Brian May<sup>14</sup>, por mencionar algunos.

En todos estos personajes, aunque separados en tiempo y espacio, el Arte y la Ciencia convergieron armónicamente; en particular, la Música y las Matemáticas, entre las cuales subyace una inherente relación que ha sido y sigue siendo motivo de curiosidad tanto artística, como científica. Manifestación de tal hecho es encontrado en el compositor e ingeniero Iannis Xenakis, quien dedicó su vida a un concepto fuera de lo común: la *Música Estocástica*. Xenakis desarrolló un método contemporáneo de composición musical basado en la Teoría de Probabilidad e inspirado por la idea de que la Música cuenta con cierto componente aleatorio. En sus experimentaciones es posible encontrar motivación y justificación:

*“Desde la antigüedad, conceptos como azar (tyche), desorden (ataxia) y desorganización fueron considerados opuestos y negación a la razón (logos), orden (taxis) y organización. Sólo ahora el conocimiento ha podido penetrar el azar y descubierto cómo separar sus grados. [...] Por lo tanto, no es sorprendente que la presencia o ausencia de causalidad, primero en la filosofía y luego en las ciencias, pueda influenciar la composición musical.”*<sup>15</sup>

– I. Xenakis, 1991

Hacia la indagación sobre tal vínculo, se fijó como objetivo específico la investigación de la estructura de obras musicales académicas desde un enfoque matemático. A través de la Probabilidad

---

<sup>8</sup>Ideó la escala musical pitagórica, antecesora de la escala convencional de la escuela europea de Música [33].

<sup>9</sup>Escribió *Compendio de Música* [8] a los 22 años, su primer libro en latín.

<sup>10</sup>Autor de *Tentamen novae theoriae musicae (Una nueva teoría musical)* [4].

<sup>11</sup>“Leibniz está convencido de que la Música posee una irrefutable estructura matemática. La armonía matemática del universo se revela por ello de modo sensible e inmediato a la percepción aún antes que a la razón. Leibniz expresó la exigencia de una reconciliación entre oído y razón, entre sensibilidad e intelecto, entre arte y ciencia”; [5, p. 3].

<sup>12</sup>*Padre de la Música*, hizo alusión a conceptos matemáticos en sus piezas, como fractales y toros [20].

<sup>13</sup>Como posteriormente haría Bartók, incorporó la sección áurea en la construcción de sus composiciones [18].

<sup>14</sup>Reconciliando ambas disciplinas al ser guitarrista del grupo musical británico *Queen*, compositor, productor musical y posgraduado en Astrofísica por Imperial College London [27].

<sup>15</sup>“Since antiquity the concepts of chance (tyche), disorder (ataxia), and disorganization were considered as the opposite and negation of reason (logos), order (taxis), and organization (systasis). It is only recently that knowledge has been able to penetrate chance and has discovered how to separate its degrees [...] It is therefore not surprising that the presence of absence of the principles of causality, first in philosophy and then in the sciences, might influence musical composition”; [35, p. 4].

y la Estadística, se buscó modelar la intención del autor<sup>16</sup> en cada pieza mediante su armonía. Misma que se pretende cuantificar y a la que, de acuerdo con la Teoría Musical, tienden las expresiones sucesivas del autor a lo largo de cada composición. Dado este argumento, toma sentido hablar de la ergodicidad de los procesos estocásticos asociados a la armonía de una obra musical, por lo que los modelos Markovianos jugaron un papel fundamental. Contando con dicha conceptualización, se establecieron como objetivos secundarios la clasificación de obras por las corrientes históricas marcadas por la Historia del Arte y por su autor, e inclusive, simulación musical.

El gran supuesto detrás del trabajo es que una pieza musical es compuesta con cierto nivel de aleatoriedad en su estructura. Con lo cual, no se pretende ignorar el hecho de que una obra artística sea una manifestación humana cargada de emociones a la pauta de las reglas de la estética. Por lo contrario, se pretende experimentar con la posible variabilidad intrínseca de una pieza al ser y sentir del autor al momento de su gestación. Todo esto, confiando en que la sofisticación de la Matemática es tal que es posible modelar dicho fenómeno tan abstracto.

El Capítulo 1 es dedicado exclusivamente a la Música, se introducen importantes conceptos y una breve noción de su historia. Se clarifican desde los principios básicos y notación, hasta algunos utilizados por la musicología contemporánea, con el fin de establecer representaciones musicales y caracterizaciones matemáticas congruentes para ambas disciplinas. Adicionalmente, en el primer capítulo se propone una vía para la formalización matemática de la Teoría Musical de Conjuntos, así como herramientas musicológicas matemáticas originales. Posteriormente, en los Capítulos 2, 3 y 4 se presenta el marco teórico en el que subyace la investigación, respectivamente: Cadenas de Markov, Campos Aleatorios y Aprendizaje Estadístico. Tras un primer enfoque mediante Cadenas de Markov, la investigación redirige sus miras hacia la generalización en Campos de Gibbs y Modelos de Markov Dinámicos de orden  $k - 1$ , modelaciones propuestas en el Capítulo 5; mismo en que se expone el análisis exploratorio de resultados y su interpretación como aportación al campo de la Musicología. En los Capítulos 6 y 7 se presenta la clasificación de obras musicales por corriente histórica y autor llevada a cabo y simulación musical, respectivamente.

Los documentos referenciados en el presente están disponibles en el repositorio en línea: [github.com/danielapdv/music-mcs](https://github.com/danielapdv/music-mcs).

---

<sup>16</sup>En general, este término es polémico en la Teoría del Arte. Cuando se utilice a lo largo del documento, se referirá a las sensaciones que el autor transmite con su Música de acuerdo con el significado que da la Teoría Musical a los recursos que utilice en una pieza.



# CAPÍTULO 1

## EN BÚSQUEDA DE LA REPRESENTACIÓN MATEMÁTICA DE UNA PIEZA MUSICAL

La Historia y Teoría Musicales son sumamente extensas; sin embargo, se pretende llevar a cabo una síntesis de los aspectos musicales a tomar en cuenta para su comprensión a un nivel básico. Con ello, se facilitará el entendimiento de los modelos propuestos en la presente investigación. Adicionalmente, las fuentes de consulta principales para este primer capítulo son: [2] para Historia de la Música<sup>1</sup>, [29] para las secciones 1.2 a 1.4, [30] y [14] en la sección 1.5 y [11], para las secciones restantes.

Música, del griego *mousiké* (arte de las musas), se puede definir como las expresiones sonoras que plasman sentimientos, emociones o ideas de manera ordenada. También se define como una sucesión de sonidos modulados para recrear el oído; el arte de combinar los sonidos de la voz humana o de los instrumentos, o de unos y otros a la vez, de suerte que produzcan deleite, conmoviendo la sensibilidad [28].

La *Música académica*, también denominada *Música culta*, corresponde al término más comúnmente, aunque erróneamente utilizado, de *Música clásica*. “Con el vocablo ‘clásica’, se abarca un periodo que comprende desde la Música de la época medieval hasta nuestros días. Por tanto, bajo esa denominación se entremezclan Músicas tan diferentes como la de la época barroca y la serial, por citar dos ejemplos” [9, p. 4]. La Música académica simplemente refiere a la Música occidental de

---

<sup>1</sup>La exposición de Historia de la Música será sumamente breve, por lo que es debido hacer referencia a bibliografía clásica de Historia de la Música, como lo es [13].

tradición escrita, razón por la cual se consideran en la investigación solo este tipo de obras, pues la modelación estará basada en la Teoría Musical construida por la escuela occidental.

## 1.1. Historia de la Música

“Desde sus inicios, la Música ha jugado un papel esencial en la interacción social, ya que suele acompañar ceremonias religiosas y de muchas otras índoles” [9, p. 3]. Pese a que su origen se mantiene en el campo de la especulación, se sostiene que posiblemente se haya desarrollado entre madre e hijo para reforzar su vínculo, convirtiéndose después en una forma de comunicación entre adultos. “Parece haber sido la más antigua de las artes, ya que el hombre y la mujer la conocían y la amaban aún antes de ser consciente y de su poder creador” [9, p. 3].

Ahora bien, se distinguen seis etapas históricas en la Música occidental europea: Edad Media, Renacimiento, Barroco, Clasicismo, Romanticismo, Modernismo y Contemporaneidad. La Música en la Edad Media (450-1450) fue altamente dependiente del cristianismo, de manera que la Música Eclesiástica constituyó su manifestación en este periodo, junto con interpretaciones *polifónicas* de textos sagrados. Se hicieron experimentaciones en la notación del *timbre* y *ritmo*, comenzaron los intentos de describir la *polifonía* y se dieron los primeros registros de Música compuesta de dos o más *melodías independientes*, así como la gestación de las primeras teorías de *consonancia* y *dissonancia*. Algunos exponentes de la época son Leonin, Perotin, Guillaume de Machaut, Francesco Landini y Guillaume Dufay.

En cuanto a la Música del Renacimiento (1450-1600), y a diferencia del resto de las artes de la época, el estilo griego no tuvo gran influencia. Más bien, evolucionaron los conceptos de disonancia y consonancia del periodo anterior, con énfasis en *sonoridades armónicas*. *La textura predominante consistió en voces soprano, alto, tenor y bajo, creando una red altamente contrapuntística en la que las líneas divergen, convergen, se cruzan, hacen eco y se imitan unas a otras; a veces con gran independencia rítmica, a veces moviéndose juntas a manera de un himno*<sup>2</sup> [2, p. 81]. Las figuras eclesiásticas y aristocráticas utilizaron el arte, incluyendo la Música, como símbolo de poder

---

<sup>2</sup>“The predominant texture consisted of soprano, alto, tenor, and bass voice parts creating a highly contrapuntal web in which the lines diverge, converge, cross, echo, and imitate each other, sometimes with great rhythmic independence, sometimes moving together in the manner of a hymn”; [2, p. 81].

y riqueza. Cabe destacar que en este periodo se publicó la primera colección impresa de Música polifónica, se fundaron los primeros conservatorios en Nápoles y Venecia y comenzó el desarrollo del violín. Finalmente, entre los compositores musicales renacentistas se encuentran Palestrina, Tomás Luis de Victoria, Johannes Ockeghem y Josquin des Prez.

El Barroco comprendió el periodo entre los años 1600 y 1750, aproximadamente. Así como en el Renacimiento, la Música era usada como símbolo de opulencia. Sin embargo, en esta época se popularizaron las composiciones; cuyas expresiones fueron caracterizadas por intensidad dramática, emoción, dinamismo y efectos de contraste. En el periodo Barroco se consolidó la *escala temperada*, se publicó la Enciclopedia de la Música<sup>3</sup>, se inauguró la primera casa de ópera pública y se inventó el pianoforte, entre otros hallazgos. Johann Sebastian Bach, Antonio Vivaldi, George Frederick Handel y Jean-Philippe Rameau son algunos de los exponentes del Barroco.

Durante el Siglo de las Luces, se dio lugar al periodo Clásico (1750-1820), en el cual la expresión se basó en los ideales y la noción de belleza. A diferencia del Barroco, el estilo musical tomó matices de restricción emocional, equilibrio y simetría. En dicho periodo se expandió el tamaño de orquesta, se modificaron los mecanismos de los instrumentos como el piano y los conciertos y publicaciones musicales llegaron a todo público. Los compositores clásicos emblemáticos son Ludwig Van Beethoven, Wolfgang Amadeus Mozart y Franz Josef Haydn.

Entre 1820 y 1900 se desarrolló el Romanticismo. Fue caracterizado por la emoción, espontaneidad e individualidad; inspiradas en la naturaleza, el pasado, sueños, opresión, injusticia, muerte y tragedia. En esta época se fundó la Royal Academy of Music de Londres, se mejoraron los mecanismos del piano, se patentó el saxofón y se hicieron las primeras grabaciones magnéticas de sonido. Entre otros, destacaron en el Romanticismo Ludwig van Beethoven (vivió también el Clasicismo), Franz Schubert, Frederic Chopin y Peter Illich Tchaikovsky.

Finalmente, hacia el Siglo XX se encuentran las épocas Moderna y Contemporánea; ambas caracterizadas por su pluralidad y, en general, respuesta en contra de la estética y valores del Romanticismo. En la Música ocurrieron acontecimientos como la creación de la *Música atonal* (Arnold Schoenberg) o la incorporación de ritmos y tonalidad de Música popular como el blues o el jazz en la Música académica (George Gershwin y Aaron Copland). Durante este periodo se dio la primera grabación de ópera, la primera transmisión de Música por radio, la ingeniería de los instrumentos

---

<sup>3</sup>Del compositor alemán Michael Praetorius, 1620.

musicales electrónicos y las grabaciones estereofónicas. Arnold Shoenberg, Bela Bartók y George Gershwin son compositores musicales representativos de dicha era, que se extiende hasta la fecha.

## 1.2. Principios musicales

La escritura musical se lleva a cabo en un arreglo de cinco líneas horizontales paralelas llamado *pentagrama*, el cual es subdividido por líneas verticales en partes denominadas *compases*. Una *nota* es cada uno de los sonidos en cuanto está producido por una vibración de frecuencia constante [28]. En la notación musical, se representan con símbolos igualmente denominados *notas*, las cuales son escritas sobre las líneas o los espacios del pentagrama. Su posición es definida por la *clave*, el símbolo que se encuentra al inicio del pentagrama; el cual es seguido por la *armadura* y el *signo de compás* o *medida* de la pieza<sup>4</sup>.



Figura 1.1: Ejemplo Pentagrama 1.



Figura 1.2: Ejemplo Pentagrama 2.

La *altura* de una nota corresponde a qué tan aguda o grave es; ésta es determinada por la frecuencia de su onda de sonido, una cualidad acústica. Por ejemplo, la nota  $La_0$  –notación que tendrá sentido más adelante– es el sonido cuya frecuencia de onda corresponde a 27.5 Hertz (ciclos por segundo) y, de hecho, la nota más grave de la escala temperada.

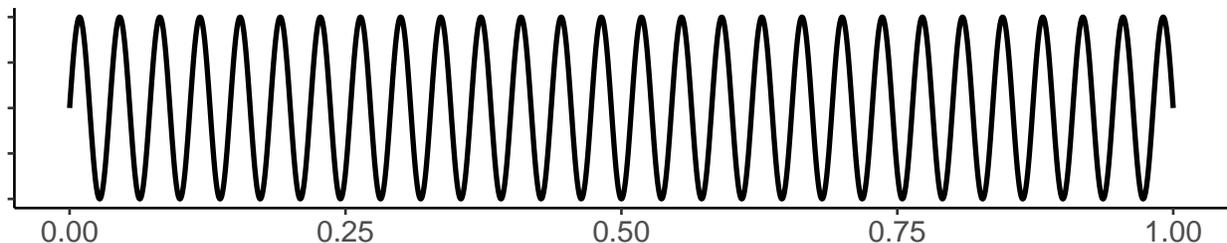


Figura 1.3: Onda de sonido de la nota  $La_0$ . El eje  $x$  corresponde al tiempo (segundos), mientras que el eje  $y$  a la amplitud de la onda (decibeles).

<sup>4</sup>Estos conceptos se definirán posteriormente.

Una nota con frecuencia al doble que otra se dice que se encuentra una *octava* arriba de ella. Ahora bien, una octava se separa en doce *semitonos* (la distancia más corta entre dos notas en la escala diatónica, la distancia entre dos teclas de piano) y ocho *tonos* (dos semitonos). Adicionalmente, aquellas notas separadas por una o más octavas reciben el mismo nombre del semitono al que corresponden, seguido por un número que denota la octava a la que pertenecen –por ejemplo,  $La_0$ –. Lo anterior implica que el sistema musical occidental tradicional da cabida a *equivalencia entre octavas*.

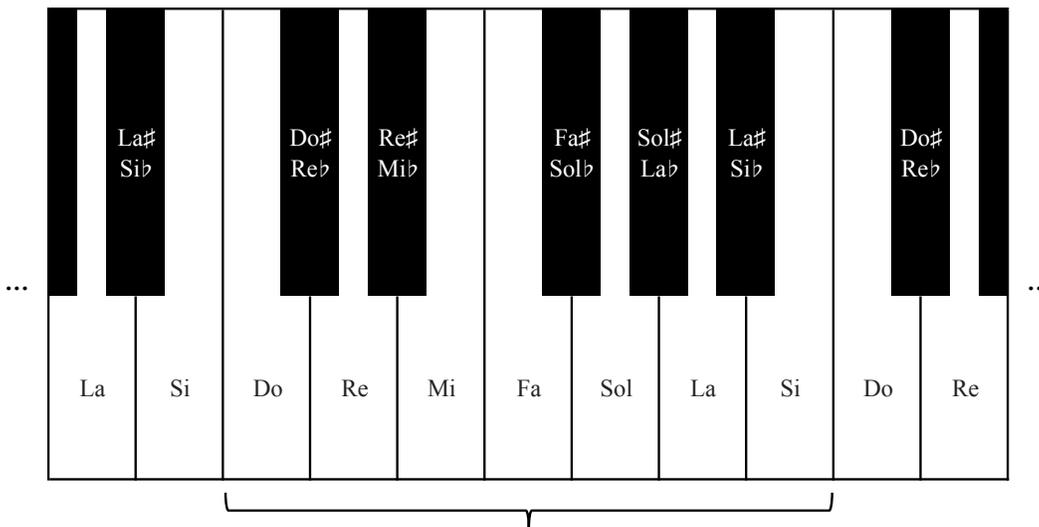


Figura 1.4: Octava en el piano con nombres de las notas.

Dentro de una octava se distinguen siete notas con diferentes alturas que se denominan *notas naturales*: *Do, Re, Mi, Fa, Sol, La* y  $Si^5$ . Cuando se anteponen los símbolos  $\flat$  ó  $\sharp$  se está refiriendo respectivamente a la nota que se encuentra un semitono abajo (*bemol*) o un semitono arriba (*sostenida*) a la nota natural indicada. También cabe destacar que existe equivalencia entre una nota sostenida y la nota que le sigue en calidad bemol, por ejemplo, son equivalentes  $Do\sharp$  y  $Re\flat$ .



Figura 1.5: Notas naturales de la cuarta octava. El símbolo (C) que sigue de la clave denota compases de 4/4.

<sup>5</sup> C, D, E, F, G, A y B en la notación inglesa.

La armadura es la lista de sostenidos y bemoles en la *clave*, las notas no especificadas en ella se llaman *alteraciones* o *accidentes*. Adicionalmente, en caso de que se especifique una nota en la armadura, se antepone el símbolo ♮ para referirse a la nota natural asociada.

Por otro lado, una característica fundamental de una nota es su *duración*, cuya notación es ilustrada en la Figura 1.6, junto con sus nombres. La duración *efectiva* de una nota en una pieza está dada por su compás y su *tempo*. Con el primer concepto se indica cuántos pulsos o tiempos hay en un compás (número superior) y el tipo de nota que dura un tiempo (número inferior); mientras que el siguiente refiere a la velocidad con que debe ser tocada y es normalmente especificado al inicio de la pieza.

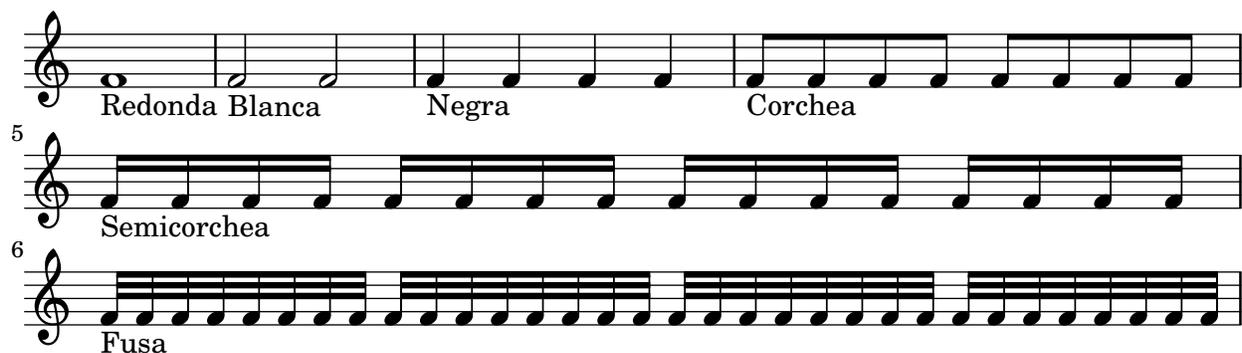


Figura 1.6: Duraciones canónicas de las notas.

Es decir, las duraciones de las notas tienen las siguientes equivalencias en términos de una nota *redonda*:

- Media duración de una nota *breve*<sup>6</sup>,
- Dos notas *blancas*,
- Cuatro notas *negras*,
- Ocho *corcheas*,
- 16 *semicorcheas*,

<sup>6</sup>La cual se omite de la Figura 1.6 puesto que, dada su equivalencia justo antes mencionada, debería doblarse la cantidad de notas ilustradas.

- 32 *fusas*, y
- 64 *semifusas*.

Un elemento esencial de la Música es el *silencio*, por lo que también tiene cabida en la notación musical con sus diferentes posibles duraciones (análogas a las de las notas).

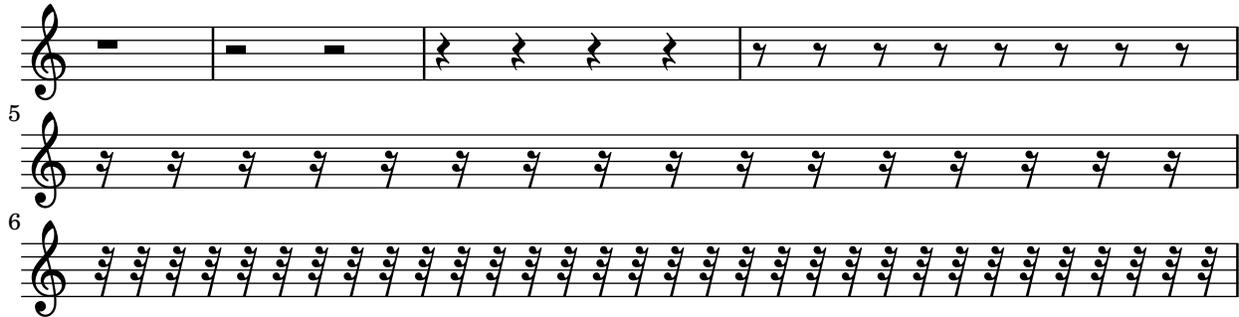


Figura 1.7: Duraciones canónicas de los silencios.

### 1.3. Elementos de la Música

Otra definición para la Música es *melodía, ritmo y armonía combinados* [28], a causa de que son las fracciones que la componen. No obstante, el *timbre* y la *textura* de la Música también son partes fundamentales de ella.

En primer lugar, la Música se concreta de manera física a través del tiempo. Por lo que se define primeramente el ritmo como *el posicionamiento de los sonidos de una pieza musical en el tiempo*<sup>7</sup> [29, p. 71].

Ahora bien, la melodía corresponde a la composición en que se desarrolla una idea musical y se realiza a través de una sucesión de notas denominada *línea melódica*. Cabe destacar que no

<sup>7</sup>“The placement of the sounds in time is the rhythm of a piece of music.”; [29, p. 71].

todas las notas en una melodía han de pertenecer a la línea melódica principal de la pieza. Ciertas notas pueden ser posicionadas con el fin de hacerla *compleja*<sup>8</sup>, por lo que toman el nombre de *notas ornamentales*. En general, las notas de este tipo son ajenas a los *acordes* en que aparecen<sup>9</sup>, de manera que no influyen en su estructura. Adicionalmente, se sitúan alrededor de las notas en la línea melódica principal y con ellas es posible generar énfasis o variedad en el ritmo.

El tercer elemento que compone a la Música es la *armonía*, de hecho, también el más desarrollado y estudiado en la Música Occidental. Tal concepto será el cimiento de los modelos propuestos en la presente investigación, por lo que la siguiente sección es dedicada de manera exclusiva a ello.

También denominado *color*, el timbre es la parte de la Música fuera de sus aspectos tonales del sonido o duración. A pesar de que una nota corresponde a un sonido causado por una vibración de frecuencia específica, como se expuso en la Sección 1.2, dos notas iguales tocadas en instrumentos diferentes suenan distinto; esto es debido a la diferencia entre sus timbres, por las frecuencias añadidas por los instrumentos al tocar una nota a causa de su mecanismo o configuración específica.

Finalmente, *cuando se describe la textura de una pieza musical, se está describiendo cuánto está sucediendo en la Música en cualquier momento*<sup>10</sup> [29, p. 80]. Los cuatro términos formales utilizados para categorizar la textura en la Música, a través de sus melodías y armonías, están dados a continuación:

- Polifónica o Contrapuntística: cuando una pieza musical cuenta con dos o más líneas melódicas simultáneas e independientes (*contrapunto*).
- Monofónica: Música compuesta por una sola línea melódica, sin armonía o *contrapunto*.
- Homofónica: textura asociada a piezas musicales que contienen una clara línea melódica y *acompañamiento*. Puede tener otras líneas melódicas, pero no independientes puesto que se suelen utilizar para *llenar* los acordes del acompañamiento.
- Heterofónica: Música caracterizada por una sola línea melódica que ocurre de manera simultánea a ejecución de variaciones suyas, sin ser líneas melódicas independientes.

---

<sup>8</sup>La complejidad musical refiere a la variedad, cambio y ocurrencia de nuevos estímulos en una pieza.

<sup>9</sup>No los generan, no son parte de ellos. Véase la siguiente sección.

<sup>10</sup>“When you describe the texture of a piece of music, you are describing how much is going on in the music at any given moment”; [29, p. 80].

## 1.4. Teoría de la armonía

J. Laborecky, en su diccionario terminológico musical define a la *armonía* en la Música como sigue:

*“La armonía musical es el estudio sobre el carácter de tonos que suenan simultáneamente, su significado, sus transiciones, la función de sus relaciones y su uso en la pieza musical. Estudia relaciones horizontales (subsecuentes) en el tiempo y relaciones verticales (concurrentes) dentro del espacio tonal<sup>11</sup>.”*

– J. Laborecky, 1984

La armonía de una pieza musical entonces refiere al uso simultáneo de notas y *acordes*, junto con su carácter, significado, transiciones y relaciones dentro de la obra. A partir de lo cual, destaca que la *Teoría Musical* de la escuela occidental ha basado su estudio en los tonos y sus relaciones conjuntas, de manera que es un<sup>12</sup> sistema meramente de *intervalos*, la distancia de un tono a otro. Nótese además que la armonía es un elemento sustancialmente diferente a la melodía; de hecho, coloquialmente podría decirse que se manifiesta en el *acompañamiento*<sup>13</sup> de la pieza.

En camino hacia un entendimiento de aquéllo que representa la armonía, deben esclarecerse ciertos conceptos, empezando por el de *intervalo*. En concordancia a lo mencionado anteriormente, pero de manera puntual, se define el intervalo entre dos notas como la razón de sus frecuencias, la cual es detectable a través de su sonido, o bien, a través de su notación. Los intervalos básicos, ejemplificados en la Figura 1.8 y ordenados por el número de semitonos de diferencia a los que refieren, son: unísono, segunda menor (*m2*), segunda mayor (*M2*), tercera menor (*m3*), tercera mayor (*M3*), cuarta justa (*P4*), tritonos (*A4* y *d5*), quinta justa (*P5*), sexta menor (*m6*), sexta mayor (*M6*), séptima menor (*m7*), séptima mayor (*M7*) y octava perfecta.

---

<sup>11</sup>“Music harmony is the study about the character of simultaneously sounding pitches, their meaning, transitions, functional relationships and usage in the musical piece. It studies horizontal (subsequent) relationships in the time and vertical (concurrent) relationships among the tone space”; [23].

<sup>12</sup>Existen otras conceptualizaciones de la Música, todas aquéllas que no se hayan gestado en Europa desde el Medioevo. Recuérdese que la Música ha pertenecido a diferentes etnias a lo largo de la historia y el mundo.

<sup>13</sup>Definido como todo elemento de la Música que no es parte de la melodía.

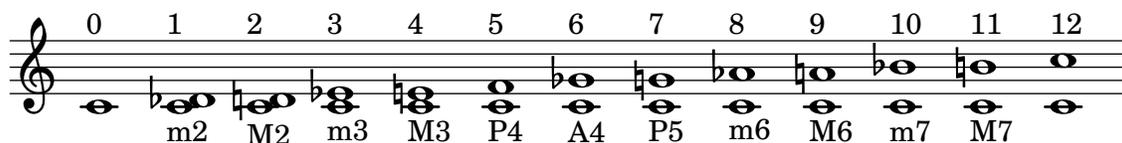


Figura 1.8: Intervalos básicos entre dos notas y su notación.

Un acorde es un conjunto de al menos tres notas, las cuales tienen intervalos entre ellas lo suficientemente amplios para poder sonar simultáneamente sin dar la sensación de densidad excesiva<sup>14</sup> [25, p. 20]. La *nota fundamental* o *raíz* del acorde es aquella mediante la cual el acorde puede ser construido apilando terceras, en la armonía tradicional occidental; además, esta configuración del acorde es llamada *posición básica* o *estado fundamental*. Pueden obtenerse otras posiciones del acorde reorganizando las notas que lo conforman: si la tercera del acorde se encuentra en la posición más baja, ésta se denomina primera inversión; mientras que si la quinta está en la posición más baja, se dice que el acorde está en su segunda inversión. Cabe hacer notar que al invertir un acorde se sigue haciendo referencia al mismo, puesto que sigue estando integrado por las mismas notas. En la Figura 1.9 se ejemplifica tal hecho con el acorde *Do Mayor*.

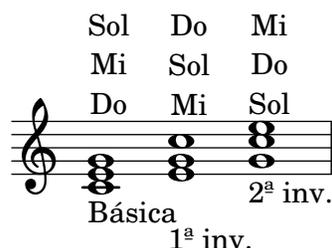


Figura 1.9: Ejemplo de inversiones del acorde *Do Mayor* (C).

Adicionalmente, cabe hacer notar que así como las escalas, los acordes toman su nombre, primero, por su nota fundamental y, luego, por el tipo y cantidad de intervalos entre sus notas en su posición básica. Los tipos de acorde más comunes son las *triadas*<sup>15</sup> y las *séptimas*<sup>16</sup> básicas, los cuales son descritos a continuación junto con su notación:

- Mayores (*Maj*): Triadas mayores, es decir, aquellas conformadas por la nota fundamental, una tercera mayor y una quinta justa.
- Menores (*min* ó *m*): Triadas menores, integradas por la fundamental, una tercera menor y una quinta justa.

<sup>14</sup>“Chord is a set of tones with the minimum of three tones, having the intervals in between them big enough, so the may sound together without the feel of excessive density”; [25, p. 20].

<sup>15</sup>Acorde en posición básica e integrado por tres notas: la fundamental, una tercera y una quinta.

<sup>16</sup>Acorde compuesto por una triada con una nota séptima agregada.

- Aumentados ( $^+$  ó *aug*): Triadas compuestas por la raíz y dos terceras mayores.
- Disminuidos ( $^-$  ó *dim*): Triadas integradas por dos terceras menores, además de la raíz del acorde.
- Séptima dominante (7): Una séptima compuesta por una triada mayor y una nota séptima menor.
- Séptima mayor (*Maj7* ó *M7*): Una séptima compuesta por una triada mayor y una nota séptima mayor.
- Séptima menor (*min7* ó *m7*): Una séptima compuesta por una triada menor y una nota séptima menor.
- Séptima aumentada ( $7\#5$  ó *aug7*): Una séptima compuesta por una triada aumentada y una nota séptima menor.
- Séptima disminuida ( $^o$  ó *dim7*): Una séptima compuesta por una triada disminuida y una nota séptima disminuida<sup>17</sup>.
- Séptima semi-disminuida ( $C^0$ ): Una séptima compuesta por una triada disminuida y una nota séptima menor.

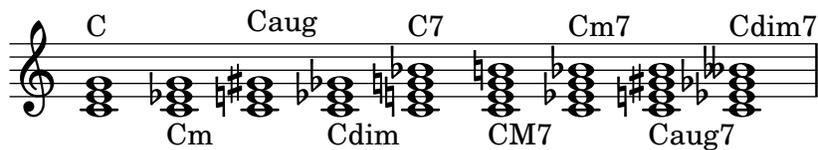


Figura 1.10: Triadas y séptimas comunes de *Do* (C).

Ahora bien, los mencionados no son los únicos acordes que existen. En general, puede *extenderse* el acorde agregando terceras o cualesquiera tonos en la escala en la que se encuentre la pieza. Asimismo, en caso de que una nota del acorde no se encuentre en la escala menor o mayor de su nota fundamental, a dicha nota se le denomina *nota alterada* y al acorde, *acorde alterado*.

Por otro lado, una *escala* es una serie de notas creciente (ordenadas de grave a agudo) o decreciente (ordenadas de agudo a grave). Las escalas *diatónicas* están integradas por ocho notas con intervalos de un semitono y un tono. Adicionalmente, se dividen en dos categorías:

<sup>17</sup>Un semitono por debajo de una séptima menor.

- Escalas mayores: Son caracterizadas por la presencia de terceras mayores. La más común es la escala de *Do mayor*, integrada por las notas naturales y mencionada en la Sección 1.2.
- Escalas menores: Son análogas a las escalas mayores, pero con la presencia principal de terceras menores. Asimismo, la escala menor más común es *La menor* y la componen las notas naturales.

El nombre de una escala es determinado por su distribución de tonos y semitonos, así como por su primer tono. Dicha nota corresponde a su *centro tonal* y es denominada *tónica*. Bajo el mismo razonamiento, cada nota de una escala se indexa respecto a su intervalo con la tónica. Tal índice es denominado el *grado* de la escala y se denota con *numerales romanos* del uno al siete, donde se hace distinción entre tonos mayores y menores (respecto la tónica) escribiéndolos con mayúsculas o minúsculas respectivamente<sup>18</sup>.

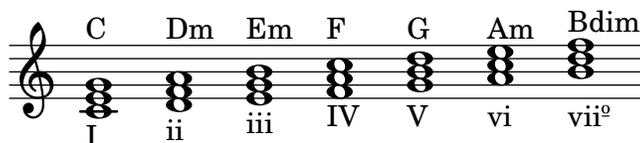


Figura 1.11: Acordes y grados de la tonalidad *Do mayor*.

Teniendo las nociones de acorde y escala, puede definirse el concepto elemental de la *armonía tonal*<sup>19</sup>: la *tonalidad* o el *tono* de una pieza musical. Basado en una escala menor o mayor, el tono de una pieza musical es un sistema de las relaciones dentro de su armonía. En realidad, cualquier acorde podría formar parte de la tonalidad de una pieza.

Sin embargo, los que aparecerán más frecuentemente son aquéllos cuyas notas fundamentales formen parte de la escala correspondiente, aquéllos a los que se aplica la numeración romana. Así, dicha numeración refiere a la *función armónica* de las notas en una tonalidad, el grado de tensión que produce respecto la tónica. Las tres principales funciones armónicas en la Música son:

- La tónica, al ser el centro tonal de la escala (grado I), da la sensación de reposo, resolución. Las armonías por lo general comienzan en la tónica.

<sup>18</sup>Véase Figura 1.11.

<sup>19</sup>Un sistema concreto de la armonía musical en el que cada parte de la pieza debe pertenecer a una tonalidad, mayor o menor, y cada armonía cuenta con una relación, ya sea cercana o distante, al centro tonal, de acuerdo con Marsík [25, p. 11].

- La triada correspondiente al grado IV del tono de la pieza es denominada *subdominante*. Normalmente, el movimiento de la tónica en la armonía se lleva cabo a través de ella.
- La triada *dominante* es aquella correspondiente al grado V y se asocia a la máxima tensión respecto la tónica, por lo que se dice que se requiere una transición a ella, un *descanso*.

De esta manera, I-IV-V-I corresponde a la *progresión armónica*<sup>20</sup> básica. El resto de los grados de la escala<sup>21</sup> se consideran variaciones de las tres funciones armónicas mencionadas, por lo que pueden sustituirse en la progresión dado su grado en la escala.

## 1.5. Consonancia y disonancia

Como se habrá dado a notar, algunos elementos de la Teoría de la Armonía están dados en función de las *sensaciones* que produce su sonoridad. Tales nociones tienen base en conceptos profundos que, de hecho, tendrán un papel importante en la modelación llevada a cabo, puesto que se pretende que sea congruente con las bases musicales. De manera que esta sección es dedicada a la exposición de dos elementos fundamentales en la Teoría de la Armonía y, por lo tanto, del presente documento: *consonancia* y *disonancia*.

En general, la consonancia se define como la *cualidad de aquellos sonidos que, oídos simultáneamente, producen efecto agradable*<sup>22</sup>. Ahora bien, es natural pensar en la disonancia como el concepto dual de la consonancia; en general, se define como *un acorde no consonante* [28], *una nota, acorde o intervalo que no encaja en las armonías triádicas que hemos aprendido a esperar de la Música*<sup>23</sup> [29, p. 84].

No obstante, dada la carga perceptual que subyace en la consonancia, históricamente se ha conceptualizado de maneras distintas. De acuerdo con William A. Sethares en [30, p. 77], existen al menos cinco:

---

<sup>20</sup>Sucesión de acordes.

<sup>21</sup>*Supertónica* (ii), *Mediante* (iii), *Submediante* (vi) y *Subtónica* (vii°).

<sup>22</sup>Definición tomada de [28], aunque de hecho, es concordante con la bibliografía sobre Teoría Musical consultada.

<sup>23</sup>“A dissonance is a note, chord or interval that does not fit into the triadic harmonies that we have learned to expect from music”; [29, p. 84].

- **Consonancia Melódica:** Es una noción meramente melódica, en el sentido de que esta cualidad es asociada a los intervalos melódicos en su contexto musical. Data del medievo, manifestándose en los cánticos eclesiásticos, siendo la primera concepción de consonancia y disonancia.
- **Consonancia Polifónica:** Al surgimiento de la Música Polifónica hacia el Renacimiento, el concepto de consonancia naturalmente se vinculó a los intervalos de dos sonidos simultáneos, dejando atrás el principio melódico del punto anterior. Asimismo, con la Consonancia Polifónica se establece por primera vez la asociación clara de consonancia a lo “agradable”. Entre los teóricos musicales de la época, se abrieron dos vertientes: algunos consideraban que la consonancia de un intervalo sería proporcional al grado en que los dos tonos pueden sonar como uno solo; mientras que otros sostenían que el oído *tiende a preferir* intervalos simples, por lo que se consideraban consonantes el unísono, terceras, cuartas, quintas, sextas y octavas.
- **Consonancia Contrapuntística:** Bajo este enfoque, el factor determinante de la consonancia de un intervalo es el contexto en el que ocurre, dado que surgió junto con el contrapunto. Se consideran<sup>24</sup> consonantes los mismos tipos de intervalo que en la Consonancia Polifónica, con la diferencia de que las segundas aumentadas se dicen disonantes y la tercera menor consonante.
- **Consonancia Funcional:** Posteriormente, el pensamiento se tornó hacia la relación entre los tonos individuales y la tónica de la escala de la pieza, considerando disonantes aquellos tonos alejados de la misma.
- **Consonancia Psicoacústica:** También denominada consonancia sensorial, es la conceptualización más reciente y está enfocada en la percepción de consonancia a través de los mecanismos del sistema auditivo humano. De esta manera, las sensaciones que definen si sonidos simultáneos son consonantes o no, son meramente físicas.

Dado que la disonancia es relacionada a la tensión (e incluso al desagrado), lo común es evolucionar hacia el reposo, *resolver* en consonancia de nuevo<sup>25</sup>, en cadencia. *La cadencia es cualquier parte en una pieza musical que tiene la sensación de un punto de terminación. Éste puede ser un*

---

<sup>24</sup>Puesto que ésta es la noción de consonancia más común hoy en día.

<sup>25</sup>He aquí el por qué de la progresión armónica básica mencionada en la sección anterior.

*fuerte, definitivo punto de detenimiento –al final de la pieza, por ejemplo, o al final de un movimiento o verso– pero también se refiere las pausas “de reposo temporal” que dan fin a las ideas musicales dentro de cada gran sección*<sup>26</sup> [29, p. 200].

Christophe Guillotel-Nothmann en [14, p. 2] muestra que la evolución del tratamiento de las disonancias puede ser considerada como uno de los factores que llevaron al nacimiento de la tonalidad. Más aún, menciona que la disonancia juega un papel crucial en la asimetría en las progresiones de las notas fundamentales, i.e., su función armónica.

Asimismo, Carl Dahlhaus<sup>27</sup> estableció que la teoría de Jean-Philippe Rameau toma como premisa filosófica el principio Aristotélico de la tendencia de la imperfección a la perfección. De manera que, *el origen de la regla de acuerdo con la cual la disonancia, asociada a la imperfección, debe ser resuelta mediante el menor movimiento en una consonancia imperfecta y luego a una consonancia perfecta, asociada con la perfección, está relacionado a este principio Aristotélico*<sup>28</sup> [14, p. 3].

Nótese entonces que, a pesar del momento histórico al que pertenezca una pieza musical, la consonancia o la disonancia que subyazca en su armonía –junto con su uso específico derivado de la concepción de estos elementos y del desarrollo de la Teoría Musical hasta su momento de creación– son determinantes de las sensaciones que transmite, que el autor busca transmitir. *Uno de los aspectos principales en la funcionalidad tonal, a saber, la tendencia direccional de las progresiones de acordes, depende de las decisiones subjetivas artísticas inherentes a la Música Occidental*<sup>29</sup> [14, p. 9].

---

<sup>26</sup>“A cadence is any place in a piece of music that has the feel of an ending point. This can be either a strong, definite stopping point –the end of the piece, for example, or the end of a movement or a verse– but it also refers to the “temporary-resting-place” pauses that round off the ends of musical ideas within each larger section”; [29, p. 200].

<sup>27</sup>Musicólogo alemán contemporáneo.

<sup>28</sup>“The origin of the rule according to which the dissonance, associated with imperfection, must be resolved by the least movement onto an imperfect consonance and then onto a perfect consonance associated with perfection is related to this Aristotelian principle”; [14, p. 3].

<sup>29</sup>“One of the main aspects of tonal functioning, namely the directional tendency of chord progressions, depends on subjective artistic choices inherent in Western music”; [14, p. 9].

## 1.6. Teoría Musical de Conjuntos

El término *Música Postonal* es utilizado para categorizar a la Música Académica que *no sigue las convenciones tradicionales de armonía tonal* [21, p. 1]. Surgió a inicios del Siglo XX tras el Romanticismo, en la época Contemporánea-Moderna; aunque cabe destacar que no toda la Música Académica Contemporánea o Moderna es Postonal. Como es mencionado en la Sección 1.1, dicha etapa se caracterizó por la negación de los *valores y usos estéticos* del periodo anterior; en específico, la Música Postonal es caracterizada por la declinación del sistema tonal desarrollado desde el Barroco. Algunas de las tendencias de esta corriente son:

- Politonalidad: La presencia de dos o más centros tonales simultáneamente en la pieza.
- Pandiatonismo: Uso de la escala diatónica en un contexto no tradicional.
- Atonalidad: Evitación de la determinación de un centro tonal.
- Música Céntrica: El establecimiento del centro tonal de la pieza a través de otros métodos fuera de la progresión armónica básica y sus sustituciones.

Dadas las nuevas maneras de composición musical, los teóricos musicales se vieron en necesidad de evolucionar en cuanto a herramientas para el análisis de la Música ya que, como puede apreciarse en las secciones anteriores, la teoría desarrollada hasta el inicio de la contemporaneidad está basada precisamente en la tonalidad y las progresiones armónicas tradicionales. Razón por la cual, surgió la *Teoría Musical de Conjuntos*, la cual será de gran utilidad para la caracterización matemática de una pieza musical.

En primera instancia, recuérdese que una octava está integrada por doce semitonos, doce *clases de notas*. De acuerdo con Allen Forte en [11, p. 3], un *conjunto tonal* es un conjunto de enteros distintos  $\{0, 1, \dots, 11\}$  que representan clases de notas (la nota Do está asociada al cero, y así ascendentemente). Ahora bien, una *clase de conjuntos* es una colección de conjuntos tonales relacionados por una transformación o grupo de transformaciones. La ventaja de las herramientas musicológicas postonales que se presentarán, es que pueden escribirse matemáticamente. Es momento de establecer que la teoría desarrollada en esta sección corresponde a la publicada por Allen Forte en [11], pero se hace una propuesta propia para su formalización al lenguaje matemático.

**Definición 1.1.** La aplicación  $C : \mathcal{N} \rightarrow \{0, 1, \dots, 11\}$  dada por  $C(n_i) = i; i \in \{0, \dots, 11\}$  se denominará *clase de notas*. Donde

$$\mathcal{N} = \{Do, Do\sharp - Reb, Re, Re\sharp - Mib, Mi, Fa, Fa\sharp - Solb, Sol, Sol\sharp - Lab, La, La\sharp - Sib, Si\}.$$

Es debido hacer notar que la notación utilizada para describir al conjunto  $\mathcal{N}$  de la definición anterior, en cuanto a notas bemoles y sostenidas, hace alusión a la equivalencia entre ellas –la cual fue mencionada en la Sección 1.2–.

**Definición 1.2.** Un *conjunto tonal* es una  $n$ -ada  $(A_1, \dots, A_n) \in \{0, 1, \dots, 11\}^n; n \leq 12$  tal que para cada  $i \in \{1, \dots, n\}, A_i = C(n_j)$  para alguna nota  $n_j \in \mathcal{N}$  (son clases de notas) y  $A_i \neq A_k$ , para  $i \neq k; k \in \{1, \dots, n\}$ .

Nótese que a través de las clases de notas y conjuntos tonales se tiene una manera de representar un acorde matemáticamente. Por ejemplo, el acorde Do Mayor estaría asociado al conjunto tonal  $[0\ 4\ 7]$ , puesto que 0, 4 y 7 son resultado de aplicar la función  $C$  a Do, Mi y Sol.

Se denotará a los conjuntos tonales como se hace convencionalmente  $[A_1 \dots A_n]$ , en lugar de la notación habitual matemática para vectores. Adicionalmente, se denotará por  $\mathcal{CT}$  a la colección de conjuntos tonales, i.e.,

$$\mathcal{CT} = \bigcup_{n=1}^{12} \{0, \dots, 11\}^n.$$

**Definición 1.3.** Una *permutación circular de un conjunto tonal*  $[A_1 \dots A_n] \in \mathcal{CT}$  es una permutación  $P_n : \{0, \dots, 11\}^n \rightarrow \{0, \dots, 11\}^n$  dada por

$$P_n(A_i) = \begin{cases} A_1 & \text{si } i = n, \\ A_{i+1} & \text{si } i \in \{1, \dots, n-1\}. \end{cases}$$

Además, se llamará a la  $i$ -ésima permutación circular de  $A$  al conjunto tonal resultante al aplicar la permutación circular  $i - 1$  veces a  $A$ .

Por ejemplo, la primera permutación circular de  $[0\ 4\ 7]$  es  $[4\ 7\ 0]$ , por lo tanto su segunda per-

mutación es [7 0 4]. Al aplicar a un conjunto tonal una permutación circular, se sigue hablando del mismo conjunto en el sentido de que sigue haciendo referencia al mismo acorde.

**Definición 1.4.** Se dice que un conjunto tonal está en *orden normal* si la diferencia entre su primer y última entrada es la menor posible de todas sus permutaciones circulares. Es decir,  $A \in \mathcal{C}\mathcal{T}$  está en orden normal si

$$A = \operatorname{argmín}_{B \in P_n^A} \mathcal{F}_n(B);$$

donde  $\mathcal{F}_n : \{0, \dots, 11\}^n \rightarrow \{0, \dots, 11\}$ ,

$$\mathcal{F}_n(B) = \begin{cases} B_n - B_1 & \text{si las entradas de B están ordenadas} \\ & \text{ascendentemente,} \\ B_n + 12 - B_1 & \text{en otro caso,} \end{cases}$$

$$\text{y } P_n^A := \{B \in \mathcal{C}\mathcal{T} \mid (P_n)^k(A) = B, \text{ p.a. } k = 0, 1, \dots, n\}.$$

El conjunto atonal asociado al acorde *FaM7* (Fa, La, Do, Mi) está dado por [5 9 0 4]; ordenando sus entradas ascendentemente resulta en el conjunto tonal [0 4 5 9]. Sus cuatro permutaciones circulares están dadas en el Cuadro 1.1, donde se muestra la razón por la que [5 9 0 4] en su orden normal es [4 5 9 0], o bien, [9 0 4 5].

$i$	$(P_4)^i(B)$	$\mathcal{F}_4(B)$
0	[0 4 5 9]	9
1	[4 5 9 0]	8
2	[5 9 0 4]	11
3	[9 0 4 5]	8

Cuadro 1.1: Permutaciones circulares del conjunto tonal [0 4 5 9].

**Definición 1.5.** El conjunto tonal de *mejor orden normal* de otro conjunto tonal está dado por:

1. Su forma normal, si es que ésta es única.

2. En caso de que su forma normal no sea única, aquella forma normal tal que la diferencia entre su  $i$ -ésima y primera entrada es la menor posible, donde  $i$  corresponde a la primera entrada tal que dichas diferencias son distintas entre sus formas normales, i.e.,

$$A = \underset{B \in FN_A}{\operatorname{argmín}} \mathcal{F}'_n(B);$$

donde  $\mathcal{F}'_n : \{0, \dots, 11\}^n \rightarrow \{0, \dots, 11\}$ ,

$$\mathcal{F}'_n(B) = \begin{cases} B_i + 12 - B_1 & \text{si } i \geq \lceil n/2 \rceil, \\ B_i - B_1 & \text{en otro caso,} \end{cases}$$

$$FN_A := \{B \in \mathcal{CT} \mid B \text{ es forma normal de } A\},$$

$$e \quad i = \operatorname{mín}\{j = 2, \dots, n-1 \mid \forall B, B' \in FN_A, B_j - B_1 \neq B'_j - B'_1\}.$$

En el caso del conjunto atonal  $[5 \ 9 \ 0 \ 4]$ , se tienen dos formas normales:  $[4 \ 5 \ 9 \ 0]$  y  $[9 \ 0 \ 4 \ 5]$ . Nótese ahora que la entrada mínima en la que las diferencias son distintas es  $i = 2$ . Luego,  $\mathcal{F}'_n([4 \ 5 \ 9 \ 0]) = 1$  y  $\mathcal{F}'_n([9 \ 0 \ 4 \ 5]) = 3$ . Por lo tanto, la mejor forma normal del conjunto tonal asociado a *FaM7* es  $[4 \ 5 \ 9 \ 0]$ .

**Definición 1.6.**  $A \in \mathcal{CT}$  está en su *forma básica* si es de la forma

$$A := [A_1 \ \dots \ A_n] = [0 \ (B_2 + 12 - B_1) \ \text{mód. } 12 \ \dots \ (B_n + 12 - B_1) \ \text{mód. } 12];$$

donde  $[B_1 \ \dots \ B_n] \in P_n^A$  corresponde a la mejor forma normal de  $A$ .

Finalmente, la forma básica del conjunto tonal  $[5 \ 9 \ 0 \ 4]$  es  $[0 \ 1 \ 5 \ 8]$ , puesto que su mejor forma normal es  $[4 \ 5 \ 9 \ 0]$ , y

$$[0 \ (5 + 12 - 4) \ \text{mód. } 12 \ (9 + 12 - 4) \ \text{mód. } 12 \ (0 + 12 - 4) \ \text{mód. } 12] = [0 \ 1 \ 5 \ 8].$$

**Definición 1.7.** La *inversión* de un conjunto tonal  $[A_1 \dots A_n] \in \mathcal{CT}$  es una transformación  $I_n : \{0, \dots, 11\}^n \rightarrow \{0, \dots, 11\}^n$  dada por

$$I_n([A_1 \dots A_n]) = [(12 - A_1) \text{ mód. } 12 \dots (12 - A_n) \text{ mód. } 12].$$

**Proposición 1.1.** Para todo  $A \in \mathcal{CT}$ ,  $(I_n \circ I_n)(A) = A$ .

*Demostración.*

Primero, notemos que si  $z = 0$ ,

$$\begin{aligned} (12 - (12 - z) \text{ mód. } 12) \text{ mód. } 12 &= (12 - (12) \text{ mód. } 12) \text{ mód. } 12 \\ &= (12) \text{ mód. } 12 = 0. \end{aligned}$$

Además, para todo  $z \in \{1, \dots, 11\}$  ocurre

$$\begin{aligned} (12 - (12 - z) \text{ mód. } 12) \text{ mód. } 12 &= (12 - (12 - z)) \text{ mód. } 12 \\ &= (z) \text{ mód. } 12 = z. \end{aligned}$$

Por lo tanto, para todo  $z \in \{0, \dots, 11\}$ ,

$$(12 - (12 - z) \text{ mód. } 12) \text{ mód. } 12 = z.$$

Por otro lado, sea  $A := [A_1 \dots A_n] \in \mathcal{CT}$ . Por lo anterior y las definiciones 1.2 y 1.7,

$$\begin{aligned} (I_n \circ I_n)(A) &= I_n(I_n(A)) = I_n(I_n([A_1 \dots A_n])) \\ &= I_n([(12 - A_1) \text{ mód. } 12 \dots (12 - A_n) \text{ mód. } 12]) \\ &= [(12 - (12 - A_1) \text{ mód. } 12) \text{ mód. } 12 \dots (12 - (12 - A_n) \text{ mód. } 12) \text{ mód. } 12] \\ &= [((12 - 12) + A_1 \text{ mód. } 12) \text{ mód. } 12 \dots ((12 - 12) + A_n \text{ mód. } 12) \text{ mód. } 12] \\ &= [A_1 \text{ mód. } 12 \dots A_n \text{ mód. } 12] = [A_1 \dots A_n] = A. \end{aligned}$$

■

**Definición 1.8.** Una *transposición*<sup>30</sup> de un conjunto tonal es una transformación  $T_n^t : \{0, \dots, 11\}^n \rightarrow$

---

<sup>30</sup>En la Teoría Musical habitual, refiere a cuando en una pieza musical se traslada de una tonalidad a otra diferente —es un recurso musical común, no sólo de la contemporaneidad—. Adicionalmente, cabe destacar que éste difiere del concepto matemático de transposición dentro de la teoría de grupos de permutaciones.

$\{0, \dots, 11\}^n$  dada para algún  $t \in \{0, \dots, 11\}$  por

$$T_n^t([A_1 \dots A_n]) = [(A_1 + t) \text{ mód. } 12 \dots (A_n + t) \text{ mód. } 12].$$

**Proposición 1.2.** Para todo  $A \in \mathcal{C}\mathcal{T}$ ,  $(T_n^{12-t} \circ T_n^t)(A) = A$ .

*Demostración.*

Sean  $A := [A_1 \dots A_n] \in \mathcal{C}\mathcal{T}$  y  $t \in \{0, \dots, 11\}$ . Por la Definición 1.8 y propiedad de los módulos aritméticos sobre  $\mathbb{Z}$ ,

$$\begin{aligned} (T_n^{12-t} \circ T_n^t)(A) &= T_n^{12-t}(T_n^t(A)) = T_n^{12-t}(T_n^t([A_1 \dots A_n])) \\ &= T_n^{12-t}([(A_1 + t) \text{ mód. } 12 \dots (A_n + t) \text{ mód. } 12]) \\ &= [(((A_1 + t) \text{ mód. } 12) + (12 - t)) \text{ mód. } 12 \dots \\ &\quad (((A_n + t) \text{ mód. } 12) + (12 - t)) \text{ mód. } 12] \\ &= [(A_1 + t + 12 - t) \text{ mód. } 12 \dots (A_n + t + 12 - t) \text{ mód. } 12] \\ &= [(A_1 + 12) \text{ mód. } 12 \dots (A_n + 12) \text{ mód. } 12] = [A_1 \dots A_n] = A. \end{aligned}$$

■

Regresando al ejemplo de  $[5\ 9\ 0\ 4] \in \mathcal{C}\mathcal{T}$  asociado al acorde  $FaM7$ , su conjunto tonal “inverso” es  $[7\ 3\ 0\ 8]$ . Además, sus once transposiciones se encuentran ilustradas en la Figura 1.12.

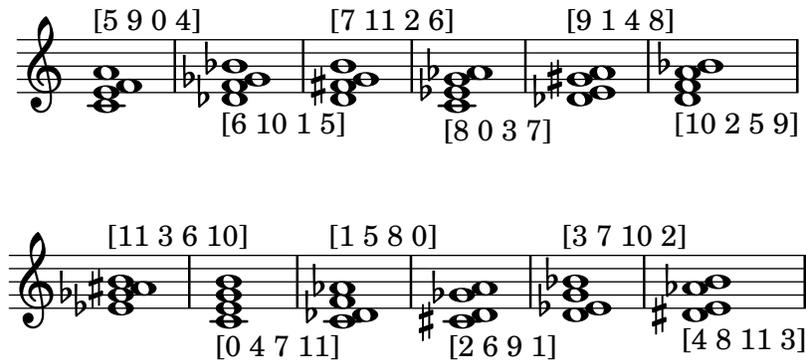


Figura 1.12: Transposiciones del conjunto tonal asociado a  $FaM7$  (notación musical y representación matemática). Para la representación de los acordes en notación musical, se escogió la inversión del acorde tal que sus notas se encontraran en la quinta octava en la escala Do Mayor.

**Definición 1.9.** Se dice que  $A, B \in \mathcal{C}\mathcal{T}$  son equivalentes si y sólo si son reducibles a la misma forma básica mediante la aplicación de transposiciones o inversiones seguidas de transposiciones. Se denotará por  $A \equiv_M B$  al hecho de que A y B sean equivalentes. Además, a una colección de

conjuntos tonales equivalentes se le llama *clase de conjunto*.

**Observación.** La equivalencia entre acordes es una relación de equivalencia ya que, claramente,

- Para todo  $A \in \mathcal{CT}$ , ocurre  $A \equiv_M A$ .
- Para todo  $A, B \in \mathcal{CT}$ , si  $A \equiv_M B$  entonces  $B \equiv_M A$ .
- Para todo  $A, B, C \in \mathcal{CT}$ , si  $A \equiv_M B$  y  $B \equiv_M C$ , entonces  $A \equiv_M C$ .

Cabe destacar, como se habrá notado con la Figura 1.12, que las herramientas postonales presentadas no están basadas en escalas o tonalidades, por lo que funcionan para análisis de Música Postonal. No obstante, como se verá más adelante, también son aplicables a la Música con armonía de fundamento tonal de épocas anteriores. De hecho, gran parte de la modelación se llevará a cabo a través de las clases de conjuntos y el siguiente concepto.

**Definición 1.10.** El *vector de clase de intervalo* asociado a un conjunto tonal  $A := [A_1 \dots A_n] \in \mathcal{CT}$  corresponde a la función

$$\mathcal{I}_n : \{0, \dots, 11\}^n \rightarrow \left\{0, \dots, \binom{n}{2}\right\}^6 \text{ dada por}$$

$$\mathcal{I}_n(A) = \left[ \sum_{(i,j) \in I_n} \mathbb{1}_{S_1}(A_i, A_j) \dots \sum_{(i,j) \in I_n} \mathbb{1}_{S_6}(A_i, A_j) \right];$$

donde

$$I_n = \{(i, j) \in \{1, \dots, n\}^2 \mid i < j\}, \text{ y}$$

$$S_k = \{(x, y) \in \{0, \dots, 11\}^2 \mid (|x - y| \bmod 12) = k\}, \text{ con } k \in \{1, \dots, 6\}.$$

Nótese que el vector de clase de intervalo es, por así decirlo, un histograma del número de semitonos que separan cada par de las notas que integran el acorde. Por ejemplo, el vector de clase de intervalo de  $[5 \ 9 \ 0 \ 4]$  está dado por  $[1 \ 0 \ 1 \ 2 \ 2 \ 0]$ .

Ahora bien, lo intuitivo (e ideal) sería que los vectores de clase de intervalo fueran una transformación invariante bajo transposiciones e inversiones seguidas de transposiciones, es decir, que indujeran una manera única de caracterizar las clases de conjunto. No obstante, existen contraejemplos de este hecho, en concreto, diecinueve. Allen Forte en [11], realizó una clasificación exhaustiva de clases de conjunto e identificó aquéllas que no cumplían con esta propiedad. Lo relativo a la clasificación de Forte puede ser consultada en el *Anexo I*.

**Definición 1.11.** Se dice que dos clases de conjunto son *Z-relativas* si comparten el mismo vector de clase de intervalo, pero no son equivalentes.

La indagación hacia los factores que determinan si las clases de conjunto son Z-relativas se desvían del objetivo de la presente investigación. Sin embargo, es un tema que se ha abordado de una manera sofisticada y matemática en documentos como [24] de John Mandereau, Daniele Ghisib, Emmanuel Amiotc, Moreno Andreatad y Carlos Agond.

La implicación que sí se sigue de la Definición 1.11 y es clara, es que si dos conjuntos tonales son equivalentes, tendrán el mismo vector de clase de intervalo puesto que son reducibles a la misma forma básica. Un cuestionamiento lógico consecuente es, en caso de que dos clases de conjunto no sean equivalentes, ¿cómo medir qué tan similares son? Existen varias propuestas en cuanto a medidas de similitud entre clases de conjunto a través de sus vectores de clase de intervalo, algunas de las cuales son enunciadas en [22, pp. 50-61].

## 1.7. Herramientas musicológicas propuestas

Las medidas de similitud entre conjuntos tonales referenciadas al final de la sección anterior están inspiradas en la psicoacústica, que no es el enfoque que se tomará en la investigación. Por lo tanto, en la presente sección se proponen medidas basadas en la consonancia contrapuntística, la vía que se seguirá, debido a que bajo tal enfoque se compuso la mayor parte de las piezas musicales académicas a analizar. Adicionalmente, a pesar de que no se habla de tonalidad en su planteamiento, la Teoría Musical de Conjuntos es un sistema basado en los intervalos de las notas que componen un acorde, lo que lo hace congruente y aplicable a la armonía tonal y la conceptualización contrapuntística de la consonancia.

De hecho, bajo tal precepto, las entradas del vector de clase de intervalo son categorizadas entre consonantes y disonantes como se muestra a continuación<sup>31</sup>:

Segunda Menor	←	m2	[ <b>Disonante,</b>	M7	→	Séptima Mayor
Segunda Mayor	←	M2	<b>Disonante,</b>	m7	→	Séptima Menor
Tercera Menor	←	m3	<b>Consonante,</b>	M6	→	Sexta Mayor
Tercera Mayor	←	M3	<b>Consonante,</b>	m6	→	Sexta Menor
Cuarta Justa	←	P4	<b>Consonante,</b>	P5	→	Quinta Justa
Tritono	←	A4	<b>Disonante</b>	] d5	→	Tritono

Cuadro 1.2: Categorización de las entradas de los vectores de clase de intervalo de acuerdo con la consonancia contrapuntística.

Lo anterior da lugar a la puntualización de dos observaciones:

- Con el vector de clase de intervalo de un conjunto tonal (acorde) se contabilizan las frecuencias de los tipos de intervalo de las clases de notas (notas) que lo componen dos a dos. El unísono o la octava perfecta se conciben como la relación de *consonancia perfecta* entre dos notas, razón por la cual se excluyen del vector de clase de intervalo, dado que lo *interesante* es analizar las sensaciones que transmiten los otros tipos de intervalo.
- A pesar de que existen 11 tipos de intervalo tonal básico (excluyendo el unísono y la octava perfecta), en este contexto, dos clases de notas pueden tener a lo más 6 semitonos de diferencia, por lo que el vector de clase de intervalo tiene únicamente 6 entradas. Nótese que esto

<sup>31</sup>Véase la Figura 1.8 y los intervalos básicos definidos en la Sección 1.4.

se debe a que si la clase de nota  $n_i$  está separada en la escala de la clase de nota  $n_j$  por un intervalo de  $x$  semitonos,  $n_j$  está separada de  $n_i$  por  $12 - x$  semitonos (ya que la construcción de las clases de notas no contempla la octava en que se encuentra la nota asociada).

Ahora bien, la naturaleza de la consonancia contrapuntística y la Teoría Musical de Conjuntos inspiró a la generación de las herramientas musicológicas propias siguientes.

**Definición 1.12.** La *medida de consonancia relativa* de un conjunto tonal  $A = [A_1 \dots A_n] \in \mathcal{CT}$  corresponde a la función  $\mathcal{C} \circ \mathcal{I}_n$ ; donde  $\mathcal{C} : \{0, \dots, \binom{n}{2}\}^6 \rightarrow [0, 1]$  está dada por

$$\mathcal{C}(V_1, \dots, V_6) = \begin{cases} \frac{(V_3+V_4+V_5)-(V_1+V_2+V_6)}{n(n-1)} + \frac{1}{2} & \text{si } n > 1, \\ \frac{(V_3+V_4+V_5)-(V_1+V_2+V_6)+1}{2} & \text{si } n = 1. \end{cases}$$

La construcción de tal medida de consonancia nace del razonamiento de que la “consonancia total” de un conjunto tonal, dada la categorización contrapuntística de sus entradas, puede expresarse como  $(V_3 + V_4 + V_5) - (V_1 + V_2 + V_6)$ , cantidad que se encuentra en  $\{-\binom{n}{2}, -\binom{n}{2} + 1, \dots, -1, 0, 1, \dots, \binom{n}{2}\}$  puesto que  $\binom{n}{2}$  es la cantidad de intervalos dentro del conjunto tonal  $[A_1 \dots A_n]$ . Realizando las operaciones necesarias para hacer que la diferencia se encuentre en el intervalo  $[0, 1]$ , se llega a la expresión de la Definición 1.12. A su vez, esta definición da introducción natural al siguiente concepto.

**Definición 1.13.** La *medida de disonancia relativa* de un conjunto tonal  $[A_1 \dots A_n] \in \mathcal{CT}$  corresponde a la función  $\mathcal{D} \circ \mathcal{I}_n$ ; donde  $\mathcal{D} : \{0, \dots, \binom{n}{2}\}^6 \rightarrow [0, 1]$  está dada por  $\mathcal{D}(V_1, \dots, V_6) = 1 - \mathcal{C}(V_1, \dots, V_6)$ .

Nótese las medidas de consonancia y disonancia definidas resultan en cantidades relativas respecto la cardinalidad de un conjunto tonal. Bajo la idea de que no sólo la proporción de intervalos consonantes determina el grado de consonancia o disonancia de un conjunto tonal, sino también su cantidad absoluta, se hace la construcción de las siguientes medidas *absolutas*.

**Definición 1.14.** La *medida de consonancia absoluta* de un conjunto tonal  $A = [A_1 \dots A_n] \in \mathcal{CT}$

corresponde a la función  $\mathcal{C}^* \circ \mathcal{I}_n$ ; donde  $\mathcal{C}^* : \{0, \dots, \binom{n}{2}\}^6 \rightarrow [0, 1]$  está dada por

$$\mathcal{C}^*(V_1, \dots, V_6) = \frac{(V_3 + V_4 + V_5) - (V_1 + V_2 + V_6)}{2 \cdot 66} + \frac{1}{2}.$$

Cabe destacar que existe un motivo por el que el primer sumando de la medida de consonancia absoluta sea dividido por 66. Bajo la construcción de un conjunto tonal, el máximo valor que la suma de las entradas del vector de clase de intervalo<sup>32</sup> puede tomar es  $\binom{12}{2} = 66$ .

**Definición 1.15.** La *medida de disonancia absoluta* de un conjunto tonal  $[A_1 \dots A_n] \in \mathcal{CT}$  corresponde a la función  $\mathcal{D}^* \circ \mathcal{I}_n$ ; donde  $\mathcal{D}^* : \{0, \dots, \binom{n}{2}\}^6 \rightarrow [0, 1]$  está dada por  $\mathcal{D}^*(V_1, \dots, V_6) = 1 - \mathcal{C}^*(V_1, \dots, V_6)$ .

Ahora bien, puede considerarse únicamente la separación de las clases de notas de un conjunto tonal mediante una estructura simplificada que únicamente contemple aquellos intervalos disonantes y consonantes. Cabe hacer notar que de esta manera se está haciendo el supuesto de que los intervalos disonantes son disonantes en la misma medida y, análogamente, para los consonantes. Por ejemplo, que un tritono es tan disonante como una segunda menor y una sexta mayor es tan consonante como una tercera mayor.

**Definición 1.16.** El *vector de clase de intervalo reducido* asociado a un conjunto tonal  $A := [A_1 \dots A_n] \in \mathcal{CT}$  es la función

$$\mathcal{I}_n^* : \{0, \dots, 11\}^n \rightarrow \left\{0, \dots, \binom{n}{2}\right\}^2 \text{ dada por}$$

$$\mathcal{I}_n^*(A) = \left[ \sum_{(i,j) \in I_n} \mathbb{1}_{S_1^*}(A_i, A_j) \quad \sum_{(i,j) \in I_n} \mathbb{1}_{S_2^*}(A_i, A_j) \right];$$

donde

$$I_n = \{(i, j) \in \{1, \dots, n\}^2 \mid i < j\},$$

$$S_1^* = \{(x, y) \in \{0, \dots, 11\}^2 \mid (|x - y| \bmod 12) = k; k \in \{3, 4, 5\}\}, \text{ y}$$

---

<sup>32</sup>La cantidad de relaciones dos a dos posibles entre las notas que lo conforman.

$$S_2^* = \{(x, y) \in \{0, \dots, 11\}^2 \mid (|x - y| \bmod 12) = k; k \in \{1, 2, 6\}\}.$$

Por ejemplo, se había mostrado que el vector de clase de intervalo del conjunto tonal [5 9 0 4] está dado por [1 0 1 2 2 0]. Entonces, el vector de clase de intervalo reducido del mismo conjunto tonal es [5 1].

Bajo la conceptualización contrapuntística de consonancia, la primer entrada de tal estructura está asociada a los intervalos consonantes, mientras que la segunda a los disonantes; por lo que se denotará en el cuerpo del presente documento por  $[C D]$  a los vectores de clase de intervalo reducidos. Adicionalmente, lo anterior da pie a la definición de otras medidas de consonancia y disonancia.

**Definición 1.17.** Se define la *medida de consonancia*  $i$ ,  $i \in \{1, 2, 3\}$  de un conjunto tonal  $A = [A_1 \dots A_n] \in \mathcal{C} \mathcal{T}$  como la función  $\mathcal{C}_i \circ \mathcal{I}_n^*$ ; donde  $\mathcal{C}_i : \{0, \dots, \binom{n}{2}\}^2 \rightarrow [0, 1]$  está dada por:

$$\begin{array}{lll} 1. & & 3. \\ \mathcal{C}_1(C, D) = \frac{C}{66}. & \mathcal{C}_2(C, D) = \begin{cases} \frac{C-D}{n \cdot (n-1)} + \frac{1}{2} & \text{si } n > 1, \\ \frac{C-D}{2} + \frac{1}{2} & \text{si } n = 1. \end{cases} & \mathcal{C}_3(C, D) = \frac{C-D}{2 \cdot 66} + \frac{1}{2}. \end{array}$$

**Definición 1.18.** Se define la *medida de disonancia*  $i$ ,  $i \in \{1, 2, 3\}$  de un conjunto tonal  $A = [A_1 \dots A_n] \in \mathcal{C} \mathcal{T}$  como la función  $\mathcal{D}_i \circ \mathcal{I}_n^*$ ; donde  $\mathcal{D}_i : \{0, \dots, \binom{n}{2}\}^2 \rightarrow [0, 1]$  está dada por:

$$\begin{array}{lll} 1. & 2. & 3. \\ \mathcal{D}_1(C, D) = -\frac{D}{66}. & \mathcal{D}_2(C, D) = 1 - \mathcal{C}_2(C, D). & \mathcal{D}_3(C, D) = 1 - \mathcal{C}_3(C, D). \end{array}$$

Adicionalmente, las medidas propuestas fungen, no sólo como cuantificadores de similitud entre conjuntos tonales, sino también pueden pensarse como un medidas del *flujo de consonancia y de disonancia* dentro de un conjunto tonal. Lo cual lleva a preguntarse por alguna medida de variación en la dinámica del *tempo armónico*<sup>33</sup> dada una sucesión de acordes. Supóngase que tenemos una sucesión de  $k$  acordes, los cuales pueden ser representados mediante su forma básica asociada a una clase de conjunto y, por ende, al vector de clase de intervalo reducido. Se propone cuantificar el intercambio de flujo de consonancia y disonancia en la sucesión desde la primera intervención hasta la última, tomando en cuenta la inmediata anterior mediante las funciones propuestas en las

<sup>33</sup>También denominado *ritmo armónico*, refiere a la regularidad con que los acordes cambian.

definiciones 1.19 y 1.20.

**Definición 1.19.** La medida de consonancia de los conjuntos tonales asociados una sucesión de  $k$  acordes  $A^1, \dots, A^k \in \mathcal{CT}$  con  $A^i = [A_1^i \dots A_{n_i}^i]$ ,  $n_i \in \{1, \dots, \binom{12}{2}\}$ ,  $i \in \{1, \dots, k\}$  corresponde a la función  $\mathcal{C}_k^* \circ (\mathcal{I}_{n_1}^*, \dots, \mathcal{I}_{n_k}^*)$ ; donde

$$\mathcal{C}_k^* : \left\{0, \dots, \binom{n}{2}\right\}^2 \times \dots \times \left\{0, \dots, \binom{n}{2}\right\}^2 \text{ (} k \text{ veces)} \rightarrow [0, 1] \text{ está dada por}$$

$$\mathcal{C}_k^*((C_1, D_1), \dots, (C_k, D_k)) = \begin{cases} \frac{66+C_k+2\cdot C_{k-1}-D_1}{2\cdot(66+C_{k-1}+D_{k-1})} & \text{si } k \geq 3, \\ \frac{66+C_2-D_1}{2\cdot 66} & \text{si } k = 2. \end{cases}$$

La motivación de la definición 1.19 subyace en que la ganancia neta de consonancia en la sucesión puede caracterizarse de esta manera:

$(C_k$	$-$	$D_{k-1})$	$+$	$(C_{k-1}$	$-$	$D_1)$
Consonancia		Disonancia		Consonancia		Disonancia
ganada		perdida		ganada		perdida
Pasando	de	k-1 a k.		Pasando	de	1 a k-1.

**Definición 1.20.** La medida de disonancia de los conjuntos tonales asociados a una sucesión de  $k$  acordes  $A^1, \dots, A^k \in \mathcal{CT}$  con  $A^i = [A_1^i \dots A_{n_i}^i]$ ,  $n_i \in \{1, \dots, \binom{12}{2}\}$ ,  $i \in \{1, \dots, k\}$  corresponde a la función  $\mathcal{D}_k^* \circ (\mathcal{I}_{n_1}^*, \dots, \mathcal{I}_{n_k}^*)$ ; donde

$$\mathcal{D}_k^* : \left\{0, \dots, \binom{n}{2}\right\}^2 \times \dots \times \left\{0, \dots, \binom{n}{2}\right\}^2 \text{ (} k \text{ veces)} \rightarrow [0, 1] \text{ está dada por}$$

$$\mathcal{D}_k^*((C_1, D_1), \dots, (C_k, D_k)) = \begin{cases} \frac{66+D_k+2\cdot D_{k-1}-C_1}{2\cdot(66+C_{k-1}+D_{k-1})} & \text{si } k \geq 3, \\ \frac{66+D_2-C_1}{2\cdot 66} & \text{si } k = 2. \end{cases}$$

Análogamente a la medida de consonancia, la definición anterior se concretó bajo la idea siguiente:

$$\begin{array}{cccc}
(D_k & - & C_{k-1}) & + & (D_{k-1} & - & C_1) \\
\text{Disonancia} & & \text{Consonancia} & & \text{Disonancia} & & \text{Consonancia} \\
\text{ganada} & & \text{perdida} & & \text{ganada} & & \text{perdida} \\
\\ 
\text{Pasando} & \text{de} & \text{k-1 a k.} & & \text{Pasando} & \text{de} & \text{1 a k-1.}
\end{array}$$

Explicación detallada sobre las medidas de consonancia y disonancia planteadas en las últimas dos definiciones, junto con sus posibles interpretaciones, se presentan en el Capítulo 5, Sección 2.2.

En la Figura 1.13 se encuentran las gráficas de las medidas de consonancia y disonancia propuestas aplicadas a las sucesiones de acordes<sup>34</sup> de *Polonaise Op. 89 en Do mayor* de Beethoven. Se excluye de las gráficas la medida  $\mathcal{C}_2$  puesto que es equivalente a la de consonancia relativa para vectores de clase de intervalo, así como sus medidas de disonancia asociadas. El mismo argumento es cierto para  $\mathcal{C}_3$  y las medidas de consonancia absolutas.

Nótese que, excluyendo la diferencia en las escalas, las diferentes medidas de consonancia y disonancia son muy parecidas, excepto por el hecho de que algunas resultan más suaves y otras más picudas. Dado que surgieron de razonamientos diferentes, sobre todo las últimas tres, esto puede ser indicador de consistencia en la interpretación de consonancia y disonancia y su cuantificación a través de los vectores de clase de intervalo (“clásicos” o reducidos).

Adicionalmente, se puede apreciar de manera clara la resolución de la disonancia a la consonancia, así como cierta periodicidad en el uso de la disonancia (lo cual es completamente congruente con las funciones armónicas de la Teoría de la Armonía). Más aún, cabe destacar que al aumentar el número de acordes  $k$  contemplados en las sucesiones para las medidas de consonancia y disonancia,  $\mathcal{C}_k^*$  y  $\mathcal{D}_k^*$ , el cambio en las valuaciones de las medidas es menor (al menos en su tendencia). Lo que, podría ser evidencia de que tales conceptos tienden a cierto punto, congruente con la teoría también. Estos patrones, que de hecho van haciéndose más claros conforme la pieza musical avanza en el tiempo, son precisamente lo que se pretende analizar y, luego, predecir con base en su dinamismo a lo largo de la obra. Asimismo, al observar la Figura 1.13, es posible encontrar la motivación para el uso de estructuras probabilísticas como herramientas para la modelación.

---

<sup>34</sup>Excluyendo notas de paso, dado que al ser ajenas al acorde, no aportan a su estructura.

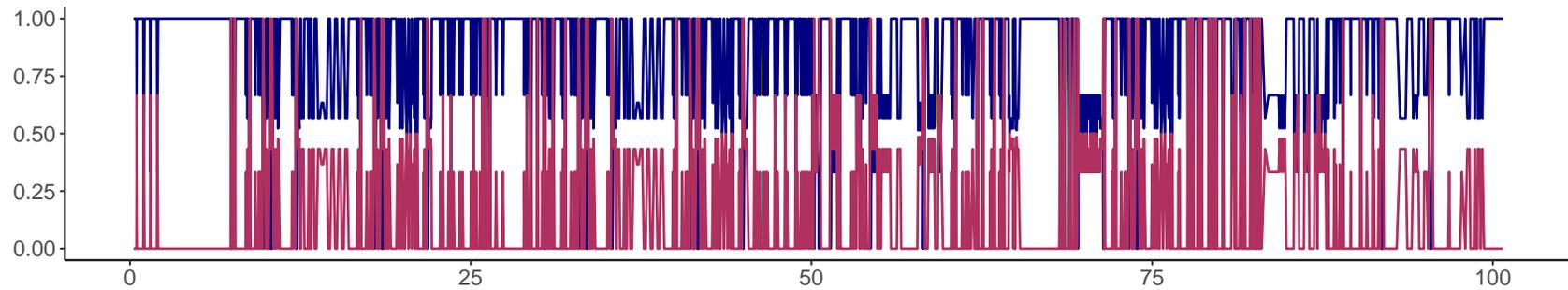
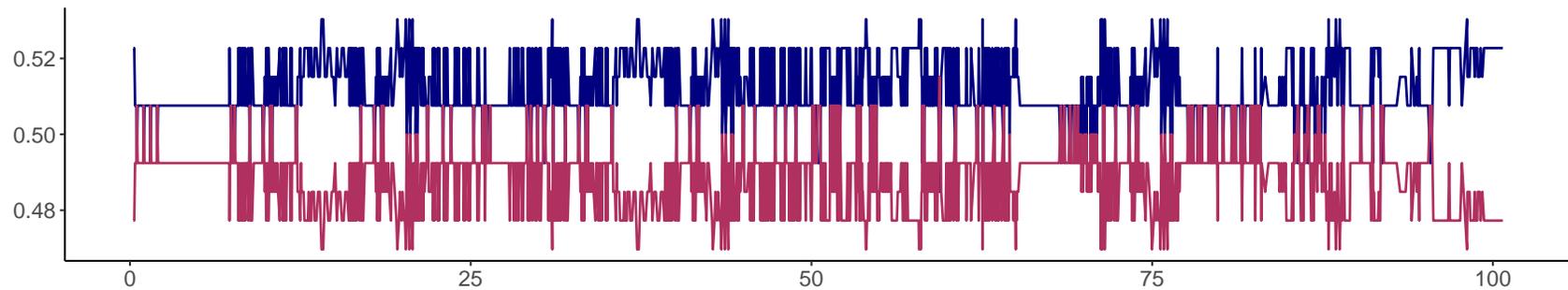
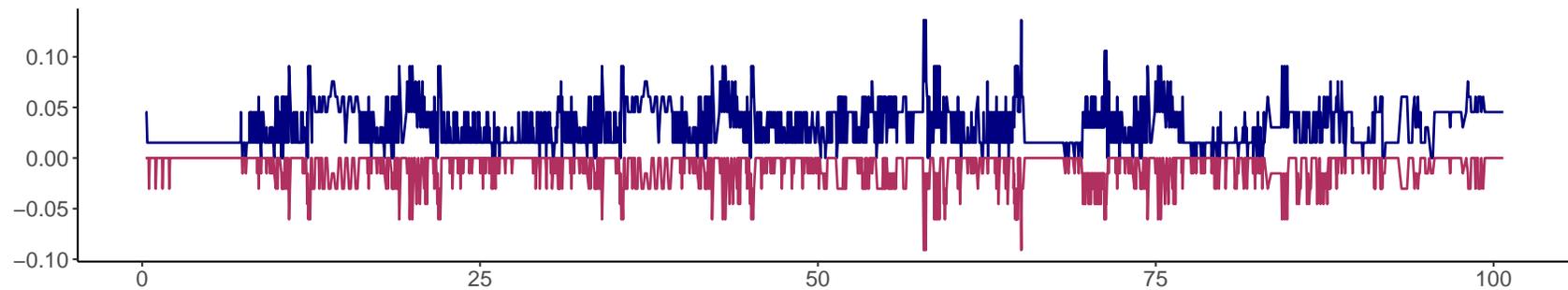
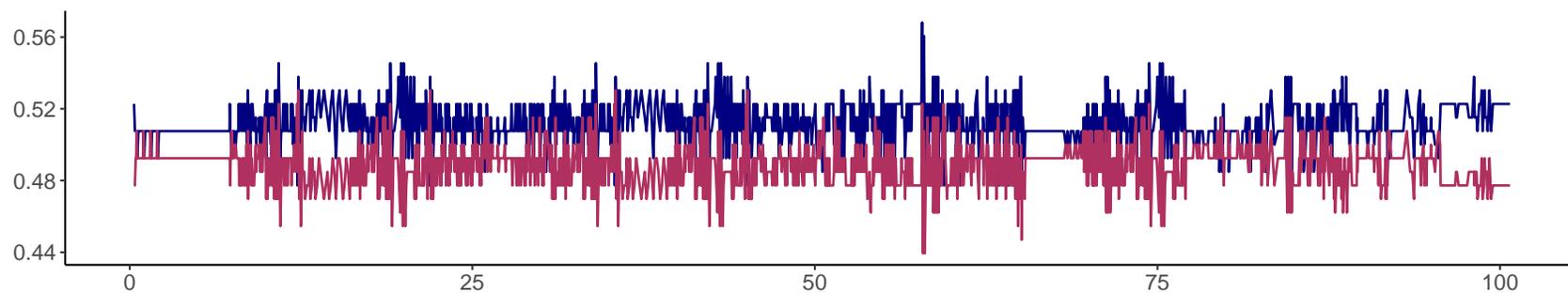
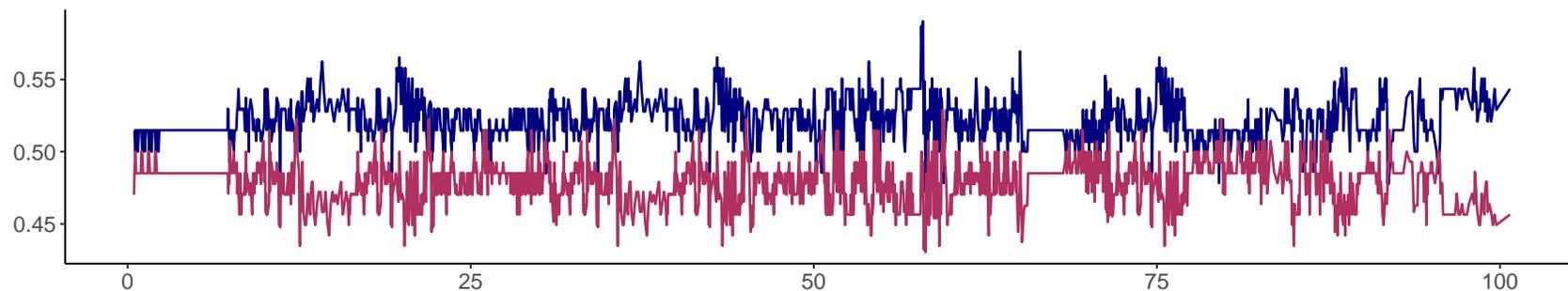
(a) Medidas relativas ( $\mathcal{C}$  y  $\mathcal{D}$ ) y medidas 2 ( $\mathcal{C}_2$  y  $\mathcal{D}_2$ ).(b) Medidas absolutas ( $\mathcal{C}^*$  y  $\mathcal{D}^*$ ) y medidas 3 ( $\mathcal{C}_3$  y  $\mathcal{D}_3$ ).(c) Medidas 1 ( $\mathcal{C}_1$  y  $\mathcal{D}_1$ ).

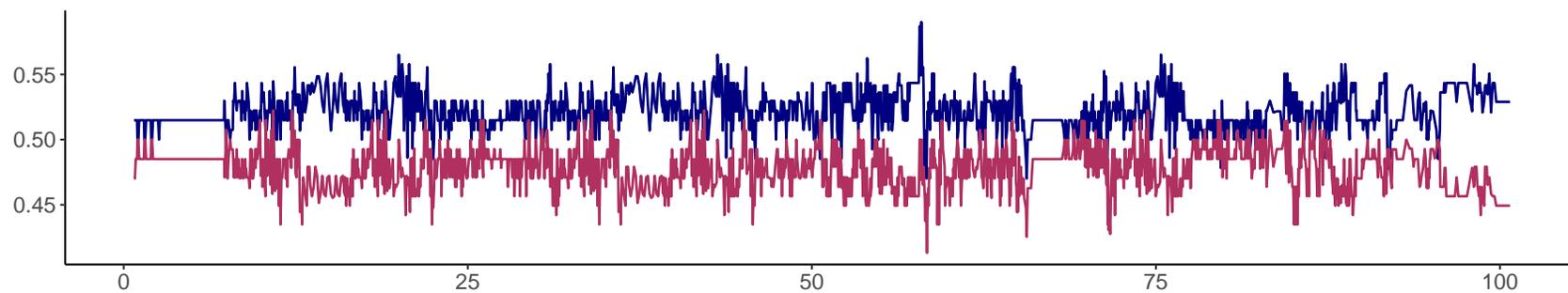
Figura 1.13: Medidas de consonancia y disonancia planteadas aplicadas a *Polonaise Op. 89 en Do mayor* de Beethoven. El eje horizontal representa el tiempo (compases); la línea azul corresponde a la medida de consonancia y la magenta a la medida de disonancia asociada a ella.



(d) Medidas de los conjuntos tonales asociados a una sucesión de 2 acordes ( $\mathcal{C}_2^*$  y  $\mathcal{D}_2^*$ ).



(e) Medidas de los conjuntos tonales asociados a una sucesión de 5 acordes ( $\mathcal{C}_5^*$  y  $\mathcal{D}_5^*$ ).



(f) Medidas de los conjuntos tonales asociados a una sucesión de 10 acordes ( $\mathcal{C}_{10}^*$  y  $\mathcal{D}_{10}^*$ ).

Figura 1.13: Medidas de consonancia y disonancia planteadas aplicadas a *Polonaise Op. 89 en Do mayor* de Beethoven. El eje horizontal representa el tiempo (compases); la línea azul corresponde a la medida de consonancia y la magenta a la medida de disonancia asociada a ella.



## CAPÍTULO 2

## CADENAS DE MARKOV

La indagación sobre Teoría Musical llevó a concebir a la Música como un sistema dinámico. La motivación subyace en el hecho de que una pieza musical no es estática o inamovible: el autor utiliza múltiples recursos a lo largo la misma con el fin de plasmar ideas y causar sensaciones a través del reposo y la tensión. Siguiendo la Teoría de la Armonía, esto es llevado a cabo con cierta intención y cierto tempo armónico, resultando en patrones que, de hecho, son observables en la Figura 1.13 y que son más marcados al transcurrir el tiempo.

Consecuentemente, cabe preguntarse por el comportamiento de las intervenciones en una obra musical al largo plazo y, luego, cómo cuantificarlo. De aquí nació el pensamiento en la ergodicidad de la armonía de una pieza y en la utilización, como primer recurso, de *Cadenas de Markov* para su modelación. Por lo tanto, en el presente capítulo se proporciona la teoría matemática de Cadenas de Markov a utilizar, así como los aspectos teóricos de la Estimación por Máxima Verosimilitud relacionada que se utilizará para la aplicación de la modelación propuesta. Es debido hacer notar que lo siguiente es contenido estándar de Probabilidad, el cual fue adaptado de los capítulos 2 y 3 de [6, pp. 53-124].

## 2.1. Conceptos básicos

**Definición 2.1.** Dado un espacio de probabilidad  $(\Omega, \mathcal{F}, \mathbb{P})$ , un *proceso estocástico* es una colección de variables aleatorias  $\{X_t : t \in T\}$  indexadas al conjunto  $T$  (espacio parametral) que toman valores en el conjunto  $S$  (espacio de estados).

**Definición 2.2.** Sea  $\{X_n\}_{n \in \mathbb{N}}$  un proceso estocástico con espacio de estados  $S$  numerable o finito. Se dice que  $\{X_n\}_{n \in \mathbb{N}}$  es una *Cadena de Markov* si satisface la *propiedad de Markov*, i.e., para todo  $n \geq 1$  y para todo  $s, x_0, x_1, \dots, x_{n-1} \in S$ ,

$$\mathbb{P}(X_n = s \mid X_0 = x_0, \dots, X_{n-1} = x_{n-1}) = \mathbb{P}(X_n = s \mid X_{n-1} = x_{n-1}).$$

Además, la evolución de una Cadena de Markov es caracterizada por sus probabilidades de transición  $\mathbb{P}(X_n = j \mid X_{n-1} = i)$ .

Considérese el clásico ejemplo de la *Ruina del jugador*, en que una persona con cierta cantidad de dinero  $X_0$  participa en un juego de apuestas de \$1. En cada ronda, se tiene cierta probabilidad  $p$  de ganar, mientras que con probabilidad  $1 - p$  se pierde. La Cadena de Markov  $\{X_n\}_{n \in \mathbb{N}}$  que denota la cantidad de dinero que tiene el jugador en la ronda  $n$  toma valores en el espacio de estados  $\mathbb{N}$  (se supondrá que la casa de apuestas tiene una cantidad infinita de dinero).

Adicionalmente, la cadena está caracterizada por las siguientes probabilidades de transición, puesto que el juego acaba cuando el jugador no tiene más dinero que apostar,

$$\mathbb{P}(X_n = j \mid X_{n-1} = i) = \begin{cases} p & \text{si } j = i + 1, \\ 1 - p & \text{si } j = i - 1, \\ 1 & \text{si } i = j = 0, \\ 0 & \text{en otro caso.} \end{cases}$$

**Definición 2.3.** Se dice que una Cadena de Markov  $\{X_n\}_{n \in \mathbb{N}}$  es *homogénea* si para todo  $n \geq 1$ ,  $\mathbb{P}(X_n = j \mid X_{n-1} = i) = \mathbb{P}(X_1 = j \mid X_0 = i)$ .

**Definición 2.4.** Se define la *matriz de transición* de una Cadena de Markov Homogénea  $\{X_n\}_{n \in \mathbb{N}}$  como  $(P)_{ij} = \mathbb{P}(X_1 = j \mid X_0 = i)$ , con  $i, j \in S$ .

Dado el ejemplo anterior, nótese que como las probabilidades de perder o ganar son iguales en cada ronda, entonces la cadena asociada al dinero del jugador es una Cadena de Markov Homogénea. De hecho, su matriz de transición está dada como sigue.

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

**Proposición 2.1 (Ecuaciones Chapman-Kolmogorov).** Para una Cadena de Markov  $\{X_n\}_{n \in \mathbb{N}}$  ocurre que para todo  $n, m \in \mathbb{N}$  tales que  $0 \leq m \leq n$ ,

$$\mathbb{P}(X_n = j \mid X_0 = i) = \sum_{k \in S} \mathbb{P}(X_m = k \mid X_0 = i) \cdot \mathbb{P}(X_n = j \mid X_m = k).$$

*Demostración.*

Primero, nótese que por el Teorema de Probabilidad Total,

$$\mathbb{P}(X_n = j, X_0 = i) = \sum_{k \in S} \mathbb{P}(X_n = j, X_m = k, X_0 = i).$$

Luego, por definición de probabilidad condicional y la propiedad de Markov,

$$\begin{aligned} \mathbb{P}(X_n = j, X_m = k, X_0 = i) &= \mathbb{P}(X_n = j \mid X_m = k, X_0 = i) \cdot \mathbb{P}(X_m = k, X_0 = i) \\ &= \mathbb{P}(X_n = j \mid X_m = k, X_0 = i) \cdot \mathbb{P}(X_m = k \mid X_0 = i) \cdot \mathbb{P}(X_0 = i) \\ &= \mathbb{P}(X_n = j \mid X_m = k) \cdot \mathbb{P}(X_m = k \mid X_0 = i) \cdot \mathbb{P}(X_0 = i). \end{aligned}$$

Finalmente, se tiene lo siguiente.

$$\begin{aligned}
\mathbb{P}(X_n = j | X_0 = i) &= \frac{\mathbb{P}(X_n = j, X_0 = i)}{\mathbb{P}(X_0 = i)} = \frac{1}{\mathbb{P}(X_0 = i)} \sum_{k \in S} \mathbb{P}(X_n = j, X_m = k, X_0 = i) \\
&= \frac{1}{\mathbb{P}(X_0 = i)} \left[ \sum_{k \in S} \mathbb{P}(X_n = j | X_m = k) \cdot \mathbb{P}(X_m = k | X_0 = i) \cdot \mathbb{P}(X_0 = i) \right] \\
&= \sum_{k \in S} \mathbb{P}(X_n = j | X_m = k) \cdot \mathbb{P}(X_m = k | X_0 = i).
\end{aligned}$$

Por lo tanto,  $\mathbb{P}(X_n = j | X_0 = i) = \sum_{k \in S} \mathbb{P}(X_m = k | X_0 = i) \cdot \mathbb{P}(X_n = j | X_m = k)$ .

■

**Notación.** Sea  $\{X_n\}_{n \in \mathbb{N}}$  una Cadena de Markov Homogénea con matriz de transición  $P$  y espacio de estados  $S$ . Para  $i, j \in S$ , se denotará  $P_{ij} := \mathbb{P}(X_1 = j | X_0 = i)$ .

**Proposición 2.2.** Para  $\{X_n\}_{n \in \mathbb{N}}$  una Cadena de Markov Homogénea con matriz de transición  $P$  ocurre que

$$\mathbb{P}(X_n = j | X_0 = i) = (P^n)_{ij}$$

*Demostración.*

Primero, utilizando las ecuaciones Chapman-Kolmogorov con  $m = 1$  se obtiene

$$\mathbb{P}(X_n = j | X_0 = i) = \sum_{k_1 \in S} \mathbb{P}(X_n = j | X_1 = k_1) \cdot \mathbb{P}(X_1 = k_1 | X_0 = i).$$

Pero, a su vez sucede que

$$\mathbb{P}(X_n = j | X_1 = k_1) = \sum_{k_2 \in S} \mathbb{P}(X_n = j | X_2 = k_2) \cdot \mathbb{P}(X_2 = k_2 | X_1 = k_1).$$

Por lo tanto,

$$\mathbb{P}(X_n = j | X_0 = i) = \sum_{k_1 \in S} \sum_{k_2 \in S} \left[ \mathbb{P}(X_n = j | X_2 = k_2) \cdot \mathbb{P}(X_2 = k_2 | X_1 = k_1) \cdot \mathbb{P}(X_1 = k_1 | X_0 = i) \right].$$

Aplicando inductivamente las ecuaciones Chapman-Kolmogorov y dado que la cadena es ho-

mogénea por hipótesis,

$$\begin{aligned} \mathbb{P}(X_n = j | X_0 = i) &= \sum_{k_1 \in S} \dots \sum_{k_{n-1} \in S} \left[ \mathbb{P}(X_n = j | X_{n-1} = k_{n-1}) \cdot \mathbb{P}(X_{n-1} = k_{n-1} | X_{n-2} = k_{n-2}) \cdot \dots \cdot \mathbb{P}(X_2 = k_2 | X_1 = k_1) \cdot \mathbb{P}(X_1 = k_1 | X_0 = i) \right] \\ &= \sum_{k_1 \in S} \dots \sum_{k_{n-1} \in S} P_{k_{n-1}j} \dots P_{k_1 k_2} P_{ik_1}. \end{aligned}$$

Finalmente, por la definición de producto de matrices,

$$\sum_{k_1 \in S} \dots \sum_{k_{n-1} \in S} P_{k_{n-1}j} \dots P_{k_1 k_2} P_{ik_1} = (P^n)_{ij}.$$

Por lo tanto,  $\mathbb{P}(X_n = j | X_0 = i) = (P^n)_{ij}$ . ■

## 2.2. Teorema de convergencia

**Definición 2.5.** Sean  $i, j \in S$ , si existe  $n \geq 0$  tal que  $\mathbb{P}(X_n = j | X_0 = i) > 0$ , se dice que el estado  $j$  es *accesible* desde el estado  $i$ . Si, además,  $i$  es accesible desde  $j$ , se dice que los estados *se comunican*.

**Definición 2.6.** Si todos los estados de una Cadena de Markov se comunican entre sí, se dice que la cadena es *conexa* o *irreducible*.

Cabe destacar que la cadena de la Ruina del jugador no es conexa, puesto que existe un estado que no se comunica con el resto, a saber, el estado 0. Esto es visualizable a continuación.

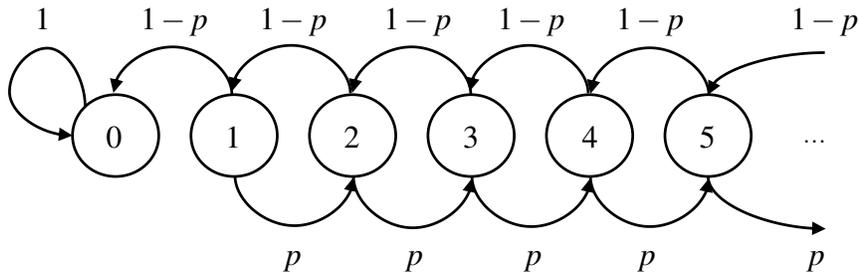


Figura 2.1: Dinámica de ganancia o pérdida del jugador: espacio de estados, sus clases de comunicación y probabilidades de transición.

**Definición 2.7.** Un estado  $i \in S$  es *recurrente* si para algún  $n \geq 1$ ,

$$\mathbb{P}(X_n = i | X_0 = i) = 1.$$

Es decir, si la probabilidad de regresar eventualmente a tal estado es unitaria.

**Definición 2.8.** Se denomina *tiempo de visita (hitting time)* de una Cadena de Markov a la variable aleatoria dada por

$$\tau_{ij} = \min\{n \geq 1 \mid X_n = j \text{ dado que } X_0 = i\}.$$

**Definición 2.9.** Un estado  $i \in S$  es *recurrente positivo* si  $\mathbb{E}(\tau_{ii}) < \infty$ .

**Definición 2.10.** Una *distribución estacionaria* para una Cadena de Markov con matriz de transición  $P$  es un vector de distribución de probabilidad tal que  $\pi P = \pi$ .

Ahora bien, la distribución estacionaria de una Cadena de Markov toma ese nombre en alusión a que, en caso de existir y ser única, corresponde a sus probabilidades límite. Es decir, a su distribución de probabilidad cuando  $n \rightarrow \infty$ ; a su comportamiento *a largo plazo*.

**Teorema 2.1.** Una Cadena de Markov irreducible es recurrente positiva (todos sus estados son recurrentes positivos) si y sólo si existe su distribución estacionaria. Más aún, la distribución estacionaria  $\pi$ , si existe, es única y sus entradas son estrictamente positivas<sup>1</sup>.

Para concluir el breve ejemplo de la Ruina del jugador, nótese que no todos los estados de la cadena son recurrentes ni recurrentes positivos puesto el estado 0 es *absorbente*, es decir, si el jugador pierde todo su dinero (o bien, desde un inicio no tenía dinero  $-X_0 = 0-$ ) no hay oportunidad de volver a jugar, ya que no tiene qué apostar. Así, por el Teorema 2.1, no existe una distribución estacionaria para la cadena; más aún, no es conexa.

**Teorema 2.2.** Una Cadena de Markov Homogénea asociada a un espacio de estados finito es recurrente positiva<sup>2</sup>.

---

<sup>1</sup>La demostración del teorema está disponible en [6, p. 104].

<sup>2</sup>Demostración consultable en [6, p. 105].

## 2.3. Estimación por Máxima Verosimilitud

En este apartado se presentan los aspectos teóricos de la estimación por Máxima Verosimilitud de la matriz de transición de una Cadena de Markov Homogénea, en medida de que será implementada para cubrir los objetivos de la presente investigación. Supóngase que se observan las realizaciones de una Cadena de Markov Homogénea  $X_0, X_1, \dots, X_n$  con espacio de estados  $S$  de cardinalidad  $m$ .

Utilizando inductivamente la definición de probabilidad condicional, la propiedad de Markov y el hecho de que la cadena es homogénea, se puede expresar su función de Verosimilitud  $\mathcal{L}$ :

$$\mathcal{L}(X_n, X_{n-1}, \dots, X_0; P) = P_{x_{n-1}x_n} \cdot \dots \cdot P_{x_0x_1} \cdot \mathbb{P}(X_0 = x_0 | P).$$

Sea  $n_{ij}$  el número de veces que la cadena tomó el estado  $i$  seguido del estado  $j$  según las observaciones; donde  $i, j \in S$ . Para simplificación de la escritura, se introduce la siguiente notación

$$n_i = \sum_{j=1}^m n_{ij} \quad \text{y} \quad \pi_{x_0} = \mathbb{P}(X_0 = x_0 | P).$$

Luego, nótese que las probabilidades  $P_{x_kx_r}$  se pueden expresar de la siguiente manera

$$P_{x_{n-1}x_n} \cdot \dots \cdot P_{x_0x_1} = \prod_{i,j \in S} P_{ij}^{n_{ij}}.$$

Ahora, puesto que la matriz de transición es una *matriz estocástica*<sup>3</sup>,  $P_{im} = 1 - \sum_{j=1}^{m-1} P_{ij}$ . Por lo que la función de Verosimilitud puede expresarse de esta manera

$$\mathcal{L}(X_n, X_{n-1}, \dots, X_0; P) = \pi_{x_0} \cdot \left[ \prod_{i=1}^m \prod_{j=1}^{m-1} P_{ij}^{n_{ij}} \right] \cdot \left[ \prod_{i=1}^m \left( 1 - \sum_{j=1}^{m-1} P_{ij} \right)^{n_{im}} \right].$$

Entonces, la función de log-Verosimilitud de la cadena está dada por

$$\ell(X_n, \dots, X_0; P) = \ln(\pi_{x_0}) + \sum_{i=1}^m \sum_{j=1}^{m-1} \left( n_{ij} \cdot \ln(P_{ij}) \right) + \sum_{i=1}^m \left( n_{im} \cdot \ln \left( 1 - \sum_{j=1}^{m-1} P_{ij} \right) \right).$$

Ahora bien, nótese que la función de log-Verosimilitud es diferenciable respecto  $P_{ij}$  con  $i \in$

---

<sup>3</sup>Sus entradas son no negativas y la suma de sus filas es unitaria.

$\{1, \dots, m\}$  y  $j \in \{1, \dots, m-1\}$ . Por lo que sus derivadas parciales existen y son

$$\frac{\partial \ell}{\partial P_{kr}} = \frac{n_{kr}}{P_{kr}} - \frac{n_{km}}{P_{km}}.$$

Igualando a cero se tiene que la función de log-Verosimilitud alcanza sus puntos críticos en

$$P_{kr} = \frac{P_{km} \cdot n_{kr}}{n_{km}}.$$

Para verificar que en dichos puntos alcanza su valor máximo, se obtienen sus derivadas parciales de segundo orden

$$\frac{\partial^2 l}{\partial P_{kr}^2} = -\frac{n_{kr}}{P_{kr}^2} - \frac{n_{km} \cdot P_{kr}}{(P_{km})^2}.$$

Dado que  $\frac{P_{km} \cdot n_{kr}}{n_{km}} > 0$ , se verifica la condición  $\frac{\partial^2 l}{\partial P_{kr}^2} \left( \frac{P_{km} \cdot n_{kr}}{n_{km}} \right) < 0$ . Por lo que, en efecto, la función de log-Verosimilitud alcanza su máximo en tales condiciones. Por lo tanto, para  $k \in \{1, \dots, m\}$  y  $r \in \{1, \dots, m-1\}$  se tienen los estimadores

$$\widehat{P}_{kr} = \frac{\widehat{P}_{km} \cdot n_{kr}}{n_{km}}.$$

Pero, como  $P_{im} = 1 - \sum_{j=1}^{m-1} P_{ij}$ , por la *propiedad de invarianza de los Estimadores Máximo Verosímiles*,

$$\widehat{P}_{km} = 1 - \sum_{j=1}^{m-1} \widehat{P}_{kj} = 1 - \sum_{j=1}^{m-1} \frac{\widehat{P}_{km} \cdot n_{kj}}{n_{km}}.$$

Expresión a la que es equivalente la siguiente ecuación

$$1 = \widehat{P}_{km} \cdot \left[ 1 + \sum_{j=1}^{m-1} \frac{n_{kj}}{n_{km}} \right] = \widehat{P}_{km} \cdot \left[ \frac{n_k}{n_{km}} \right].$$

Por lo que, finalmente, el Estimador por Máxima Verosimilitud de la matriz de transición de una Cadena de Markov Homogénea  $\widehat{P}$  tiene entradas dadas como sigue para  $k \in \{1, \dots, m\}$ ,

$$(\widehat{P})_{kr} = \begin{cases} \frac{n_{kr}}{n_k} & \text{si } r \in \{1, \dots, m-1\}, \\ 1 - \sum_{j=1}^{m-1} \widehat{P}_{kj} & \text{si } r = m. \end{cases}$$

## CAPÍTULO 3

# CAMPOS ALEATORIOS

Hasta el momento, ha quedado establecida la intención de conceptualizar matemáticamente una pieza musical como un sistema dinámico. Pese a que las Cadenas de Markov son una vía adecuada para la modelación, cabe cuestionarse si la Propiedad de Markov<sup>1</sup> es suficiente para caracterizar el fenómeno en cuestión. Es decir, si la mejor manera de pensar en la dinámica de las intervenciones en una obra musical es que acontecen en función únicamente de la inmediata anterior... Afirmación que probablemente cualquier músico o musicólogo refutaría. De este hecho nace la motivación para modelarla de una forma más general, extendiendo la afirmación planteada a más de una intervención antecedente, mediante Campos Aleatorios y Modelos de Markov de orden  $m - 1$ .

Dado lo anterior, se da pie a la introducción de la teoría de Campos Aleatorios, de Gibbs y Markov, en el presente capítulo. Para ello, fue tomado como referencia el Capítulo 7 de [6, pp. 253-322].

**Definición 3.1.** Sea  $S$  un conjunto finito de elementos denotados por  $s$  (*sitios*) y sea  $\Lambda$  un conjunto finito (*espacio fase*). Un *Campo Aleatorio* sobre  $S$  con *fases* en  $\Lambda$  es una colección de variables aleatorias  $X = \{X(s)\}_{s \in S}$  que toman valores en  $\Lambda$ .

Una aplicación sencilla de Campos Aleatorios son las redes neuronales artificiales. Considérese un Campo Aleatorio  $X$  sobre  $S = \mathbb{Z}_m^2$ ,  $m \in \mathbb{Z}^+$ , con fases en  $\Lambda = \{0, 1\}$ . Éste representará una red neuronal compuesta por  $m^2$  neuronas ordenadas en una red: cada sitio representa una neurona y las fases refieren a si ésta se encuentra en estado inhibitorio (0) o excitatorio (1). A lo largo del

---

<sup>1</sup>En el contexto de Cadenas de Markov.

presente capítulo, se utilizará  $m = 4$  para el desarrollo de este ejemplo.

**Definición 3.2.** Una *configuración* de un Campo Aleatorio es un elemento del *espacio de configuraciones*  $x \in \Lambda^S$  de la forma

$$x = \{x(s) \mid s \in S\};$$

donde, para todo  $s \in S$ ,  $x(s) \in \Lambda$ . Para una configuración  $x \in \Lambda^S$  y un subconjunto  $A \subset S$ , se define la restricción de  $x$  a  $A$  por  $x(A) = \{x(s) \mid s \in A\}$ .

La representación de la red neuronal  $X$  y una de sus posibles configuraciones se muestran en la Figura 3.1.

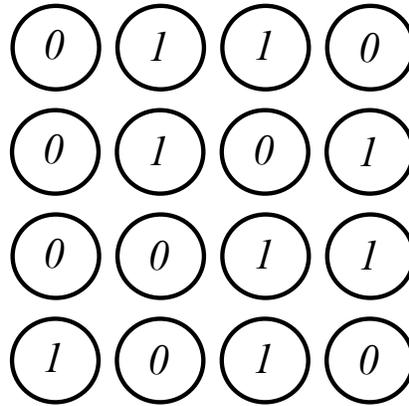


Figura 3.1: Ejemplo de configuración de la red neuronal artificial de 16 neuronas modelada mediante el Campo Aleatorio  $X$ , con  $S = \mathbb{Z}_4^2$ ,  $\Lambda = \{0, 1\}$ .

**Definición 3.3.** Un *sistema de vecindades* sobre  $S$  es una familia de subconjuntos de  $S$ ,  $N = \{N_s\}_{s \in S}$ , tal que para todo  $s, t \in S$  ocurre:

1.  $s \notin N_s$ .
2. Si  $t \in N_s$ , entonces  $s \in N_t$ .

Al subconjunto  $N_s$  se le llama *vecindad* del sitio  $s$  y la pareja  $(S, N)$  es llamada una gráfica o topología.

Para la red neuronal  $X$ , considérese la vecindad tipo  $\alpha$ , sobre la cual para un sitio  $s := (s_1, s_2) \in S$ ,  $N_s^\alpha = \{(s_1 + 1, s_2), (s_1 - 1, s_2), (s_1, s_2 + 1), (s_1, s_2 - 1)\}$ . Además, si  $s \in N_t$ , ocurre sinapsis entre la neurona  $s$  y la neurona  $t$ .

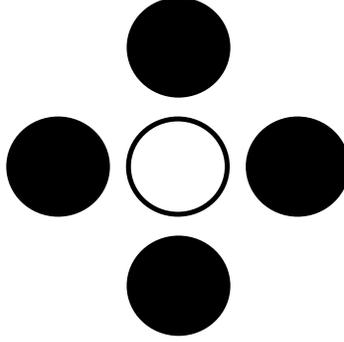


Figura 3.2: Sistema de vecindades tipo  $\alpha$  sobre un conjunto de sitios bidimensional. Los círculos negros representan los sitios vecinos al sitio blanco.

### 3.1. Campos Aleatorios de Markov

La Propiedad Markoviana, definida a continuación para Campos Aleatorios, será fundamental para la modelación, por lo que se centrará la atención en aquéllos que poseen dicha propiedad.

**Definición 3.4.** Un Campo Aleatorio  $X$  es un *Campo Aleatorio de Markov* respecto al sistema de vecindades  $N$  si para todo  $s \in S$ , dado  $X(N_s)$ ,

$$X(s) \perp X(S \setminus \tilde{N}_s);$$

donde  $\tilde{N}_s = N_s \cup \{s\}$ . Es decir, si para todo  $s \in S$  y para todo  $x \in \Lambda^S$ ,

$$\mathbb{P}(X(s) = x(s) \mid X(S \setminus s) = x(S \setminus s)) = \mathbb{P}(X(s) = x(s) \mid X(N_s) = x(N_s)).$$

Nótese que lo anterior implica que, en un Campo Aleatorio de Markov, las fases en que se realice un sitio son determinadas únicamente por las fases de sus vecinos; lo cual es caracterizado a través de su especificación local.

**Definición 3.5.** La *característica local* de un campo aleatorio de Markov  $X$  en el sitio  $s$  corresponde a la función  $\pi^s : \Lambda^S \rightarrow [0, 1]$  dada por

$$\pi^s(x) = \mathbb{P}(X(s) = x(s) \mid X(N_s) = x(N_s)).$$

A la familia  $\{\pi^s\}_{s \in S}$  se le conoce como *especificación local* de  $X$ .

De esta manera, la característica local de la red neuronal en el sitio  $s := (s_1, s_2) \in \mathbb{Z}_4^2$  está dada como sigue de acuerdo al sistema de vecindades  $N^\alpha$

$$\begin{aligned} \pi^s(x) = \mathbb{P}(X(s) = x(s) \mid & X(s_1 + 1, s_2) = x(s_1 + 1, s_2), X(s_1 - 1, s_2) = x(s_1 - 1, s_2), \\ & X(s_1, s_2 + 1) = x(s_1, s_2 + 1), X(s_1, s_2 - 1) = x(s_1, s_2 - 1)). \end{aligned}$$

**Definición 3.6.** La distribución de probabilidad  $\pi$  en el espacio finito de configuraciones  $\Lambda^S$  con  $S = \{1, \dots, k\}$  satisface la *condición de positividad* si para todo  $i \in S$  y  $x_i \in \Lambda$  ocurre que si la distribución marginal  $\pi_i$  del sitio  $i$  es nula, entonces para todo  $y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_k \in \Lambda$ ,

$$\pi(y_1, \dots, y_{i-1}, x_i, y_{i+1}, \dots, y_k) = 0.$$

**Teorema 3.1.** Dos distribuciones de un Campo Aleatorio de Markov con espacio de configuraciones finito  $\Lambda^S$ , que satisfacen la condición de positividad y tengan la misma especificación local, son idénticas.

*Demostración.*

Sea  $X$  un Campo Aleatorio de Markov con espacio de configuraciones finito  $\Lambda^S$  tal que satisface la condición de positividad. Denótese al conjunto de sitios por  $S = \{s_1, \dots, s_k\}$ , lo cual tiene sentido puesto que  $\#(S) < \infty$  por hipótesis. De esta manera, cualquier configuración  $z \in \Lambda_S$  se puede escribir  $z = (z(s_1), \dots, z(s_k))$ . Dado lo anterior y por definición de probabilidad condicional, para  $x, y \in \Lambda^S$  arbitrarias y  $\pi$  distribución del campo,

$$\begin{aligned} \pi(x) &= \pi(x(s_1), \dots, x(s_k)) \\ &= \pi(x(s_k) \mid x(s_1), \dots, x(s_{k-1})) \cdot \pi(x(s_1), \dots, x(s_{k-1})) \\ &= \pi(x(s_k) \mid x(s_1), \dots, x(s_{k-1})) \cdot \frac{\pi(x(s_1), \dots, x(s_{k-1}), y(s_k))}{\pi(y(s_k) \mid x(s_1), \dots, x(s_{k-1}))}. \end{aligned}$$

Pero, a su vez,

$$\begin{aligned}\pi(x(s_1), \dots, x(s_{k-1}), y(s_k)) &= \pi(x(s_{k-1}) \mid x(s_1), \dots, x(s_{k-2}), y(s_k)) \cdot \\ &\quad \cdot \pi(x(s_1), \dots, x(s_{k-2}), y(s_k)) \\ &= \pi(x(s_{k-1}) \mid x(s_1), \dots, x(s_{k-2}), y(s_k)) \cdot \\ &\quad \cdot \frac{\pi(x(s_1), \dots, x(s_{k-2}), y(s_{k-1}), y(s_k))}{\pi(y(s_{k-1}) \mid x(s_1), \dots, x(s_{k-2}), y(s_k))}.\end{aligned}$$

Por lo que se tiene lo siguiente

$$\begin{aligned}\pi(x) &= \frac{\pi(x(s_k) \mid x(s_1), \dots, x(s_{k-1}))}{\pi(y(s_k) \mid x(s_1), \dots, x(s_{k-1}))} \cdot \frac{\pi(x(s_{k-1}) \mid x(s_1), \dots, x(s_{k-2}), y(s_k))}{\pi(y(s_{k-1}) \mid x(s_1), \dots, x(s_{k-2}), y(s_k))} \cdot \\ &\quad \cdot \pi(x(s_1), \dots, x(s_{k-2}), y(s_{k-1}), y(s_k)).\end{aligned}$$

Realizando este procedimiento  $k - 2$  veces más se obtiene la siguiente expresión, la cual está bien definida debido a la hipótesis de positividad,

$$\pi(x) = \prod_{i=1}^k \frac{\pi(x(s_i) \mid x(s_1), \dots, x(s_{i-1}), y(s_{i+1}), \dots, y(s_k))}{\pi(y(s_i) \mid x(s_1), \dots, x(s_{i-1}), y(s_{i+1}), \dots, y(s_k))} \cdot \pi(y).$$

Por otra parte, sean  $\pi_1$  y  $\pi_2$  dos distribuciones sobre el campo con la misma especificación local  $\{V_C\}_{C \subset S}$  tales que satisfacen la condición de positividad. Por lo anterior, se tiene que para todo  $x \in \Lambda^S$  y para  $y \in \Lambda^S$  fijo y arbitrario,

$$\begin{aligned}\frac{\pi_1(x)}{\pi_2(x)} &= \frac{\pi_1(y)}{\pi_2(y)} \cdot \left[ \prod_{i=1}^k \frac{\pi_1(x(s_i) \mid x(s_1), \dots, x(s_{i-1}), y(s_{i+1}), \dots, y(s_k))}{\pi_1(y(s_i) \mid x(s_1), \dots, x(s_{i-1}), y(s_{i+1}), \dots, y(s_k))} \right] \cdot \\ &\quad \cdot \left[ \prod_{i=1}^k \frac{\pi_2(x(s_i) \mid x(s_1), \dots, x(s_{i-1}), y(s_{i+1}), \dots, y(s_k))}{\pi_2(y(s_i) \mid x(s_1), \dots, x(s_{i-1}), y(s_{i+1}), \dots, y(s_k))} \right]^{-1}.\end{aligned}$$

Ahora bien, ambas funciones de probabilidad comparten especificaciones locales, de manera que para todo  $i \in \{1, \dots, k\}$ ,

$$\begin{aligned}\pi_1(x(s_i) \mid x(s_1), \dots, x(s_{i-1}), y(s_{i+1}), \dots, y(s_k)) \\ = \pi_2(x(s_i) \mid x(s_1), \dots, x(s_{i-1}), y(s_{i+1}), \dots, y(s_k)),\end{aligned}$$

$$\begin{aligned}
y \quad \pi_1(y(s_i) \mid x(s_1), \dots, x(s_{i-1}), y(s_{i+1}), \dots, y(s_k)) \\
= \pi_2(y(s_i) \mid x(s_1), \dots, x(s_{i-1}), y(s_{i+1}), \dots, y(s_k)).
\end{aligned}$$

De manera que para todo  $x \in \Lambda^S$  y para  $y \in \Lambda^S$  fijo y arbitrario,

$$\frac{\pi_1(x)}{\pi_2(x)} = \frac{\pi_1(y)}{\pi_2(y)},$$

lo que implica que  $\frac{\pi_1(x)}{\pi_2(x)}$  es constante. Finalmente, dado que  $\pi_1$  y  $\pi_2$  son funciones de probabilidad, forzosamente  $\frac{\pi_1(x)}{\pi_2(x)} = 1$ . Por lo tanto, son distribuciones idénticas, puesto que para toda configuración del campo  $x \in \Lambda^S$ ,  $\pi_1(x) = \pi_2(x)$ . ■

## 3.2. Campos Aleatorios de Gibbs

Dada la noción de Campos Aleatorios respecto un sistema de vecindades, se vuelve *interesante* la caracterización de las funciones de probabilidad conjunta y condicionales<sup>2</sup> sobre el mismo, lo cual es posible bajo el concepto de *Campos Aleatorios de Gibbs*.

**Definición 3.7.** Un *clique* de la gráfica  $(S, N)$  es un conjunto de la forma:

1.  $\{s\}$  con  $s \in S$ , ó
2.  $C \subset S$  con al menos dos elementos y tal que cualesquiera dos sitios en  $C$  son vecinos.

Además, se le llama *clique maximal* al *clique*  $C$  tal que para todo  $s \notin C$ ,  $C \cup \{s\}$  no es un *clique*.

Nótese que los *cliques* que induce la vecindad  $N^\alpha$  sobre el campo asociado a la red neuronal  $X$  son de la forma  $\{(s_1, s_2)\}, \{(s_1, s_2), (s_1 + 1, s_2)\} \subset \mathbb{Z}_4^2$ . Asimismo, aquéllos de la forma  $\{(s_1, s_2), (s_1 + 1, s_2)\}$  son *cliques* maximales.

**Definición 3.8.** El *potencial de Gibbs* sobre  $\Lambda^S$  relativo al sistema de vecindades  $N$  de un Campo Aleatorio es una colección  $\{V_C\}_{C \subset S}$  de funciones  $V_C : \Lambda^S \rightarrow \mathbb{R} \cup \{\infty\}$  tales que

---

<sup>2</sup>Especificaciones locales, en caso de ser Markoviano.

1. Si  $C \subset S$  no es un *clique*, entonces  $V_C \equiv 0$ .
2. Para todo  $x, x' \in \Lambda^S$  y para todo  $C \subset S$ , si  $x(C) = x'(C)$ , entonces  $V_C(x) = V_C(x')$ .

Cabe hacer notar que cada función incluida en el potencial de Gibbs depende de la configuración del campo restringida al *clique* correspondiente. De esta manera, una interpretación para el potencial de Gibbs es que explica las fases a lo largo del campo a través de las interacciones de los sitios que pertenecen a la misma vecindad.

A manera de ejemplo ilustrativo, considérese el modelo de redes neuronales artificiales. Inspiradas en las biológicas, en las redes neuronales artificiales, las neuronas hacen sinapsis con cierta fuerza  $w_{s,t}$ . Si  $w_{s,t} > 0$ , el traspaso de información entre las dos neuronas es exitoso, por lo que se dice que la sinapsis fue excitatoria; en el caso contrario, se dice que fue inhibitoria. Por lo que, una manera de caracterizar las interacciones de las neuronas en la red, pensando en ella como un Campo Aleatorio de Gibbs, es a través de la definición siguiente del potencial:

$$V_C(x) = \begin{cases} (w_{s,t} + w_{t,s}) \cdot x(s) \cdot x(t) & \text{si } C \ni \{s, t\}, \\ -h_s \cdot x(s) & \text{si } C \ni \{s\}; \end{cases}$$

donde  $h_s > 0$  corresponde al umbral de la neurona  $s$ <sup>3</sup>.

**Definición 3.9.** La función de energía  $\varepsilon : \Lambda^S \rightarrow \mathbb{R} \cup \{\infty\}$  que deriva del potencial de Gibbs  $\{V_C\}_{C \subset S}$  está dada por

$$\varepsilon(x) = \sum_C V_C(x).$$

**Definición 3.10.** Un Campo Aleatorio  $X$  con espacio de configuraciones  $\Lambda^S$  es un *Campo Aleatorio de Gibbs* respecto al sistema de vecindades  $N$  si sus configuraciones siguen una función de probabilidad  $\pi_T$  de la forma

$$\pi_T(x) = \frac{1}{Z_T} \exp\left\{-\frac{1}{T} \varepsilon(x)\right\};$$

---

<sup>3</sup>Valor mínimo que debería tomar la fuerza de la sinapsis excitatoria entre  $s$  y otra neurona para un traspaso de información exitoso.

donde  $T > 0$  es la temperatura,  $\varepsilon(x)$  la energía de la configuración  $x$  y  $Z_T$  una constante de normalización (llamada *función partición*), es decir,

$$Z_T = \sum_{\bar{\lambda} \in \Lambda^S} \exp\left\{-\frac{1}{T} \varepsilon(\bar{\lambda})\right\}.$$

A tal función de probabilidad se le denomina *distribución de Gibbs*.

La función de energía asociada al potencial de Gibbs de la red neuronal  $X$  está dada por

$$\varepsilon(x) = \sum_{s \in \mathbb{Z}_4^2} \sum_{t \in N_s^\alpha} (w_{s,t} + w_{t,s}) \cdot x(s) \cdot x(t) - \sum_{s \in \mathbb{Z}_4^2} h_s \cdot x(s).$$

Por lo que su distribución de Gibbs puede escribirse como sigue

$$\pi(x) = \frac{1}{Z} \exp\left\{-\sum_{s \in \mathbb{Z}_4^2} \sum_{t \in N_s^\alpha} (w_{s,t} + w_{t,s}) \cdot x(s) \cdot x(t) - \sum_{s \in \mathbb{Z}_4^2} h_s \cdot x(s)\right\}; \text{ donde}$$

$$Z = \sum_{\bar{\lambda} \in \{0,1\}^{\mathbb{Z}_4^2}} \exp\left\{-\sum_{s \in \mathbb{Z}_4^2} \sum_{t \in N_s^\alpha} (w_{s,t} + w_{t,s}) \cdot \bar{\lambda}(s) \cdot \bar{\lambda}(t) - \sum_{s \in \mathbb{Z}_4^2} h_s \cdot \bar{\lambda}(s)\right\}.$$

**Teorema 3.2.** Si  $X$  es un Campo Aleatorio con distribución de Gibbs  $\pi$ ; donde  $\varepsilon(x)$  deriva del potencial de Gibbs  $\{V_C\}_{C \subset S}$  relativo a un sistema de vecindades  $N$ , entonces  $X$  es Markoviano respecto a  $N$ . Es decir, los Campos Aleatorios de Gibbs son Campos Aleatorios de Markov. Más aún,

$$\pi^s(x) = \frac{\exp\{-\sum_{C \ni s} V_C(x)\}}{\sum_{\lambda \in \Lambda} \exp\{-\sum_{C \ni s} V_C(\lambda, x(S \setminus s))\}}.$$

*Demostración.*

Sea  $X$  un Campo Aleatorio de Gibbs con espacio de configuraciones  $\Lambda^S$  y potencial  $\{V_C\}_{C \subset S}$  y sea  $s \in S$ . Por definición de probabilidad condicional,

$$\mathbb{P}(X(s) = x(s) \mid X(S \setminus s) = x(S \setminus s)) = \frac{\mathbb{P}(X(s) = x(s), X(S \setminus s) = x(S \setminus s))}{\mathbb{P}(X(S \setminus s) = x(S \setminus s))} = \frac{\pi(x)}{\sum_{\lambda \in \Lambda} \pi(\lambda, x(S \setminus s))}.$$

Luego, nótese que

$$\pi(x) = \frac{1}{Z} \exp\left\{-\sum_{C \subset S} V_C(x)\right\} = \frac{1}{Z} \exp\left\{-\sum_{C \ni s} V_C(x) - \sum_{C \not\ni s} V_C(x)\right\}.$$

Asimismo, para  $\lambda \in \Lambda$  ocurre,

$$\begin{aligned} \pi(\lambda, x(S \setminus s)) &= \frac{1}{Z} \exp\left\{-\sum_{C \subset S} V_C(\lambda, x(S \setminus s))\right\} \\ &= \frac{1}{Z} \exp\left\{-\sum_{C \ni s} V_C(\lambda, x(S \setminus s)) - \sum_{C \not\ni s} V_C(\lambda, x(S \setminus s))\right\}. \end{aligned}$$

Cabe hacer notar que para todos los *cliques* tales que  $s \notin C$  en la expresión anterior,  $V_C(\lambda, x(S \setminus s)) = V_C(x)$ . De esta manera,

$$\mathbb{P}(X(s) = x(s) \mid X(S \setminus s) = x(S \setminus s)) = \frac{\exp\{-\sum_{C \ni s} V_C(x)\}}{\sum_{\lambda \in \Lambda} \exp\{-\sum_{C \ni s} V_C(\lambda, x(S \setminus s))\}}.$$

Ahora bien, la expresión anterior está en términos meramente de  $X(s)$  y  $X(N_s)$  puesto que involucra únicamente a las configuraciones del campo restringidas a aquellos *cliques* de los que  $s$  es elemento. Por lo tanto, puede considerarse como una función  $g$  de  $x(s)$  y  $x(N_s)$ , es decir,

$$\mathbb{P}(X(s) = x(s) \mid X(S \setminus s) = x(S \setminus s)) = g(x(s), x(N_s)).$$

Por otro lado,

$$\begin{aligned} \mathbb{P}(X(s) = x(s), X(N_s) = x(N_s)) &= \sum_{\bar{\lambda} \in \Lambda^{S \setminus \tilde{N}_s}} \mathbb{P}(X(s) = x(s), X(N_s) = x(N_s), X(S \setminus \tilde{N}_s) = \bar{\lambda}) \\ &= \sum_{\bar{\lambda} \in \Lambda^{S \setminus \tilde{N}_s}} \mathbb{P}(X(s) = x(s) \mid X(N_s) = x(N_s), X(S \setminus \tilde{N}_s) = \bar{\lambda}) \cdot \\ &\quad \cdot \mathbb{P}(X(N_s) = x(N_s), X(S \setminus \tilde{N}_s) = \bar{\lambda}) \\ &= \sum_{\bar{\lambda} \in \Lambda^{S \setminus \tilde{N}_s}} \mathbb{P}(X(s) = x(s) \mid X(S \setminus s) = x(S \setminus s)) \cdot \\ &\quad \cdot \mathbb{P}(X(N_s) = x(N_s), X(S \setminus \tilde{N}_s) = \bar{\lambda}) \\ &= g(x(s), x(N_s)) \cdot \sum_{\bar{\lambda} \in \Lambda^{S \setminus \tilde{N}_s}} \mathbb{P}(X(N_s) = x(N_s), X(S \setminus \tilde{N}_s) = \bar{\lambda}) \\ &= g(x(s), x(N_s)) \cdot \mathbb{P}(X(N_s) = x(N_s)). \end{aligned}$$

Utilizando la definición de probabilidad condicional,

$$\mathbb{P}(X(s) = x(s) \mid X(N_s) = x(N_s)) = g(x(s), x(N_s)).$$

Consecuentemente,

$$\mathbb{P}(X(s) = x(s) \mid X(N_s) = x(N_s)) = \mathbb{P}(X(s) = x(s) \mid X(S \setminus s) = x(S \setminus s)).$$

De donde se sigue por la Definición 3.4 que el Campo Aleatorio  $X$  es Markoviano y, además,

$$\pi^s(x) = \frac{\exp\{-\sum_{C \ni s} V_C(x)\}}{\sum_{\lambda \in \Lambda} \exp\{-\sum_{C \ni s} V_C(\lambda, x(S \setminus s))\}}.$$

■

Habiendo definido a la red neuronal  $X$  como un Campo de Gibbs, el teorema anterior implica que la red es un Campo de Markov. Más aún, para todo sitio es posible expresar de manera analítica su característica local, lo cual se llevará a cabo a continuación. Primero, considérese el sitio  $s = (s_1, s_2) \in \mathbb{Z}_4^2$ . Recuérdese que la construcción del campo  $X$  fue hecha con vecindades tipo  $\alpha$  y, así, los *cliques* sobre el campo de los que  $s$  es elemento son

$$\{(s_1, s_2)\}, \{(s_1, s_2), (s_1 + 1, s_2)\} \text{ y } \{(s_1, s_2), (s_1 - 1, s_2)\}.$$

Utilizando la notación  $s' = (s_1 + 1, s_2)$  y  $s'' = (s_1 - 1, s_2)$ , se tiene

$$\begin{aligned} \sum_{C \ni s} V_C(x) &= V_{\{s\}}(x) + V_{\{s, s'\}}(x) + V_{\{s, s''\}}(x) \\ &= -h_s \cdot x(s) + (w_{ss'} + w_{s's}) \cdot x(s) \cdot x(s') + (w_{ss''} + w_{s''s}) \cdot x(s) \cdot x(s''). \end{aligned}$$

De manera que

$$\begin{aligned} \sum_{C \ni s} V_C(0, x(S \setminus s)) &= -h_s \cdot 0 + (w_{ss'} + w_{s's}) \cdot 0 \cdot x(s') + (w_{ss''} + w_{s''s}) \cdot 0 \cdot x(s'') = 0, \text{ y} \\ \sum_{C \ni s} V_C(1, x(S \setminus s)) &= -h_s + (w_{ss'} + w_{s's}) \cdot x(s') + (w_{ss''} + w_{s''s}) \cdot x(s''). \end{aligned}$$

Luego, por el Teorema 3.2, característica local de la red neuronal  $X$  está dada por

$$\begin{aligned}\pi^s(x) &= \frac{\exp\{-\sum_{C \ni s} V_C(x)\}}{\sum_{\lambda \in \{0,1\}} \exp\{-\sum_{C \ni s} V_C(\lambda, x(S \setminus s))\}} \\ &= \frac{\exp\{x(s) \cdot [h_s - (w_{ss'} + w_{s's}) \cdot x(s') - (w_{ss''} + w_{s''s}) \cdot x(s'')]\}}{1 + \exp\{h_s - (w_{ss'} + w_{s's}) \cdot x(s') - (w_{ss''} + w_{s''s}) \cdot x(s'')\}}.\end{aligned}$$

**Teorema 3.3. (Hammersley-Clifford)** Sea  $\pi$  la distribución de un Campo Aleatorio de Markov respecto a la topología  $(S, N)$  tal que satisface la condición de positividad. Entonces, para alguna función de energía  $\varepsilon$  derivada de un potencial de Gibbs  $\{V_C\}_{C \subset S}$  asociado a  $(S, N)$ ,

$$\pi(x) = \frac{1}{Z} \exp\{-\varepsilon(x)\}.$$

Es decir, los Campos Aleatorios de Markov (cuya distribución cumple la condición de positividad) son Campos Aleatorios de Gibbs<sup>4</sup>.

### 3.3. Estimación paramétrica

En este apartado se llevará a cabo el desarrollo teórico de la estimación paramétrica sobre Campos Aleatorios por Máxima Verosimilitud, la cual será aplicada para el cumplimiento de los objetivos de la investigación. Considérese un Campo Aleatorio  $X$  sobre  $S$ , con fases en  $\Lambda$  y tal que su espacio de configuraciones sea finito. Supóngase que  $X$  es un Campo de Gibbs (y, por lo tanto, un Campo de Markov) respecto al sistema de vecindades  $N$ , tal que éste genera  $k$  *cliques* de formas diferentes<sup>5</sup>. Entonces, en general, puede denotarse al potencial de Gibbs del campo como sigue

$$\{V_C\}_{C \subset S} = \{V_{C_1}\}_{C_1 \subset S} \cup \dots \cup \{V_{C_k}\}_{C_k \subset S};$$

donde  $\{V_{C_j}\}_{C_j \subset S}$ ,  $j \in \{1, \dots, k\}$ , corresponde al potencial de Gibbs sobre el campo restringido a los *cliques* de la  $j$ -ésima forma generada por  $N$ .

<sup>4</sup>La demostración del teorema es consultable en [6, p. 262].

<sup>5</sup>Por ejemplo, recuérdese que la vecindad tipo  $\alpha$  induce *cliques* de un solo tipo además de los conjuntos unitarios, a saber, aquéllos de la forma  $\{(s_1, s_2), (s_1 + 1, s_2)\}$ ; para  $(s_1, s_2) \in S$ .

Luego, el Campo Aleatorio  $X$  estará caracterizado mediante su distribución de Gibbs en forma paramétrica

$$\pi(x; \underline{\theta}) = \frac{1}{Z(\underline{\theta})} \cdot e^{-\langle \underline{\theta}, \underline{V}_C(x) \rangle} = \frac{1}{Z(\underline{\theta})} \cdot \exp \left\{ - \sum_{j=1}^k \left[ \theta_j \cdot \sum_{C_j \subset S} V_{C_j}(x) \right] \right\};$$

donde  $\underline{\theta} = (\theta_1, \dots, \theta_k) \in (\mathbb{R}^+)^k$  corresponde al vector de parámetros del modelo y

$$\underline{V}_C(x) = \left( \sum_{C_1 \subset S} V_{C_1}(x), \dots, \sum_{C_k \subset S} V_{C_k}(x) \right) \in \mathbb{R}^k,$$

de manera que  $\varepsilon = \langle \underline{\theta}, \underline{V}_C(x) \rangle$  es la energía que deriva del potencial de Gibbs.

Supóngase que se cuenta con una serie de observaciones del campo  $X_1, \dots, X_n$  que se denotará por  $\underline{X}$ . Se determinará el Estimador Máximo Verosímil de  $\underline{\theta}$  suponiendo, en primer lugar, que las observaciones son independientes y, luego, bajo la restricción de no independencia con las  $m - 1$  observaciones inmediatas anteriores –es decir, considerando el campo como un *Modelo de Markov de orden  $m - 1$* –.

### 3.3.1. Estimación para observaciones independientes

Suponiendo observaciones del campo independientes, la función de Verosimilitud de  $\underline{X}$  estará dada por

$$\begin{aligned} \mathcal{L}(\underline{X}; \underline{\theta}) &= \prod_{i=1}^n \pi(x_i; \underline{\theta}) = \prod_{i=1}^n \left( \exp \left\{ - \sum_{j=1}^k \left[ \theta_j \cdot \sum_{C_j \subset S} V_{C_j}(x_i) \right] \right\} \cdot \frac{1}{Z(\underline{\theta})} \right) \\ &= Z(\underline{\theta})^{-n} \exp \left\{ - \sum_{i=1}^n \sum_{j=1}^k \left[ \theta_j \cdot \sum_{C_j \subset S} V_{C_j}(x_i) \right] \right\}. \end{aligned}$$

Por lo que su función de log-Verosimilitud está dada como sigue

$$\ell(\underline{X}; \underline{\theta}) = -n \ln Z(\underline{\theta}) - \sum_{i=1}^n \sum_{j=1}^k \left[ \theta_j \cdot \sum_{C_j \subset S} V_{C_j}(x_i) \right].$$

Nótese que ésta es diferenciable respecto  $\underline{\theta}$  puesto que es composición de sumas, productos por constantes y la función exponencial. Ahora bien, para encontrar el Estimador Máximo Verosímil

de  $\underline{\theta}$ , es necesario encontrar sus derivadas parciales. Primero,

$$\frac{\partial Z}{\partial \theta_l} = \frac{\partial}{\partial \theta_l} \left[ \sum_{\bar{\lambda} \in \Lambda^s} e^{-\langle \underline{\theta}, \underline{V_C}(\bar{\lambda}) \rangle} \right] = \sum_{\bar{\lambda} \in \Lambda^s} \left[ e^{-\langle \underline{\theta}, \underline{V_C}(\bar{\lambda}) \rangle} \cdot (-1) \cdot \sum_{C_l \subset S} V_{C_l}(\bar{\lambda}) \right].$$

De esta manera,

$$\begin{aligned} \frac{\partial}{\partial \theta_l} \left[ -n \ln Z(\underline{\theta}) \right] &= \frac{n}{Z(\underline{\theta})} \cdot \sum_{\bar{\lambda} \in \Lambda^s} \left[ e^{-\langle \underline{\theta}, \underline{V_C}(\bar{\lambda}) \rangle} \cdot \sum_{C_l \subset S} V_{C_l}(\bar{\lambda}) \right] = n \cdot \sum_{\bar{\lambda} \in \Lambda^s} \left[ \left( \sum_{C_l \subset S} V_{C_l}(\bar{\lambda}) \right) \cdot \frac{e^{-\langle \underline{\theta}, \underline{V_C}(\bar{\lambda}) \rangle}}{Z(\underline{\theta})} \right] \\ &= n \cdot \mathbb{E} \left[ \sum_{C_l \subset S} V_{C_l}(X) \mid \underline{\theta} \right], \end{aligned}$$

puesto que  $\mathbb{E}[g(X)] = \sum_x g(x) \cdot \mathbb{P}(X = x)$ .

Posteriormente,

$$\frac{\partial}{\partial \theta_l} \left[ \sum_{i=1}^n \sum_{j=1}^k \left( \theta_j \cdot \sum_{C_j \subset S} V_{C_j}(x_i) \right) \right] = \sum_{i=1}^n \sum_{C_l \subset S} V_{C_l}(x_i).$$

Por lo que se tiene el siguiente resultado

$$\frac{\partial \ell}{\partial \theta_l} = n \cdot \mathbb{E} \left[ \sum_{C_l \subset S} V_{C_l}(X) \mid \underline{\theta} \right] - \sum_{i=1}^n \sum_{C_l \subset S} V_{C_l}(x_i).$$

Igualando a cero la ecuación anterior para encontrar el punto crítico de la función de log-Verosimilitud, se obtiene que el Estimador Máximo Verosímil de  $\underline{\theta}$  es aquel que satisface para cada  $l \in \{1, \dots, k\}$  lo siguiente

$$\frac{1}{n} \sum_{i=1}^n \left( \sum_{C_l \subset S} V_{C_l}(x_i) \right) = \mathbb{E} \left[ \sum_{C_l \subset S} V_{C_l}(X) \mid \underline{\theta} \right],$$

es decir, que la esperanza de la suma de los potenciales restringidos a cada tipo de *clique* sea igual a su media empírica. Nótese que tal estimador debe ser aproximado utilizando métodos numéricos.

Para corroborar que el punto crítico encontrado es máximo, primero se obtiene lo siguiente, con base en la derivada parcial  $\frac{\partial Z}{\partial \theta_l}$  encontrada anteriormente

$$\begin{aligned}
\frac{\partial}{\partial \theta_l} \left( \frac{1}{Z(\underline{\theta})} \right) &= -\frac{1}{Z(\underline{\theta})^2} \cdot \frac{\partial Z}{\partial \theta_l} = -\frac{1}{Z(\underline{\theta})^2} \cdot \sum_{\bar{\lambda} \in \Lambda^s} \left[ e^{-\langle \underline{\theta}, V_C(\bar{\lambda}) \rangle} \cdot (-1) \cdot \sum_{C_l \subset S} V_{C_l}(\bar{\lambda}) \right] \\
&= \frac{1}{Z(\underline{\theta})} \cdot \sum_{\bar{\lambda} \in \Lambda^s} \left[ \left( \sum_{C_l \subset S} V_{C_l}(\bar{\lambda}) \right) \cdot \frac{e^{-\langle \underline{\theta}, V_C(\bar{\lambda}) \rangle}}{Z(\underline{\theta})} \right] = \frac{1}{Z(\underline{\theta})} \cdot \mathbb{E} \left[ \sum_{C_l \subset S} V_{C_l}(X) \mid \underline{\theta} \right].
\end{aligned}$$

Luego, es posible obtener la segunda derivada parcial de la función de log-Verosimilitud, que también es diferenciable respecto de  $\underline{\theta}$ , para  $\theta_l, l \in \{1, \dots, k\}$  como sigue

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \theta_l^2} &= \frac{\partial}{\partial \theta_l} \left( n \cdot \mathbb{E} \left[ \sum_{C_l \subset S} V_{C_l}(X) \mid \underline{\theta} \right] - \sum_{i=1}^n \sum_{C_l \subset S} V_{C_l}(x_i) \right) \\
&= n \cdot \frac{\partial}{\partial \theta_l} \left( \sum_{\bar{\lambda} \in \Lambda^s} \left[ \left( \sum_{C_l \subset S} V_{C_l}(\bar{\lambda}) \right) \cdot \frac{e^{-\langle \underline{\theta}, V_C(\bar{\lambda}) \rangle}}{Z(\underline{\theta})} \right] \right) \\
&= n \cdot \sum_{\bar{\lambda} \in \Lambda^s} \left[ \left( \sum_{C_l \subset S} V_{C_l}(\bar{\lambda}) \right) \cdot \left( \frac{e^{-\langle \underline{\theta}, V_C(\bar{\lambda}) \rangle}}{Z(\underline{\theta})} \cdot \mathbb{E} \left[ \sum_{C_l \subset S} V_{C_l}(X) \mid \underline{\theta} \right] - \frac{e^{-\langle \underline{\theta}, V_C(\bar{\lambda}) \rangle}}{Z(\underline{\theta})} \cdot \sum_{C_l \subset S} V_{C_l}(\bar{\lambda}) \right) \right] \\
&= n \cdot \mathbb{E} \left[ \sum_{C_l \subset S} V_{C_l}(X) \mid \underline{\theta} \right] \cdot \sum_{\bar{\lambda} \in \Lambda^s} \left[ \left( \sum_{C_l \subset S} V_{C_l}(\bar{\lambda}) \right) \cdot \frac{e^{-\langle \underline{\theta}, V_C(\bar{\lambda}) \rangle}}{Z(\underline{\theta})} \right] + \\
&\quad - n \cdot \sum_{\bar{\lambda} \in \Lambda^s} \left[ \left( \sum_{C_l \subset S} V_{C_l}(\bar{\lambda}) \right)^2 \cdot \frac{e^{-\langle \underline{\theta}, V_C(\bar{\lambda}) \rangle}}{Z(\underline{\theta})} \right] \\
&= n \cdot \mathbb{E} \left[ \sum_{C_l \subset S} V_{C_l}(X) \mid \underline{\theta} \right]^2 - n \cdot \mathbb{E} \left[ \left( \sum_{C_l \subset S} V_{C_l}(X) \right)^2 \mid \underline{\theta} \right].
\end{aligned}$$

Por lo tanto,

$$\frac{\partial^2 \ell}{\partial \theta_l^2} = -n \cdot \text{Var} \left[ \sum_{C_l \subset S} V_{C_l}(X) \mid \underline{\theta} \right] < 0,$$

siempre y cuando el potencial  $\{V_{C_j}\}_{C_j \subset S}$  no esté conformado por funciones constantes únicamente<sup>6</sup>. Bajo esta condición, se cumple que el punto crítico encontrado es máximo.

---

<sup>6</sup>Nótese que de ser así, además de ser poco *interesante*, la realización cualquiera de las diferentes configuraciones sería equiprobable.

### 3.3.2. Estimación para observaciones secuenciales

Supóngase que  $X_1, \dots, X_n$  corresponde a una secuencia observada que es realización de un Modelo de Markov de orden  $m - 1$  (con  $m \in \{2, \dots, n\}$ ), es decir, tal que para  $j \in \{m, \dots, n\}$ ,

$$\begin{aligned} \mathbb{P}(X_j = x_j \mid X_{j-1} = x_{j-1}, \dots, X_1 = x_1) \\ = \mathbb{P}(X_j = x_j \mid X_{j-1} = x_{j-1}, \dots, X_{j-m+1} = x_{j-m+1}). \end{aligned}$$

Sea  $X$  un Campo de Gibbs con sitios en  $\{1, \dots, m\}$  y fases en  $\Lambda$  finito, respecto al sistema de vecindades  $N$  que genera  $k$  cliques de formas diferentes. Nótese que se puede “acomodar” la secuencia observada en configuraciones sucesivas de la forma

$$(X_1, \dots, X_m), (X_2, \dots, X_{m+1}), \dots, (X_{n-m+1}, \dots, X_n);$$

de manera que puede modelarse la dependencia de orden  $m - 1$  con Campos Aleatorios de Gibbs o Markov. Con el fin de reducir la notación, expresiones de la forma  $\mathbb{P}(X_{n_1} = x_{n_1}, \dots, X_{n_r} = x_{n_r})$  se denotarán por  $\mathbb{P}(x_{n_1}, \dots, x_{n_r})$ ; donde  $r, n_i \in \{1, \dots, n\}$  con  $i \in \{1, \dots, r\}$ .

La función de Verosimilitud de la secuencia está dada como sigue. Utilizando sucesivamente la definición de probabilidad condicional y que  $X$  es un modelo Markoviano de orden  $m - 1$ ,

$$\begin{aligned} \mathcal{L}(\underline{X}; \underline{\theta}) &= \mathbb{P}(x_1, \dots, x_n; \underline{\theta}) = \mathbb{P}(x_n \mid x_{n-1}, \dots, x_1, \underline{\theta}) \cdot \mathbb{P}(x_{n-1}, \dots, x_1, \underline{\theta}) \\ &= \mathbb{P}(x_n \mid x_{n-1}, \dots, x_{n-m+1}, \underline{\theta}) \cdot \mathbb{P}(x_{n-1} \mid x_{n-2}, \dots, x_1, \underline{\theta}) \cdot \mathbb{P}(x_{n-2}, \dots, x_1, \underline{\theta}) \\ &= \mathbb{P}(x_n \mid x_{n-1}, \dots, x_{n-m+1}, \underline{\theta}) \cdot \mathbb{P}(x_{n-1} \mid x_{n-2}, \dots, x_{n-m}, \underline{\theta}) \cdot \mathbb{P}(x_{n-2}, \dots, x_1, \underline{\theta}) \\ &\quad \vdots \\ &= \left[ \prod_{i=m}^n \mathbb{P}(x_i \mid x_{i-1}, \dots, x_{i-m+1}, \underline{\theta}) \right] \cdot \mathbb{P}(x_{m-1}, \dots, x_1; \underline{\theta}). \end{aligned}$$

Luego, en términos de la distribución de Gibbs de  $X$ , para  $i \in \{m, \dots, n\}$

$$\begin{aligned} \mathbb{P}(x_{i-1}, \dots, x_{i-m+1}; \underline{\theta}) &= \sum_{\lambda \in \Lambda} \pi(\lambda, x_{i-1}, \dots, x_{i-m+1}; \underline{\theta}) \\ &= \sum_{\lambda \in \Lambda} \frac{1}{Z(\underline{\theta})} \exp\{-\langle \underline{\theta}, \underline{V}_C(\lambda, x_{i-1}, \dots, x_{i-m+1}) \rangle\}. \end{aligned}$$

Y, de manera análoga,

$$\begin{aligned}\mathbb{P}(x_i \mid x_{i-1}, \dots, x_{i-m+1}, \underline{\theta}) &= \frac{\pi(x_i, \dots, x_{i-m+1}; \underline{\theta})}{\sum_{\lambda \in \Lambda} \pi(\lambda, x_{i-1}, \dots, x_{i-m+1}; \underline{\theta})} \\ &= \frac{\exp\{-\langle \underline{\theta}, \underline{V}_C(x_i, \dots, x_{i-m+1}) \rangle\}}{\sum_{\lambda \in \Lambda} \exp\{-\langle \underline{\theta}, \underline{V}_C(\lambda, x_{i-1}, \dots, x_{i-m+1}) \rangle\}}.\end{aligned}$$

Con lo que se obtiene que

$$\begin{aligned}\mathcal{L}(\underline{X}; \underline{\theta}) &= \left[ \prod_{i=m}^n \frac{\exp\{-\langle \underline{\theta}, \underline{V}_C(x_i, \dots, x_{i-m+1}) \rangle\}}{\sum_{\lambda \in \Lambda} \exp\{-\langle \underline{\theta}, \underline{V}_C(\lambda, x_{i-1}, \dots, x_{i-m+1}) \rangle\}} \right] \\ &\quad \cdot \left[ \sum_{\lambda \in \Lambda} \frac{1}{Z(\underline{\theta})} \exp\{-\langle \underline{\theta}, \underline{V}_C(\lambda, x_{m-1}, \dots, x_1) \rangle\} \right].\end{aligned}$$

Por lo que la función de log-Verosimilitud de la secuencia, que puede optimizarse con técnicas de Análisis Numérico, está dada como sigue

$$\begin{aligned}\ell(\underline{X}; \underline{\theta}) &= - \sum_{i=m}^n \langle \underline{\theta}, \underline{V}_C(x_i, \dots, x_{i-m+1}) \rangle - \sum_{i=m}^n \left[ \ln \left( \sum_{\lambda \in \Lambda} \exp\{-\langle \underline{\theta}, \underline{V}_C(\lambda, x_{i-1}, \dots, x_{i-m+1}) \rangle\} \right) \right] + \\ &\quad - \ln Z(\underline{\theta}) + \ln \left( \sum_{\lambda \in \Lambda} \exp\{-\langle \underline{\theta}, \underline{V}_C(\lambda, x_{m-1}, \dots, x_1) \rangle\} \right).\end{aligned}$$

## CAPÍTULO 4

# APRENDIZAJE ESTADÍSTICO

Dado que uno de los principales objetivos de la presente investigación es la clasificación, esta sección está dedicada a la exposición de las herramientas de *Aprendizaje Estadístico* que se utilizarán para lograr dicha tarea.

En primera instancia, se conoce como Aprendizaje Estadístico al conjunto de herramientas computacionales y estadísticas utilizadas con el objetivo de predecir o describir cierto fenómeno observado con base en sus atributos. En general, se asume que existe cierta relación entre una variable respuesta  $Y$  y  $p$  predictores  $X_1, \dots, X_p$ , es decir,

$$Y = f(X) + \varepsilon;$$

donde  $\varepsilon$  corresponde a cierto error aleatorio de media cero e independiente de los atributos observados. Así, el objetivo principal del Aprendizaje Estadístico es determinar  $f$ , una función desconocida que *representa la información sistemática que X provee sobre Y*<sup>1</sup> [19, p. 16].

Tales herramientas han sido desarrolladas bajo dos enfoques: *Aprendizaje Supervisado* y *no Supervisado*. El primero hace alusión a las situaciones en que se cuenta con variables respuesta (salida) que son determinadas por los atributos (variables de entrada). Mientras que *en un problema de Aprendizaje no Supervisado se observan solamente los atributos y no se tienen mediciones de la salida. La tarea es describir cómo está organizada o aglomerada la información*<sup>2</sup> [16, p. 2].

<sup>1</sup>“ $f$  represents the systematic information that X provides about Y”; [19, p. 16].

<sup>2</sup>“In the unsupervised learning problem, we observe only the features and have no measurements of the outcome.

Cabe destacar que, hoy en día, la cantidad de recursos a los cuales recurrir para la clasificación son bastantes y su desarrollo sigue a la alza. De manera que únicamente se hará uso de algunos de ellos: *Regresión Multinomial*, *Regresión Ordinal*, *Análisis de Discriminante Lineal*, *Clasificador  $k$ -Vecinos más Cercanos*, *Árboles de Clasificación*, *Análisis de Componentes Principales* y *Análisis de Conglomerados*. Asimismo, dado que la bibliografía al respecto es vasta y se utilizarán meramente como herramienta para la concreción de la clasificación, en esta sección se llevará a cabo una brevísima descripción de los distintos métodos a utilizar. La teoría al respecto es ampliamente desarrollada en [1], [16], [17], [19], y [34] –por mencionar algunos–, los cuales fueron tomados como referencia principal de la exposición siguiente.

Finalmente, se considerará una muestra aleatoria de tamaño  $n$  de realizaciones de una variable respuesta categórica (clase) denotada por  $Y$  que toma valores representados por el conjunto  $\{1, \dots, k\}$ ; y de  $p$  variables predictoras (atributos) independientes denotadas por  $X = (X_1, \dots, X_p)$ .

## 4.1. Regresión Multinomial

Dadas observaciones con una variable de salida  $Y$  y  $X$  una matriz de predictores, los *Modelos Lineales Generalizados describen la variables respuesta utilizando un miembro de la familia exponencial*<sup>3</sup> [17, p. 30]. Con los GLM –*Generalised Linear Models*, en inglés– se busca aproximar una transformación  $g$  de la esperanza condicional  $\mathbb{E}(Y | X = x)$  a través de un modelo lineal de predictores, i.e., son de la forma

$$g[\mathbb{E}(Y | X = x)] = \beta_0 + x^t \beta;$$

donde  $\beta_0 \in \mathbb{R}$  y  $\beta \in \mathbb{R}^p$  corresponden a los parámetros del modelo. Tal tipo de transformaciones son denominadas *link functions* (funciones liga) y son estrictamente monótonas.

Ahora bien, Regresión Multinomial es un caso particular de los GLM en el que la variable respuesta  $Y$  toma valores cualitativos tipificados en  $k$  clases. Bajo este esquema, el modelo a ajustar

---

Our task is rather to describe how the data are organized or clustered”; [16, p. 2].

<sup>3</sup>“(These models) describe the response variable using a member of the exponential family”; [17, p. 30].

se puede escribir como sigue

$$\ln \frac{\mathbb{P}(Y = i | X = x)}{\mathbb{P}(Y = k | X = x)} = \beta_{0i} + x^t \beta_i;$$

para cada una de las clases  $i \in \{1, \dots, k-1\}$ . De manera que, al realizar la transformación inversa, es equivalente a ajustar el siguiente modelo para cada valor que tome  $Y$  excepto uno –denominado *clase base* y que puede escogerse indistintamente de todas las categorías–, en este caso,  $k$

$$\mathbb{P}(Y = i | X = x) = \frac{\exp\{\beta_{0i} + x^t \beta_i\}}{1 + \sum_{m=1}^{k-1} \exp\{\beta_{0m} + x^t \beta_m\}}.$$

La justificación de lo anterior subyace en el hecho de que al aplicar la función exponencial a la expresión del modelo, se obtiene que para cada  $i \in \{1, \dots, k-1\}$ ,

$$\mathbb{P}(Y = i | X = x) = \mathbb{P}(Y = k | X = x) \cdot \exp\{\beta_{0i} + x^t \beta_i\}.$$

Luego, nótese lo siguiente

$$\begin{aligned} \mathbb{P}(Y = k | X = x) &= 1 - \sum_{m=1}^{k-1} \mathbb{P}(Y = m | X = x) \\ &= 1 - \mathbb{P}(Y = k | X = x) \cdot \sum_{m=1}^{k-1} \exp\{\beta_{0m} + x^t \beta_m\}, \end{aligned}$$

a lo que es equivalente la ecuación

$$\mathbb{P}(Y = k | X = x) = \left[ 1 + \sum_{m=1}^{k-1} \exp\{\beta_{0m} + x^t \beta_m\} \right]^{-1}.$$

Cabe hacer notar que, pese a que lo natural sería ajustar un modelo de la forma

$$\mathbb{P}(Y = i | X = x) = \frac{\exp\{\beta_{0i} + x^t \beta_i\}}{\sum_{m=1}^k \exp\{\beta_{0m} + x^t \beta_m\}},$$

para cada  $i \in \{1, \dots, k\}$ , se estaría sobreespecificando el modelo puesto que  $\mathbb{P}(Y = k | X = x)$  corresponde al complemento de la probabilidad de pertenecer a las  $k-1$  clases restantes<sup>4</sup>.

---

<sup>4</sup>Es posible llevarlo a cabo imponiendo las restricciones adecuadas computacionalmente, como es logrado mediante Redes Neuronales.

Posteriormente, para una muestra aleatoria de tamaño  $n$ , la función de log-Verosimilitud de las observaciones puede expresarse como sigue

$$\begin{aligned}\ell(Y; X, \{\beta_{0i}, \beta_i\}_{i=1}^k) &= \sum_{m=1}^n \ln \mathbb{P}(Y = y_m \mid x_m, \{\beta_{0i}, \beta_i\}_{i=1}^k) \\ &= \sum_{m=1}^n \left[ \sum_{i=1}^k \left( \mathbb{1}_{\{i\}}^{(y_m)} \cdot (\beta_{0i} + x_m^t \beta_i) \right) - \ln \left( 1 + \sum_{r=1}^{k-1} \exp\{\beta_{0r} + x_m^t \beta_r\} \right) \right].\end{aligned}$$

De lo que sigue estimar los parámetros para cada clase maximizando la expresión anterior. Comúnmente, tal optimización es llevada a cabo mediante métodos numéricos, en lugar de la obtención de la solución explícita. Adicionalmente, *el modelo ajustado por Máxima Verosimilitud es invariante bajo la categoría que se escoja como clase base y las estimaciones de los parámetros son equivariantes (la solución para una clase base puede ser obtenida con base en la solución para otra)*<sup>5</sup> [17, p. 36].

A diferencia de Regresión Lineal Múltiple, la prueba de hipótesis respecto la significancia de los parámetros del modelo

$$H_0 : \beta_i = 0 \quad \text{vs.} \quad H_1 : \beta_i \neq 0;$$

con  $i \in \{1, \dots, p\}$ , corresponde a una prueba de Wald en lugar de una prueba  $t$ . De esta manera, la estadística a considerar para las pruebas de significancia del modelo está dada por

$$Z = \frac{\hat{\beta}_i}{\widehat{se}(\hat{\beta}_i)} \sim N(0, 1),$$

bajo la hipótesis nula, y donde  $\widehat{se}(\hat{\beta}_i)$  es el error estándar estimado de  $\hat{\beta}_i$ .

---

<sup>5</sup>“The model fit by maximum-likelihood is invariant to the choice of this base class, and the parameter estimates are equivariant (the solution for one base can be obtained from the solution for another)”; [17, p. 36].

## 4.2. Regresión Ordinal

Regresión Ordinal, también llamada un *Modelo Logit Acumulativo*, es otro caso particular de los GLM en el que la variable respuesta  $Y$  es una variable ordinal, a diferencia del modelo anterior en que  $Y$  se asume nominal. De esta manera, es más adecuado ajustar un modelo de la forma

$$\ln \frac{\mathbb{P}(Y \leq j | X = x)}{1 - \mathbb{P}(Y \leq j | X = x)} = \beta_{0j} + x^t \beta + \varepsilon;$$

para cada una de las clases  $j \in \{1, \dots, k - 1\}$ .

O bien, equivalentemente al aplicar la transformación inversa, ajustar el siguiente modelo para cada valor que tome  $Y$  excepto la última categoría ( $k$ ), de manera análoga a la Regresión Multinomial

$$\mathbb{P}(Y \leq j | X = x) = \frac{\exp\{\beta_{0j} + x^t \beta\}}{1 + \exp\{\beta_{0j} + x^t \beta\}} + \varepsilon;$$

con  $j \in \{1, \dots, k - 1\}$ .

Es de hacer notar que *el modelo logit para la probabilidad acumulada  $j$  se ve como un modelo de regresión logística binaria en el que las categorías 1 a  $j$  se combinan para formar una sola categoría y las categorías  $j + 1$  a  $J$  forman una segunda categoría*<sup>7</sup> [1, p. 180].

Adicionalmente, en este contexto se hace el supuesto de momios proporcionales—*proportional odds assumption*, en inglés—, que refiere a que  $X$  tiene el mismo efecto sobre todas las probabilidades acumulativas  $\mathbb{P}(Y \leq j)$ , i.e. se estima un mismo vector de parámetros  $\beta$  para todas las probabilidades acumuladas. Asimismo, como se deriva de la expresión del modelo, para el intercepto  $\beta_{0i}$  en cada caso no se impone dicha restricción; sí puede diferir  $\beta_{0i}$  en la estimación de los odds para cada categoría de  $Y$ . Nótese que, de lo contrario, para todo  $j \in \{1, \dots, k - 1\}$ ,

$$\mathbb{P}(Y = j | X = x) = \mathbb{P}(Y = k | X = x).$$

En este caso, la Función de Verosimilitud a maximizar para la estimación de los parámetros del

---

<sup>6</sup>Categorías  $j + 1$  a  $k$ , en la notación que se está usando.

<sup>7</sup>“A model for cumulative logit  $j$  looks like a binary logistic regression model in which categories 1– $j$  combine to form a single category and categories  $j + 1$  to  $J$  form a second category”; [1, p. 180].

modelo dadas  $n$  observaciones, está dada por

$$\begin{aligned}\mathcal{L}(Y; X, \{\beta_{0m}\}_{m=1}^k, \beta) &= \prod_{i=1}^n \mathbb{P}(Y = y_i \mid x_i, \{\beta_{0m}\}_{m=1}^k, \beta) \\ &= \prod_{i=1}^n \left[ \mathbb{P}(Y \leq y_i \mid x_i, \{\beta_{0m}\}_{m=1}^k, \beta) - \mathbb{P}(Y \leq y_i - 1 \mid x_i, \{\beta_{0m}\}_{m=1}^k, \beta) \right].\end{aligned}$$

### 4.3. Métodos de Regularización

De manera alternativa a la Selección de Modelos convencional, se hace uso de este tipo de método con la finalidad de mejorar el poder predictivo de los Modelos Lineales Generalizados, o bien, hacer selección de variables para reducir su dimensionalidad –lo cual se vuelve imperativo si  $p \geq n$ –. En general, los *Métodos de Regularización* consisten en añadir un monto de penalización sobre la magnitud de los coeficientes de regresión a la función de costo a optimizar para su estimación. De manera que se hace selección de variables reduciendo su magnitud o anulando algunos de ellos, razón por la cual toman tal nombre, *Shrinkage Methods* en inglés. Los métodos de Regularización que se utilizarán como vía alternativa para el ajuste de los GLM presentados anteriormente son *Lasso Regression* y *Ridge Regression*.

Explícitamente, se busca minimizar el Error Cuadrático Medio dividido por dos aumentado a la penalización planteada, es decir, la función objetivo está dada por

$$\frac{1}{2 \cdot n} \cdot \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j \cdot x_{ij} \right)^2 + \lambda \cdot \|\underline{\beta}\|_p^p;$$

donde  $\lambda$  corresponde a un parámetro de penalización que puede ser escogido por *Cross-Validation*<sup>8</sup> y  $\|\cdot\|_p$  la norma  $p$ , con  $p \in \mathbb{R}$ . Si se fija  $p = 1$ , el método corresponde a Regresión por Lasso; mientras que si  $p = 2$ , a Regresión Ridge.

Por último, es debido hacer notar que la utilización de métodos de regularización conlleva un gran costo en relación al diagnóstico del modelo, puesto que éste sigue siendo un problema sin solución. Robert Tibshirani, uno de los autores de esta teoría, ha remarcado que aún es un reto

<sup>8</sup>Véase Sección 4.9. Validación de modelos y remuestreo.

el desarrollo de herramientas y teoría que permitan a estos métodos ser usados en la práctica estadística: errores estándar,  $p$ -values e intervalos de confianza que tomen en consideración la naturaleza adaptativa de la estimación<sup>9</sup> [32, p. 43].

Adicionalmente, puesto que se añade una penalización a la función de costo a minimizar, el Teorema de Gauss-Markov no es válido en este caso; i.e., los estimadores de los coeficientes  $\underline{\beta}$  no cuentan con la cualidad de ser insesgados. No obstante, los Métodos de Regularización suelen tener un buen poder predictivo debido a que la varianza de sus estimaciones resulta ser menor.

## 4.4. Análisis de Discriminante

El *Análisis de Discriminante* es un método de clasificación supervisada que consiste en la asignación de una observación a aquella clase  $i$ , para cuyos atributos, la probabilidad  $\mathbb{P}(Y = i | X = x)$  sea mayor en comparación al resto de las categorías. El supuesto a considerar bajo este esquema es que el vector aleatorio  $X = (X_1, \dots, X_p)$  sigue una distribución Normal Multivariada con vector de medias  $\mu_i$  (para cada categoría  $i$ ) y matrices de covarianza  $\Sigma_i$ .

### 4.4.1. Análisis de Discriminante Lineal

Bajo la construcción anterior y el supuesto de que  $\Sigma$  es igual para todas las categorías de la variable respuesta  $Y$ , el método se denomina *Análisis de Discriminante Lineal*. Dado que  $Y$  pertenece a la clase  $i$ , la función de densidad conjunta de  $X$  está dada por

$$f_i(x) = \frac{1}{(2 \cdot \pi)^{p/2} \cdot |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu_i)^t \Sigma^{-1} (x - \mu_i) \right\}.$$

Luego, por el Teorema de Bayes, las funciones de probabilidad a posteriori para  $i \in \{1, \dots, k\}$  están dadas por

$$\mathbb{P}(Y = i | X) = \frac{\pi_i \cdot f_i(x)}{\sum_{m=1}^k \pi_m \cdot f_m(x)};$$

---

<sup>9</sup>“(Some challenges:) Develop tools and theory that allow these methods to be used in statistical practice: standard errors,  $p$ -values and confidence intervals that account for the adaptive nature of the estimation”; [32, p. 43].

donde  $\pi_i$  corresponde a la función de probabilidad a priori de  $Y$ .

El clasificador asignará una observación a la clase  $i$  para la cual  $\mathbb{P}(Y = i | X)$  sea mayor o, equivalentemente, para la cual sea mayor la siguiente expresión

$$\ln(\pi_i \cdot f_i(x)) = \ln \pi_i - \frac{p}{2} \cdot \ln(2 \cdot \pi) - \frac{1}{2} \cdot |\Sigma| - \frac{1}{2} (x - \mu_i)^t \Sigma^{-1} (x - \mu_i).$$

Nótese que, utilizando propiedades de matrices, se da la siguiente ecuación

$$(x - \mu_i)^t \Sigma^{-1} (x - \mu_i) = x^t \Sigma^{-1} x - x^t \Sigma^{-1} \mu_i - \mu_i^t \Sigma^{-1} x + \mu_i^t \Sigma^{-1} \mu_i.$$

Adicionalmente,

$$x^t \Sigma^{-1} \mu_i = (\Sigma^{-1} x)^t (\mu_i)^t = \mu_i^t \Sigma^{-1} x = \mu_i^t \Sigma^{-1} x;$$

y, por lo tanto,

$$-\frac{1}{2} (x - \mu_i)^t \Sigma^{-1} (x - \mu_i) = -\frac{1}{2} x^t \Sigma^{-1} x + x^t \Sigma^{-1} \mu_i - \frac{1}{2} \mu_i^t \Sigma^{-1} \mu_i.$$

De manera que maximizar  $\mathbb{P}(Y = i | X)$  es equivalente a maximizar la expresión

$$\delta_i(x) = x^t \Sigma^{-1} \mu_i - \frac{1}{2} \mu_i^t \Sigma^{-1} \mu_i + \ln \pi_i;$$

la cual es conocida como *función discriminante* y determina las fronteras de decisión del clasificador, i.e., el conjunto de  $x \in \mathbb{R}^p$  tales que  $\delta_i(x) = \delta_j(x)$ , para  $i \neq j$ . El hecho de que sea una función lineal en términos de  $x$  es motivo del nombre *Discriminante Lineal*.

Posteriormente, para cada  $i \in \{1, \dots, k\}$  se construye el estimador  $\hat{\delta}_i(x)$  con base en las estimaciones de  $\mu_i$ ,  $\sigma^2$  y  $\pi_i$  siguientes, a menos que se cuente con información a priori respecto la distribución de  $Y$ :

$$\hat{\mu}_i = \frac{1}{n_i} \cdot \sum_{\{m | y_m=i\}} x_m, \quad \hat{\sigma}^2 = \frac{1}{n-k} \cdot \sum_{i=1}^k \sum_{\{m | y_m=i\}} (x_m - \hat{\mu}_i)^2, \quad y \quad \hat{\pi}_i = \frac{n_i}{n};$$

donde  $n_i$  corresponde a la cantidad de observaciones pertenecientes a la clase  $i$  en la muestra.

#### 4.4.2. Análisis de Discriminante Cuadrático

Este método es análogo al descrito en la sección anterior, con la diferencia de que mediante el mismo no se hace la suposición de que la matriz de varianzas y covarianzas de la distribución conjunta Normal Multivariada de los predictores es igual para cada clase, i.e.,  $\Sigma_i$  no es necesariamente igual a  $\Sigma_j$ , con  $i, j \in \{1, \dots, k\}$ ,  $i \neq j$ . De esta forma, la función de probabilidad a posteriori para  $i \in \{1, \dots, k\}$  puede expresarse de igual manera que para el Análisis de Discriminante Lineal. Además, maximizarla es equivalente a maximizar la función discriminante, la cual en este caso está dada como sigue

$$\begin{aligned}\delta_i(x) &= -\frac{1}{2}(x - \mu_i)^t \Sigma_i^{-1} (x - \mu_i) - \frac{1}{2} \ln |\Sigma_i| + \ln \pi_i \\ &= -\frac{1}{2} x^t \Sigma_i^{-1} x + x^t \Sigma_i^{-1} \mu_i - \frac{1}{2} \mu_i^t \Sigma_i^{-1} \mu_i - \frac{1}{2} \ln |\Sigma_i| + \ln \pi_i.\end{aligned}$$

Adicionalmente, tal función es cuadrática respecto  $x$ , motivo por el cual toma su nombre el método. Finalmente, se procede a obtener  $\hat{\delta}_i(x)$  para  $i \in \{1, \dots, k\}$  con base en los estimadores de  $\hat{\Sigma}_i$ ,  $\hat{\mu}_i$  y  $\hat{\pi}_i$  dados de la misma manera que para Análisis de Discriminante Lineal –excepto el de  $\Sigma$ , en que sólo se hace la adecuación necesaria para ser estimador de la matriz de varianzas y covarianzas por clase–.

#### 4.5. $k$ -Vecinos Más Cercanos.

El método  $k$ -Vecinos más Cercanos – $k$ -Nearest Neighbours ( $k$ -NN), en inglés– corresponde a un clasificador *basado en memoria que no requiere el ajuste de ningún modelo*<sup>10</sup> [16, p. 463]. Éste consiste en la clasificación de una nueva observación  $x^*$  a la categoría predominante en los  $k$  puntos más cercanos a ella, i.e., a la moda de sus  $k$  vecinos más cercanos. Por lo general, bajo la asunción de que  $X$  toma valores reales, se utiliza la distancia Euclidiana para la detección de los  $k$  vecinos más cercanos de  $x^*$ ; sin embargo, no existe restricción para la utilización de otras métricas.

---

<sup>10</sup>“(These classifiers) are *memory-based*, and require no model to be fit”; [16, p. 463].

## 4.6. Árboles de Clasificación

Los *Árboles de Decisión* son métodos basados en la estratificación del espacio de predictores en hipercubos disjuntos con el fin de predecir la salida de nuevas observaciones a través de la media o moda de los puntos que se encuentren en la región a la que pertenezcan. Los *Árboles de Clasificación* son aquellos *Árboles de Decisión* en que la variable respuesta es categórica.

Pese a que lo ideal sería obtener todas las particiones posibles del espacio de predictores y escoger alguna combinación de ellas que minice cierto criterio de error, sería una labor muy compleja computacionalmente hablando, sobre todo a medida que se tienen más observaciones –lo cual es deseable– y más predictores. De manera que la implementación de los *Árboles de Clasificación* es llevada a cabo mediante un algoritmo de *partición binaria recursiva* que consiste en lo siguiente:

1. En primera instancia, para el atributo  $X_i$ , se divide el espacio de predictores en las dos regiones  $\{X \mid X_i < s\}$  y  $\{X \mid X_i \geq s\}$ ; donde  $s$  es una constante que minimiza el criterio de error utilizado y es llamada punto de corte. Esto es llevado a cabo para cada predictor ( $i \in \{1, \dots, p\}$ ) y se escoge la partición, dada por el predictor y el punto de corte asociado, en que el error de predicción sea menor, es decir, se construye la primera bifurcación del árbol.
2. Tras obtener la primera partición, se repite el proceso anterior partiendo cada una de las dos regiones resultantes de la misma manera (para cada predictor  $X_1, \dots, X_p$ ) por separado. Una vez más, se escoge aquella que minimice la medida de error utilizada, particionando así el espacio de predictores en tres regiones.
3. Se reproduce el algoritmo recursivamente –en cada paso dividiendo en dos cada región definida hasta ese momento– hasta cumplir con cierto criterio. Por ejemplo, que en cada región estén contenidas al menos cinco observaciones, como es sugerido en [19, p. 307].

Principalmente, se utilizan tres criterios de error a minimizar para la construcción de los *Árboles de Clasificación*:

- Tasa de error de clasificación:

$$\frac{1}{n_j} \cdot \sum_{x_m \in R_j} \mathbb{1}_{\{y_m \neq \hat{y}_m\}} = 1 - \hat{p}_{ji}.$$

- Coeficiente de Gini:

$$\sum_{i=1}^k \hat{p}_{ji} \cdot (1 - \hat{p}_{ji}).$$

- Entropía:

$$-\sum_{i=1}^k \hat{p}_{ji} \cdot \ln \hat{p}_{ji}.$$

Donde  $R_j$  corresponde a la  $j$ -ésima región en que el espacio de predictores está dividido,  $n_j$  la cantidad de puntos contenidos en ella y  $\hat{p}_{ji}$  a la proporción de observaciones pertenecientes a la clase  $i$  que se encuentran en la  $j$ -ésima región del espacio de predictores, es decir,

$$\hat{p}_{ji} = \frac{1}{n_j} \cdot \sum_{x_m \in R_j} \mathbb{1}_{\{y_m=i\}}.$$

Existen otras vías para el ajuste de un Árbol de Decisión, las cuales serán descritas a continuación: *Pruning*, *Bagging*, *Boosting* y *Random Forests*.

#### 4.6.1. *Pruning*

Cabe destacar que, como el algoritmo de partición binaria recursiva se detiene de acuerdo a una regla arbitraria en este contexto, puede resultar en un modelo saturado, afectando su capacidad de predicción (aumentando su varianza). Como solución a tal problemática, se desarrolló un método por el cual se *poda* el Árbol de Decisión  $T$  obtenido (se eliminan bifurcaciones sobre sus nodos finales), mediante un criterio de costo-complejidad, construyendo un sub-árbol que resulte en un menor error de predicción.

Similarmente a Regresión por Lasso, considérese la expresión siguiente, denominada *criterio de costo-complejidad*, que penaliza el error de predicción  $g$  de un sub-árbol  $T' \subseteq T$  mediante su cantidad de nodos terminales  $|T'|$

$$C_\alpha(T') = \sum_{m=1}^{|T'|} \sum_{x_i \in R_j} g(y_i, \hat{y}_i) + \alpha \cdot |T'|;$$

donde  $\alpha \geq 0$  corresponde a un parámetro que *gobierna el intercambio entre el tamaño del árbol y*

su bondad de ajuste. Valores mayores de  $\alpha$  resultan en árboles más pequeños  $T_\alpha$ , y viceversa para valores menores de  $\alpha$ <sup>11</sup> [16, p. 308].

Nótese que para cada valor de  $\alpha$ , puede encontrarse un sub-árbol de  $T^* \subseteq T$  tal que  $C_\alpha(T^*)$  se minimice. Posteriormente, se procede a fijar  $\hat{\alpha}$  escogiendo aquel valor que minimiza la estimación del primer término de  $C_\alpha$  obtenida mediante *Cross-Validation*<sup>12</sup>, siendo  $T_{\hat{\alpha}}$  el árbol resultante de la poda.

### 4.6.2. *Bagging*

Este método es un *procedimiento general para reducir la varianza de algún método de aprendizaje estadístico*<sup>13</sup> [19, p. 316]. Aplicado al tópico en cuestión, consiste en construir un Árbol de Clasificación para cada uno de  $m > 1$  conjuntos de entrenamiento generados por *Bootstrapping*<sup>14</sup> con base en la muestra. Posteriormente, la categorización asignada a la observación  $j$ -ésima estará dada por la moda de las predicciones asociadas la misma observación en cada árbol generado, i.e.,

$$\hat{y}_j = \operatorname{argm\acute{a}x}_{i \in \{1, \dots, k\}} \sum_{r=1}^m \mathbb{1}_{\{\hat{y}_{jr}=i\}};$$

donde  $\hat{y}_{jr}$  denota la predicción de la observación  $j$  bajo el Árbol de Clasificación  $r$ .

Cabe destacar que, *Bagging* no involucra poda, por lo que cada árbol obtenido es alto en varianza. Sin embargo, son modelos insesgados y, al hacer la predicción como se menciona, se reduce considerablemente la varianza de la clasificación en conjunto.

### 4.6.3. *Random Forests*

Un problema asociado a *Bagging* es la correlación entre los árboles construidos, lo cual tiene un efecto de alza en la varianza del modelo. Con *Random Forests* se pretende reducir la correlación

---

<sup>11</sup>“(The tuning parameter  $\alpha \geq 0$ ) governs the tradeoff between tree size and its goodness of fit to the data. Large values of  $\alpha$  result in smaller trees  $T_\alpha$ , and conversely for smaller values of  $\alpha$ ”; [16, p. 308].

<sup>12</sup>Véase Sección 4.9 Validación de Modelos y Remuestreo.

<sup>13</sup>“(*Bagging*) is a general-purpose procedure for reducing the variance of a statistical learning method”; [19, p. 316].

<sup>14</sup>Véase Sección 4.9. Validación de Modelos y Remuestreo.

entre los árboles para reducir la varianza de la clasificación.

El procedimiento que *Random Forests* involucra es parecido al de *Bagging*: Se generan  $m > 1$  muestras de tamaño  $n$  por *Bootstrap* y se ajustan modelos de árboles para cada una, prediciendo de la misma forma descrita anteriormente. Para ello, en cada iteración del algoritmo de partición binaria recursiva, en lugar de elegir la siguiente partición con base en los  $p$  predictores, se lleva a cabo con base en  $l \leq p$  de ellos. Estos son escogidos aleatoria y equiprobablemente al inicio de cada iteración y el valor en qué fijarlo sugerido por los autores es de  $\sqrt{p}$ , sin embargo, puede estimarse mediante *Cross-Validation*.

#### 4.6.4. *Boosting*

Un enfoque alternativo para mejorar el rendimiento de los Árboles de Decisión es el *Boosting*. Pese a que su creación fue inspirada en *Bagging*, mediante éste, los Árboles de Regresión *son construidos secuencialmente: cada árbol es construido utilizando información de árboles construidos anteriormente*<sup>15</sup> [19, p. 321] a través de los residuos  $r$  del modelo. Luego, se añade la información obtenida del último árbol ajustado para actualizar los residuos en el modelo. El algoritmo está dado como sigue [19, p. 323], donde  $B$  es el número de árboles a construir,  $\lambda > 0$  es un parámetro de *encogimiento*,  $d$  es el número de particiones utilizadas para ajustar cada árbol y  $\hat{f}$  corresponde al árbol a ajustar iterativamente:

1. Para cada  $i$  en el conjunto de entrenamiento, asignar  $\hat{f}(x) = 0$  y  $r_i = y_i$ .
2. Para  $b \in \{1, \dots, B\}$ :
  - a) Ajustar un árbol  $\hat{f}_b$  de  $d$  particiones con  $(X, r)$  como conjunto de entrenamiento.
  - b) Actualizar  $\hat{f}(x) \leftarrow \hat{f}(x) + \lambda \cdot \hat{f}_b(x)$ .
  - c) Actualizar los residuos  $r_i \leftarrow r_i - \lambda \cdot \hat{f}_b(x_i)$ .
3. Regresar el modelo

$$\hat{f}(x) = \sum_{b=1}^B \lambda \cdot \hat{f}_b(x).$$

---

<sup>15</sup>“(Boosting works in a similar way, except that the trees) are grown *sequentially*: each tree is grown using information from previously grown trees”; [19, p. 321].

Pese a que existe más de un algoritmo para la construcción de Árboles de Clasificación con *Boosting* –como puede consultarse en [12]–, comúnmente el procedimiento utilizado es la aplicación de la transformación *logit* sobre la variable respuesta  $Y$  (dicotómica). Ahora bien, para el caso en que  $Y$  toma valores en un conjunto de más de dos categorías, se reestructura la base de datos: se transforma la variable respuesta en variables *dummy* y se ajustan en cada iteración del algoritmo  $k$  Árboles de Clasificación (uno para cada clase). Finalmente, la predicción para la observación  $i$  en el conjunto de entrenamiento asociado al  $b$ -ésimo árbol de *Boosting* ( $\hat{f}_b$ ) corresponde a la moda de las predicciones de los  $k$  árboles ajustados.

## 4.7. Análisis de Componentes Principales

El Análisis de Componentes Principales (PCA –*Principal Components Analysis*, en inglés–) es un método mediante el cual se pretende  *sintetizar*  la información brindada por  $X$ , que se encuentra en un espacio de dimensión  $p$ , en uno de menor dimensión. Para ello, se busca una transformación lineal de  $X$ ,

$$\Phi_i = \sum_{j=1}^p \phi_{ij} \cdot X_j;$$

donde  $\Phi_i$  es denominado el  $i$ -ésimo *Componente Principal* de  $X$ ,  $i \in \{1, \dots, p\}$  y es tal que se cumpla:

- Cada  $\Phi_i$  está normalizado, i.e., para todo  $j \in \{1, \dots, p\}$

$$\sum_{j=1}^p \phi_{ij}^2 = 1.$$

- Las varianzas de los Componentes Principales son no crecientes respecto su índice

$$\text{var}(\Phi_1) \geq \dots \geq \text{var}(\Phi_p).$$

- Los Componentes Principales son linealmente independientes, i.e., ortogonales.

En este contexto, se hace el supuesto de que las variables  $X_1, \dots, X_p$  están centradas (media nula). Entonces, nótese que para  $i \in \{1, \dots, p\}$

$$\text{var}(\Phi_i) = \text{var}\left(\sum_{j=1}^p \phi_{ij} \cdot X_j\right) = \mathbb{E}\left[\left(\sum_{j=1}^p \phi_{ij} \cdot X_j\right)^2\right].$$

De manera que contando con una muestra de tamaño  $n$ ,

$$\widehat{\text{var}}(\Phi_i) = \widehat{\mathbb{E}}\left[\left(\sum_{j=1}^p \phi_{ij} \cdot X_j\right)^2\right] = \frac{1}{n} \cdot \sum_{k=1}^n \left(\sum_{j=1}^p \phi_{ij} \cdot x_{jk}\right)^2;$$

donde  $x_{jk}$  denota la  $k$ -ésima observación de la  $j$ -ésima variable  $X$ .

Lo que implica que obtener el primer Componente Principal equivale a estimar el vector de *loadings*  $(\phi_{11}, \dots, \phi_{1p})$  optimizando la siguiente función objetivo

$$\begin{aligned} & \max_{\phi_{11}, \dots, \phi_{1p}} \left\{ \frac{1}{n} \cdot \sum_{k=1}^n \left( \sum_{j=1}^p \phi_{ij} \cdot x_{jk} \right)^2 \right\} \\ & \text{s.a. } \sum_{j=1}^p \phi_{1j}^2 = 1. \end{aligned}$$

Análogamente, para el segundo Componente Principal debe maximizarse

$$\begin{aligned} & \max_{\phi_{21}, \dots, \phi_{2p}} \left\{ \frac{1}{n} \cdot \sum_{k=1}^n \left( \sum_{j=1}^p \phi_{ij} \cdot x_{jk} \right)^2 \right\} \\ & \text{s.a. } \sum_{j=1}^p \phi_{2j}^2 = 1 \text{ y } \Phi_1 \perp \Phi_2. \end{aligned}$$

Posteriormente, el tercer Componente Principal puede ser encontrado de la misma manera añadiendo la restricción de ser ortogonal respecto los primeros dos componentes; y así sucesivamente.

Escribiendo la transformación lineal buscada de manera matricial como  $\Phi = A (X_1 \dots X_p)^t$ , puede resolverse el problema de optimización mediante descomposición espectral, resultando en  $A = (v_1 \dots v_p)$ , donde  $v_i$  es el vector columna que corresponde al  $i$ -ésimo eigenvector de la matriz

$\hat{\Sigma}$  de varianzas-covarianzas muestral de  $X$

$$\hat{\Sigma} = \begin{bmatrix} \widehat{\sigma}_1^2 & \widehat{\sigma}_{12} & \dots & \widehat{\sigma}_{1p} \\ \widehat{\sigma}_{21} & \widehat{\sigma}_2^2 & \dots & \widehat{\sigma}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\sigma}_{p1} & \widehat{\sigma}_{p2} & \dots & \widehat{\sigma}_p^2 \end{bmatrix}; \text{ donde}$$

$$\widehat{\sigma}_i^2 = \frac{1}{n-1} \cdot \sum_{k=1}^n (x_{ik} - \bar{x}_i)^2 \quad \text{y} \quad \widehat{\sigma}_{ij} = \widehat{\sigma}_{ji} = \frac{1}{n-1} \cdot \sum_{k=1}^n (x_i - \bar{x}_{ik})(x_j - \bar{x}_{jk}),$$

para  $i, j \in \{1, \dots, p\}, i \neq j$ .

Adicionalmente, la varianza del  $i$ -ésimo Componente Principal, el cual es determinado por el  $i$ -ésimo eigenvector de  $\hat{\Sigma}$ , está dada por el eigenvalor asociado  $\lambda_i$ . Luego, es posible encontrar la proporción de varianza de  $X$  explicada por el  $i$ -ésimo Componente Principal como sigue

$$\frac{\lambda_i}{\sum_{j=1}^p \lambda_j}.$$

Cabe destacar que la solución está bien definida puesto que  $\hat{\Sigma}$  es simétrica y, entonces, diagonalizable por el *Teorema de Descomposición Espectral*. Razón por la cual, los Componentes Principales son únicos (numéricamente), excepto tal vez por los signos de los *loadings*, ya que el *vector de loadings de cada Componente Principal especifica su dirección en el espacio  $p$ -dimensional: cambiar el signo no tiene ningún efecto puesto que su dirección no cambia*<sup>16</sup> [19, p. 382].

Este método brinda información esencial sobre los predictores  $X$ ; lo cual se vuelve crucial en casos en que  $p$  es grande y lo convierte en una herramienta poderosa de análisis exploratorio y selección de modelos. Adicionalmente, en conjunto con el Análisis de Conglomerados, el PCA puede ser utilizado para clasificación dado que *tiende a encontrar una representación de baja dimensión de las observaciones que explican una buena fracción de la varianza*<sup>17</sup> [19, p. 385].

<sup>16</sup>“Each principal component loading vector specifies a direction in  $p$ -dimensional space: flipping the sign has no effect as the direction does not change”; [19, p. 382].

<sup>17</sup>“(PCA) looks to find a low-dimensional representation of the observations that explain a good fraction of the variance”; [19, p. 385].

## 4.8. Análisis de Conglomerados

Dadas las observaciones de los atributos  $X$ , con el Análisis de Conglomerados se busca asignarlas a distintos grupos –*clusters*– mutuamente excluyentes. En específico, el objetivo de éste método es aglomerar las observaciones en grupos con base en sus disimilitudes dos a dos: minimizándolas entre miembros del mismo grupo y maximizándolas para elementos de grupos diferentes.

Primero, se introduce el concepto de disimilitud basada en los atributos para dos observaciones  $x_i, x_{i'}$  con  $i, i' \in \{1, \dots, n\}$

$$D(x_i, x_{i'}) = \sum_{l=1}^p d(x_{il}, x_{i'l});$$

donde  $x_{il}$  denota a la  $l$ -ésima observación del atributo  $i$  y  $d(x_{il}, x_{i'l})$  corresponde a una medida de disimilitud.

Luego, se busca asignar a cada observación  $i$  una clase  $C(i) = j$  con  $j \in \{1, \dots, k\}$ ; donde  $k < n$  es el número de diferentes conglomerados (decidido en un principio). Además, el objetivo es minimizar la disimilitudes dentro del conglomerado –*within clusters*–

$$W(C) = \frac{1}{2} \sum_{j=1}^k \sum_{i, i' \in I_j} d(x_i, x_{i'});$$

donde  $I_j$  es el conjunto de índices de las observaciones que pertenecen a la clase  $j$ , i.e.,  $I_j = \{i \in \{1, \dots, n\} \mid C(i) = j\}$ . O, equivalentemente, maximizar las disimilitudes entre los conglomerados –*between clusters*–

$$B(C) = \frac{1}{2} \sum_{j=1}^k \sum_{i \in I_j} \sum_{i' \in I_j^c} d(x_i, x_{i'}).$$

Nótese además que  $W(C) = T - B(C)$ , donde  $T$  es la dispersión total de los puntos (constante independientemente del resultado del Análisis de Conglomerados).

Tras este breve preámbulo, se mencionará uno de los métodos utilizados para el Análisis de Conglomerados: *Hierarchical Clustering*. Éste toma su nombre del hecho de que *produce representaciones jerárquicas en las cuales los conglomerados a cada nivel de la jerarquía son creados*

uniendo conglomerados del próximo nivel más bajo. Al nivel más bajo, cada conglomerado contiene una sola observación. Al nivel más alto, hay solo un conglomerado que contiene a toda la muestra<sup>18</sup> [16, p. 520]. Cada nivel resultante del Análisis de Conglomerados Jerárquico es representado gráficamente en un árbol binario denominado *dendrograma*. Existen dos vías para su realización:

1. **Aglomerante:** Se comienza desde el nivel más bajo de jerarquía, agrupando observaciones recursivamente y dos a dos, i.e., empezando con tantos conglomerados de un elemento como tamaño de la muestra, en cada paso se va agrupando dos de ellos con base en su medida de disimilitud. Para esto, se definen medidas de disimilitud entre grupos  $d(G, H)$ ; donde  $G$  y  $H$  son dos *clusters*. Tres maneras diferentes de *Agglomerative Clustering* son:

- *Single Linkage:* También denominada la técnica del vecino más cercano, forma los grupos con base en aquél que sea más similar a él, i.e., utiliza la medida de disimilitud

$$d_{SL}(G, H) = \min_{i \in G, i' \in H} d(x_i, x_{i'}).$$

- *Complete Linkage:* En contraparte de Single Linkage, éste método es denominado la técnica del vecino más lejano ya que, en este caso,

$$d_{CL}(G, H) = \max_{i \in G, i' \in H} d(x_i, x_{i'}).$$

- *Average Linkage:* Se utiliza la disimilitud media entre los grupos como criterio de selección

$$d_{AL}(G, H) = \frac{1}{\#(G) \cdot \#(H)} \sum_{i \in G} \sum_{i' \in H} d(x_i, x_{i'}).$$

2. **Divisivo:** Parte del nivel más alto, y va dividiendo en dos recursivamente cada uno de los conglomerados existentes.

---

<sup>18</sup>“They (hierarchical clustering methods) produce hierarchical representations in which the clusters at each level of the hierarchy are created by merging clusters at the next lower level. At the lowest level, each cluster contains a single observation. At the highest level there is only one cluster containing all of the data”; [16, p. 520].

## 4.9. Validación de Modelos y Remuestreo

En las secciones anteriores, se ha utilizado el término *conjunto de entrenamiento*, el cual refiere a un subconjunto de las muestras observadas de  $Y$  y de  $X$ . En general, lo más conveniente es *dividir* la muestra en dos partes: el conjunto de entrenamiento, con el que se ajusta el modelo, y el conjunto de prueba, con la que éste se evalúa. La motivación de lo anterior subyace en que, naturalmente, la evaluación del modelo tiende a ser *optimista* al probar su bondad de ajuste con las observaciones por medio de las cuales fue ajustado, i.e., tiende a sobrevalorar los indicadores de bondad de ajuste o, por su contraparte, subestimar los errores de predicción.

De forma similar, los métodos de remuestreo *permiten obtener información que no estaría disponible si sólo se ajustara el modelo una vez usando la muestra de entrenamiento original*<sup>19</sup> [19, p. 175]. Dado que, en la práctica y por lo general, no es posible obtener varias muestras aleatorias independientes entre sí, el objetivo de tales métodos es obtener varias muestras a partir de una sola con el fin de construir estimadores para variables de interés. Por ejemplo, se puede obtener la media, varianza e intervalos de confianza de las tasas de error de clasificación de globales, de entrenamiento y de prueba de un modelo de clasificación.

Dos técnicas de remuestreo mediante las cuales se prosigue a la validación de estadísticas asociadas a un modelo de Aprendizaje Supervisado, son:

- *K-Fold Cross-Validation*: Se particiona la muestra en  $K$  subconjuntos de igual cardinalidad. Bajo este enfoque, la validación del modelo consiste en la evaluación sobre cada partición  $j \in \{1, \dots, K\}$  de las estadísticas de interés asociadas al modelo ajustado sobre su complemento como conjunto de entrenamiento y, luego, obtener sus estimaciones validadas correspondientes a sus medias muestrales. Por lo general, los valores del  $K$  utilizados para ello son 10, 5 ó 1 (correspondiendo este caso al método denominado *Leave-one-out Cross-Validation*).
- *Bootstrap*: Se lleva a cabo muestreo con reemplazo de la muestra  $n$  veces, conformando así *otra* muestra, sobre la cual se ajusta el modelo y se evalúa. Este proceso es repetido  $m$  veces y, de manera análoga al método anterior, se promedian las estadísticas obtenidas.

---

<sup>19</sup>“(Such an approach may) allow us to obtain information that would not be available from fitting the model only once using the original training sample”; [19, p. 175].



## CAPÍTULO 5

# MODELACIÓN MATEMÁTICA DE OBRAS MUSICALES

En este capítulo se presenta la modelación probabilística realizada. Cabe hacer notar que tal enfoque es tomado partiendo de la idea de que una pieza musical es un sistema dinámico no determinista, con lo cual no se pretende insinuar que una obra se compone al azar o de manera errática. Por el contrario, la complejidad de su estructura ha quedado clara en el Capítulo 1, de manera que se ha pensado en la Probabilidad como una herramienta que facilite su comprensión objetiva.

Adicionalmente, se sostiene que tal conceptualización no contradice el contexto musical: la intención del autor –claramente no aleatoria y en la mayoría de los casos desconocida por el expectador– puede ser plasmada de múltiples maneras. Pensándolo matemáticamente, esta intención es concretada escogiendo una gran cantidad de parámetros a su libertad entera: tempo, escala, tonalidad, timbre, modulaciones, transposiciones, alteraciones, línea melódica, progresiones armónicas, expresión, etc. El uso específico de tales recursos en la obra puede pensarse que sucede con cierta probabilidad derivada de tendencias dadas, ya sea a partir del contexto histórico y geográfico del autor, o de sus usos musicales y estilo particulares; así como simplemente su *personalidad* y *sentir* al momento de la gestación de la pieza. Lo cual sería extremadamente difícil de representar mediante un modelo determinista.

Dicho lo anterior, cabe preguntarse sobre la naturaleza de la esencia de una obra: ¿Tendrá patrones armónicos<sup>1</sup>? ¿Qué tan diferentes son entre obras, autores y corrientes? ¿Qué comportamiento muestran a corto y largo plazo? ¿La armonía de una pieza alcanza un punto en específico? ¿Hay

---

<sup>1</sup>La respuesta inmediata es sí, conociendo los conceptos de progresión armónica y cadencia.

alguna tendencia en el uso de progresiones armónicas? ¿Escalas? ¿Acordes? ¿En qué medida recurre un autor a la disonancia? ¿Cuánto depende el uso de tales recursos musicales de la corriente histórica a la que el autor pertenece? Todas estas prerrogativas buscan responderse mediante la modelación presentada, brindando evidencia cuantitativa de tales hechos.

Ahora bien, la dificultad inicial para la propuesta de un modelo que explique tal fenómeno es su congruencia con la naturaleza de la Música. El marco de referencia que se utilizará es la Teoría Musical, razón por la cual se decidió dirigir la investigación meramente a obras musicales académicas, garantizando de tal manera que sigan la misma línea de composición. En específico, se utilizarán 1,456 obras musicales académicas, 1,156 de ellas contenidas en la base de datos *Classical Archives Corpus* de la Universidad de Yale [36] y el resto<sup>2</sup> pertenecientes al corpus de la paquetería para Python desarrollada por el MIT, *music21*<sup>3</sup>. Serán consideradas únicamente obras para piano (y sus antecesores), orquesta, cuarteto de cuerdas y ensamble de vientos<sup>4</sup> con duraciones no menores a 200 compases<sup>5</sup> y compuestas por los autores: Bach, Bartók, Beethoven, Chopin, Debussy, Dvorák, Handel, Haydn, Joplin, Liszt, Mozart, Palestrina, Rachmaninov, Rameau, Ravel, Schubert, Tchaikovsky y Vivaldi, entre otros. En el Cuadro 5.1 se muestra la composición de la base de datos en cuanto a autores y corrientes históricas musicales; los títulos y autores de cada obra se encuentran en el *Anexo V*, el cual es consultable en [github.com/danielapdv/music-mcs](https://github.com/danielapdv/music-mcs).

Corriente histórica	Autores	Obras
Renacimiento	1	300
Barroco	4	315
Clasicismo	3	300
Romanticismo	6	339
Modernismo	60	202
Total	74	1,456

Cuadro 5.1: Cantidad de obras musicales y autores en la base de datos a utilizar por corriente histórica.

Como vía para caracterizar una pieza se recurrirá a su armonía; en medida que tanto el ritmo y la melodía son elementos claramente indispensables para su conformación pero, como fue esta-

<sup>2</sup>Obras compuestas por Palestrina, que representarán el total de las obras musicales del Renacimiento a utilizar.

<sup>3</sup>Desarrollada por Michael Cuthbert, Christopher Ariza, Benjamin Hogue y Josiah Wolf Oberholtzer.

<sup>4</sup>Nótese que obras para instrumentos solistas, exceptuando el piano (y sus antecesores), no brindarán mucha información para la modelación que se planea llevar a cabo, puesto que carecen de acompañamiento.

<sup>5</sup>Ya que se pretende llevar a cabo estimaciones, es necesario que las obras musicales contempladas sean de un largo *suficientemente grande*.

blecido en el primer capítulo, mediante la armonía se da carácter y significado a las transiciones y relaciones de los sonidos simultáneos de una obra. Así, la modelación deberá caracterizar la esencia armónica de una obra musical, por lo que se centrarán esfuerzos en caracterizar su tempo y funciones armónicas, para lo cual la concepción como sistema dinámico es ideal.

En primera instancia, sería lógico recurrir a la numeración romana para modelar las progresiones armónicas de la pieza a través de su función armónica respecto la tónica. Sin embargo, presentaría una gran dificultad práctica: sin conocer de voz directa del autor la tonalidad en que se encuentra la pieza, junto con el grado de los acordes a lo largo de ella, habrían de determinarse, lo cual podría ser llevado a cabo de manera cualitativa únicamente por personas dedicadas a la Música con oído experimentado. Cabe destacar que se han implementado algoritmos para resolver tal dificultad, como se expone en [26] ó [31]. No obstante, aún no distinguen ni tratan de la mejor manera las notas ornamentales y notas fundamentales, tampoco se extiende la determinación de un acorde a más allá de uno de cualidad menor, mayor o séptima mayor. De cualquier forma, de tener acceso a tal información, aún habría que lidiar con algunos detalles sutiles, pero cruciales, que no permitirían caracterizar la armonía de manera eficiente, lo que lleva a descartar esta vía de modelación:

1. La equivalencia entre acordes y escalas a través de inversiones y transposiciones.
2. No sería adecuada para la Música Postonal, en general, sería erróneo aplicar tal enfoque a la Música del Modernismo. Así como posiblemente a la Música Renacentista, puesto que en ese entonces los conceptos musicales a utilizar aún no se encontraban desarrollados.
3. No funcionaría para el manejo de modulaciones.

Habiendo establecido lo anterior, surge la motivación de encontrar una forma de modelar el tempo y funciones armónicas de manera más general, es decir, *immune* a la tonalidad y escalas –e, incluso, signo de compás– específicas en que se encuentren las obras musicales, más no que las ignore. Las notas ornamentales representan un primer gran obstáculo; pese a que sería tentador contemplar únicamente el acompañamiento de una pieza para de alguna forma extraer su armonía, sería erróneo, ya que las notas de la línea melódica (no ornamentales) sí alteran el acorde que se esté tocando en el acompañamiento. Por lo que, sin alternativa encontrada para esquivar eficientemente esta dificultad, la determinación de la naturaleza ornamental o no de una nota, se decidió incluirlas en el análisis.

Figura 5.1: Primeros cuatro compases de *Polonaise Op. 89 en Do Mayor* de Beethoven y su transformación para fines de la modelación. El tercer pentagrama corresponde a la transformación llevada a cabo para cada intervención.

Otro factor a considerar anteriormente mencionado es la equivalencia entre acordes a través de su altura, inversión y transposición, lo cual podría llevar a la representación repetida. Para evitarlo, se transformó cada acorde observado a su posición básica y se situó en la octava cuatro. En concreto, cada pieza fue transformada como es ejemplificado en la Figura 5.1, específicamente:

1. En cada intervención, se construye un acorde con todas las notas tocadas en ella, tanto en melodía como acompañamiento, resultando en un sólo acorde por intervención.
2. Luego, cada uno de los acordes resultantes son llevados a su posición básica y a la cuarta octava; con el fin de que no se presenten notas repetidas a diferentes alturas.

De esta manera, se procederá a caracterizar la armonía de una pieza mediante su tempo y funciones armónicas indirectamente, a través de las intervenciones transformadas a lo largo de ella; siendo necesario dirigir la modelación con miras hacia la detección de patrones en su uso para

poder cuantificar su variación. Finalmente, el supuesto inicial que considerará es que las intervenciones de una pieza se hacen dependiendo de un número fijo de intervenciones antecedentes, es decir, cumplen con la Propiedad Markoviana.

## 5.1. Caracterización mediante Cadenas de Markov

En primera instancia, se propone un modelo orientado mayormente hacia la explicación del tempo armónico. Para ello, se centrará en las siguientes tres características de cada intervención: duración, raíz del acorde y su Clase Forte (Véase *Anexo I. Clasificación de Allen Forte*). Nótese que un acorde está totalmente dado por su nota fundamental –que, en adición, al observar su evolución en el tiempo provee información sobre la escala y tonalidad utilizadas, debido a que se verán reflejadas en su frecuencia de aparición– y los intervalos entre ésta y el resto de las notas que lo componen –información resumida en la clasificación de Allen Forte del conjunto tonal de un acorde–. Contemplar tales características, junto con los patrones de sus duraciones, brinda información sobre la medida en que los cambios ocurren en la armonía de la pieza puesto que, mediante ellos, se refleja el uso de las funciones armónicas (notas fundamentales), así como la consonancia y disonancia dentro de cada acorde y respecto al cambio de su inmediato anterior.

En [14], Christophe Guillotel-Nothmann menciona:

*“La teoría de vectores armónicos identifica a la asimetría de las progresiones de las fundamentales como una de las características sobresalientes de la armonía tonal. Uno de los posibles orígenes de este fenómeno es el tratamiento de la disonancia — mediante el cual las reglas de preparación y resolución tienen repercusiones en las secuencias armónicas.”*<sup>6</sup>

– Guillotel-Nothmann, 2009

Entonces, bajo la motivación de que las sucesivas intervenciones en una pieza llevan a su cénit, i.e., tienden a cierto punto armónico, se propone caracterizar a cada obra musical mediante las probabilidades límite de las Cadenas de Markov asociadas a tales conceptos –con los que se pretende

---

<sup>6</sup>“The theory of harmonic vectors identifies the asymmetry of root progressions as one of the outstanding characteristics of tonal harmony. One of the possible origins of this phenomenon is the treatment of the dissonance — through which the rules of preparation and resolution have repercussions on harmonic sequences”; [14, p. 9].

representar la asimetría de las progresiones armónicas—.

### 5.1.1. Procesos estocásticos a considerar

Considérese tres Cadenas de Markov Homogéneas asociadas a las intervenciones a lo largo de una obra musical, que representen su duración  $\{D_n\}_{n \in \mathbb{N}}$ , la raíz del acorde  $\{R_n\}_{n \in \mathbb{N}}$  y la Clase Forte de su conjunto tonal  $\{C_n\}_{n \in \mathbb{N}}$  respectivamente. Sus espacios de estados, junto con una breve descripción, se muestran en el Cuadro 5.2. Nótese que describir tales conceptos con Cadenas de Markov permitirá la modelación de la tendencia de las intervenciones de una pieza a largo plazo, su armonía, a través de sus probabilidades límite.

Cadena	Notación	Espacio de estados	Denotación
Duraciones	$\{D_n\}_{n \in \mathbb{N}}$	Breve, redonda, blanca, negra, corchea, semicorchea, fusa, semifusa y otra.	Duración del acorde o silencio asociada a la $n$ -ésima intervención. En “otra” se encuentran otras duraciones menos comunes.
Raíces	$\{R_n\}_{n \in \mathbb{N}}$	0, 1, 2, ..., 12.	Semitono en una octava asociado a la raíz o nota fundamental de la $n$ -ésima intervención. ej. 1 refiere a la nota <i>Do</i> , 2 a <i>Reb</i> . El estado 0 representa silencio.
Clase Forte	$\{C_n\}_{n \in \mathbb{N}}$	Todas las clases dadas en el <i>Anexo I</i> .	Clase Forte del conjunto tonal asociado a la $n$ -ésima intervención.

Cuadro 5.2: Cadenas de Markov a considerar para la primera modelación, sus espacios de estados y su denotación.

Es tiempo de hacer notar que, bajo esta línea de pensamiento, lo ideal sería considerar la Cadena de Markov  $\{(D_n, R_n, C_n)\}_{n \in \mathbb{N}}$  para la síntesis de la información de cada intervención, la modelación de sus características en conjunto. Sin embargo, tal conceptualización resultaría en una cadena con matriz de transición con entradas mayormente nulas, dadas todas las combinaciones posibles de la terna, de manera que no podría garantizarse su irreducibilidad y, luego, su ergodicidad. Adicionalmente, su espacio de estados tendría cardinalidad  $10 \cdot 13 \cdot 278 = 36,140$ , haciendo de la estimación de su matriz de transición un proceso computacional sumamente lento. Por ello, se realizó la modelación mediante las tres cadenas de manera independiente, pese a que éste representa un gran

supuesto, musicalmente hablando.

Utilizando la paquetería `music21`, se obtuvieron las realizaciones de cada una de las Cadenas de Markov propuestas para las intervenciones transformadas –como fue descrito en la sección anterior– de las obras a considerar en la base de datos por pieza musical. Posteriormente, fueron estimadas sus matrices de transición por Máxima Verosimilitud, como fue planteado en el Capítulo 2, Sección 2.3, añadiendo un suavizamiento. Recuérdese que la estimación por Máxima Verosimilitud de la entrada  $i, j$  de la matriz de transición corresponde a la cantidad de veces que el estado  $i$  fue seguido del  $j$  dividido por la cantidad de veces que se visitó el estado  $i$ , i.e.,

$$\frac{n_{ij}}{n_i}.$$

El suavizamiento añadido corresponde a asignar a la entrada  $i, j$  de la matriz el siguiente término, en lugar del anterior

$$\frac{n_{ij} + 1}{n_i}.$$

Nótese que de esta manera se soluciona la problemática derivada de la cantidad de realizaciones observadas de la cadena, la cual tiene impacto directo en la cantidad de entradas nulas de la estimación de la matriz. Asimismo, a medida que dicha cantidad tiende a infinito, el sesgo aumentado de  $\frac{1}{n_i}$  tiende a cero.

Adicionalmente, fueron estimadas las Cadenas de Markov asociadas a los mismos conceptos para las corrientes históricas musicales determinadas por la Musicología (Renacimiento, Barroco, Clasicismo, Romanticismo y Modernismo); así como asociadas a los autores: Bach, Beethoven, Chopin, Handel, Haydn, Mozart, Palestrina y Vivaldi<sup>7</sup>. Esto, notando que la cadena asociada a cada corriente o autor puede ser obtenida generalizando la expresión de la función de Verosimilitud para todas las obras que correspondan en cada caso; obteniendo los mismos estimadores, pero extendiendo el conteo de las veces que una intervención fue seguida de otra a todas las piezas musicales que comprendan cada caso (asumiendo que son independientes entre sí). También, cabe destacar que bajo este segundo método de estimación no fueron incluidos suavizamientos puesto que, dado que la cardinalidad de observaciones aumenta al considerar más de una obra, no se consideró necesario.

---

<sup>7</sup>Autores cuya cantidad de obras en la base de datos se prestó para tales estimaciones: 105, 100, 114, 93, 100, 100, 300 y 104, respectivamente.

### 5.1.2. Probabilidades límite

Nótese que  $\{D_n\}_{n \in \mathbb{N}}$ ,  $\{R_n\}_{n \in \mathbb{N}}$  y  $\{C_n\}_{n \in \mathbb{N}}$  son Cadenas de Markov Homogéneas con espacios de estados finitos, por lo que son recurrentes positivas de acuerdo con el Teorema 2.2. Asimismo, el método de estimación de sus matrices de transición resulta en matrices con entradas estrictamente positivas, por lo que cuentan con la característica de ser irreducibles y aperiódicas. De manera que, por el Teorema 2.1, existen sus probabilidades límite, son únicas y sus entradas son estrictamente positivas.

Ahora bien, si las probabilidades límite  $\pi$  existen, por la Proposición 2.2,

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n = j \mid X_0 = i) = \lim_{n \rightarrow \infty} (P^n)_{ij} = \pi_j;$$

con lo cual pueden obtenerse numéricamente las probabilidades límite para cada cadena. Nótese que el proceder natural sería obtener el eigenvector asociado al primer eigenvalor de la matriz de transición, el cual determina la distribución estacionaria. Por la forma de las matrices de transición y, de hecho, por el teorema de convergencia, tal eigenvalor existe. Sin embargo, este método resulta poco práctico puesto que al obtener computacionalmente el determinante de las matrices de transición, se obtienen valores nulos: no porque el determinante verdaderamente sea nulo, sino porque calcularlo resulta de la multiplicación de números cercanos a cero (las probabilidades), perdiéndose por la precisión de la máquina.

La modelación propuesta en la presente sección fue implementada y aplicada a la base de datos descrita al inicio del capítulo. Los resultados generales obtenidos son analizados en la Sección 5.3 y, en adición, en el *Anexo II* se describe el proceso de implementación y recabación de resultados para una obra musical en específico: *Polonaise Op. 89 en Do Mayor*, de Beethoven.

## 5.2. Caracterización mediante Campos Aleatorios de Gibbs y Modelos de Markov

A pesar de ser un buen primer método de acercamiento, la modelación anterior cuenta con algunas desventajas: no considera las cualidades mencionadas en conjunto, supone que cada intervención está dada únicamente por la inmediata anterior e intenta manejar las equivalencias de acordes por transposiciones e inversiones, pero aún depende de la raíz del acorde. De manera que en esta sección se busca la modelación de piezas musicales de una forma más general, mediante el cual las fluctuaciones en la armonía de la pieza puedan ser medidas abstractamente sin importar su escala y tonalidad.

La vía elegida para ello son los conceptos de consonancia y disonancia contrapuntísticas. En primera instancia, históricamente se ha pensado en la disonancia como la conceptualización musical de la imperfección a razón de que causa *sensaciones desagradables* en el ser humano<sup>8</sup>. Razón por la cual, el Principio Aristotélico de la tendencia de la imperfección a la perfección es manifestado en la Música mediante el concepto de cadencia, resolución de la disonancia mediante regreso a la tónica, reposo; por lo que se sostiene que cobra sentido la observación de las fluctuaciones entre la consonancia y disonancia del sistema a largo plazo, pues contextualmente se asegura su tendencia en el método de composición musical a la consonancia.

*“Uno de los parámetros que afectan las progresiones armónicas es discutido por los teóricos de la armonía del Siglo XVIII: las reglas para la preparación y resolución de la disonancia; reglas en las cuales Rameau basa en parte su teoría de nota fundamental. [...] En la presente etapa de investigación, la asimetría de las progresiones de las fundamentales parece surgir de este principio filosófico [Principio Aristotélico de la Imperfección a la Perfección] que es la verdadera razón de las reglas contrapuntísticas de las progresiones de intervalo consonantes y disonantes.”*<sup>9</sup>

– Guillotel-Nothmann, 2009

---

<sup>8</sup>A tal grado que durante la Edad Media, se evitaba el uso de acordes disonantes al ser asociados a lo *diabólico*.

<sup>9</sup>“One of the parameters that affects chord progressions is discussed by the theorists of harmony of the 18th century: the rules for the preparation and resolution of dissonance; rules on which Rameau bases partly his theory of fundamental bass. [...] At the present stage of research, the asymmetry of root progressions seems to rise from this philosophical principle which underlies the contrapuntal rules for consonant and dissonant interval progressions”; [14, p. 1].

Ahora bien, la consonancia o disonancia (en su concepción contrapuntística) causadas por un acorde están determinadas meramente por los intervalos entre los tonos que lo integran. Además, un concepto mediante el cual es posible la representación única de los intervalos de una intervención es la de vector de clase de intervalo, como fue descrito en la Sección 1.6. Por esta razón, para la segunda propuesta de modelación de la armonía de una obra musical se considerarán las fluctuaciones en la consonancia y disonancia con base en los conceptos de Teoría Musical de Conjuntos y cuantificadas mediante las medidas propuestas de consonancia y disonancia de los conjuntos tonales asociados a una sucesión de acordes. Para ello, se supondrá que cada intervención es dependiente de las  $k - 1$  inmediatas anteriores, por lo que la modelación presentada en esta sección se llevará a cabo mediante Campos de Markov de orden  $k - 1$  caracterizados por su Medida de Gibbs.

### 5.2.1. Topología del campo y dinamismo

Bajo esta modelación se considerarán secuencias de longitud  $k \geq 2$  de los vectores de clase de intervalo reducido<sup>10</sup>  $\mathcal{I}_{n_1}^*(A_1), \dots, \mathcal{I}_{n_k}^*(A_k)$  asociados a los conjuntos tonales  $A_1, \dots, A_k \in \mathcal{C}\mathcal{T}$  de las intervenciones a lo largo de una obra musical. Sea  $X$  el Campo Aleatorio dado por la topología  $(S, N)$ ; con sitios  $S = \{1, \dots, k\} \times \{1, 2\}$ , el espacio fase  $\Lambda = \{0, \dots, 66\}$  (pues el valor mínimo y máximo que pueden tomar las entradas de un vector de clase de intervalos reducido es de 0 y 66, respectivamente) y el sistema de vecindades trivial, i.e., para todo  $s \in S$ ,  $N_s = S \setminus \{s\}$ .

Es decir, las configuraciones del campo estarán dadas por las entradas de los vectores de clase de intervalo reducidos de cada secuencia, haciéndolo un campo dinámico. Además, se modelará cada pieza de acuerdo a tales secuencias observadas bajo el supuesto de que el vector de una intervención es dependiente de los vectores asociados a las  $k - 1$  intervenciones anteriores, lo cual es equivalente a que  $X$  sea un Modelo de Markov de orden  $k - 1$ , i.e.,

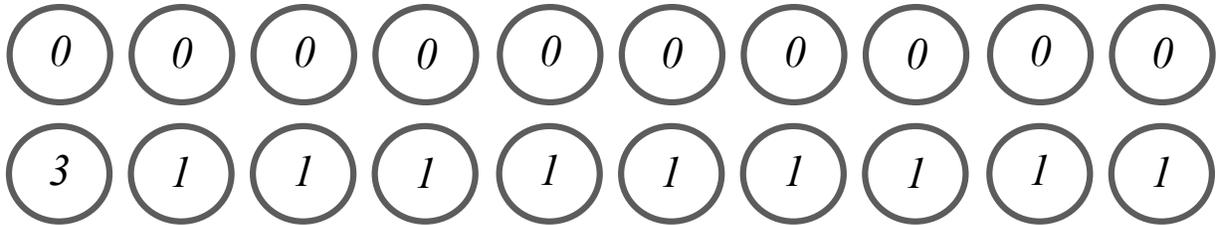
$$\begin{aligned} \mathbb{P}(\mathcal{I}_{n_m}^*(A_m) \mid \mathcal{I}_{n_{m-1}}^*(A_{m-1}), \dots, \mathcal{I}_{n_1}^*(A_1)) &= \\ &= \mathbb{P}(\mathcal{I}_{n_m}^*(A_m) \mid \mathcal{I}_{n_{m-1}}^*(A_{m-1}), \dots, \mathcal{I}_{n_{m-k+1}}^*(A_{m-k+1})); \end{aligned}$$

donde  $m \geq k$  y  $A_i$  corresponde al conjunto tonal asociado a la  $i$ -ésima intervención (acorde) en la pieza musical.

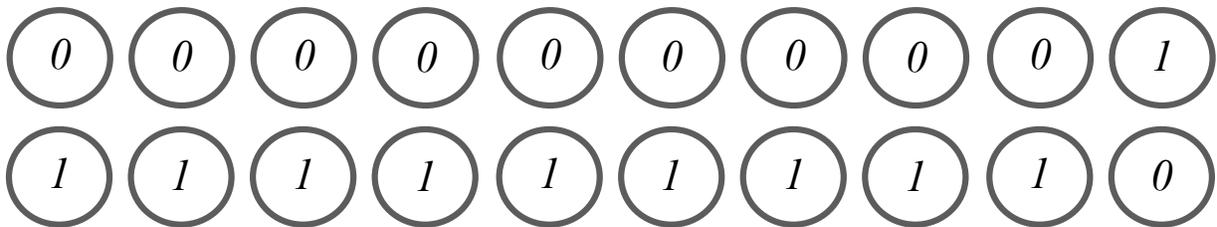
---

<sup>10</sup>La razón de no contemplar los vectores de clase de intervalo tradicionales se hará notar en las siguientes secciones.

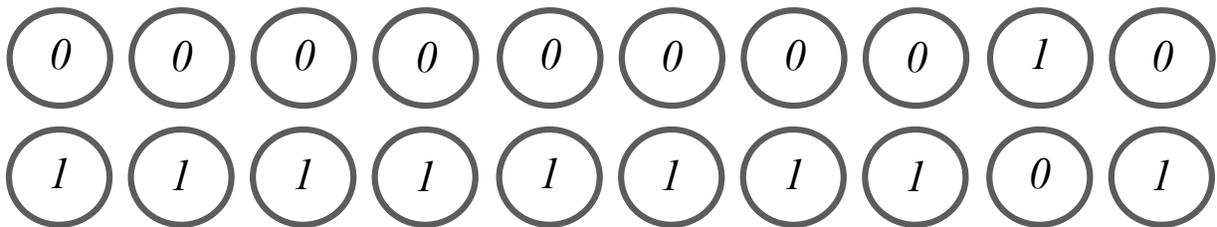
En la Figura 5.2 se muestran las primeras 3 configuraciones observadas al considerar el Campo Aleatorio  $X$  caracterizado de la manera propuesta asociado a la obra *Polonaise Op. 89 en Do Mayor* de Beethoven, con  $k = 10$ . Cabe destacar que si quisiese ver al campo definido como un modelo gráfico probabilístico, éste correspondería a un gráfico completo en que cada vértice es un sitio.



(a) Primera secuencia observada.



(b) Segunda secuencia observada.



(c) Tercera secuencia observada.

Figura 5.2: Primeras tres configuraciones observadas del Campo Aleatorio  $X$  con  $k = 10$  asociado a *Polonaise Op. 89 en Do Mayor* de Beethoven. La primera fila corresponde a la entrada del vector de clase de intervalo reducido de cada intervención relacionada a la disonancia, y la segunda a la entrada del vector relacionada con la consonancia. Las intervenciones sucesivas se representan de izquierda a derecha cronológicamente.

## 5.2.2. Potencial de activación y energía del campo

Habiendo establecido la topología del Campo Aleatorio a considerar, se procederá a caracterizarlo mediante su potencial de activación, i.e., se asumirá que es también un Campo de Gibbs<sup>11</sup>. Dada la motivación del inicio de la sección, se pretende cuantificar las fluctuaciones de consonancia y disonancia en una obra musical, para lo cual serán utilizadas las medidas de consonancia y disonancia para conjuntos tonales asociados a una sucesión de  $k$  acordes planteadas en las definiciones 1.19 y 1.20. Entonces, se definen los potenciales de activación del campo  $X$  para cada *clique*  $C \subseteq S$  como sigue

$$V_C(x) = \begin{cases} \mathcal{C}_k^*(\mathcal{J}_{n_1}^*(A_1), \dots, \mathcal{J}_{n_k}^*(A_k)) & \text{si } C = \{(1, 2), (k-1, 1), (k-1, 2), (k, 1)\} \\ \mathcal{D}_k^*(\mathcal{J}_{n_1}^*(A_1), \dots, \mathcal{J}_{n_k}^*(A_k)) & \text{si } C = \{(1, 1), (k-1, 1), (k-1, 2), (k, 2)\} \\ 0 & \text{e.o.c.} \end{cases}$$

$$= \begin{cases} \frac{66+x(k,1)+2 \cdot x(k-1,1)-x(1,2)}{2 \cdot [66+x(k-1,1)+x(k-1,2)]} & \text{si } C = \{(1, 2), (k-1, 1), (k-1, 2), (k, 1)\} \text{ y } k \geq 3 \\ \frac{66+x(k,2)+2 \cdot x(k-1,2)-x(1,1)}{2 \cdot [66+x(k-1,1)+x(k-1,2)]} & \text{si } C = \{(1, 1), (k-1, 1), (k-1, 2), (k, 2)\} \text{ y } k \geq 3 \\ \frac{66+x(2,1)-x(1,2)}{2 \cdot 66} & \text{si } C = \{(1, 2), (2, 1)\} \text{ y } k = 2 \\ \frac{66+x(2,2)-x(1,1)}{2 \cdot 66} & \text{si } C = \{(1, 1), (2, 2)\} \text{ y } k = 2 \\ 0 & \text{e.o.c.} \end{cases} .$$

Luego, la energía del campo para una configuración  $x \in \Lambda^S$  se puede expresar de la siguiente forma paramétrica

$$\varepsilon(x) = \theta_{\mathcal{C}} \cdot V_{\mathcal{C}}(x) - \theta_{\mathcal{D}} \cdot V_{\mathcal{D}}(x);$$

donde  $\mathcal{C}$  denota el *clique* asociado a la consonancia de la secuencia  $(\{(1, 2), (k-1, 1), (k-1, 2), (k, 1)\})$ ,  $\mathcal{D}$  el *clique* asociado a su disonancia  $(\{(1, 1), (k-1, 1), (k-1, 2), (k, 2)\})$  y  $\theta_{\mathcal{C}}, \theta_{\mathcal{D}} \in [0, \infty)$  corresponden los parámetros asociados a cada potencial respectivamente.

Nótese que bajo este enfoque se estará modelando el flujo de consonancia-disonancia ganada o perdida al realizar una intervención en la pieza respecto la consonancia-disonancia acumulada en las  $k-1$  intervenciones inmediatas anteriores. Con el fin de mostrar la cuantificación llevada a

<sup>11</sup>Lo cual no contradice los Teoremas 3.2 y 3.3 por su construcción, como se verá posteriormente.

cabo por las medidas propuestas como potenciales de activación, se muestran en la Figura 5.3 los valores de  $V_{\mathcal{C}}$  y  $V_{\mathcal{D}}$  para ciertas secuencias de acordes de tamaño  $k = 3$  que representan casos clave; y variando los valores de las entradas asociadas a la consonancia  $C$  y a la disonancia expresada como  $66 - C$  del tercer vector de clase de intervalo reducido en la sucesión de acordes.

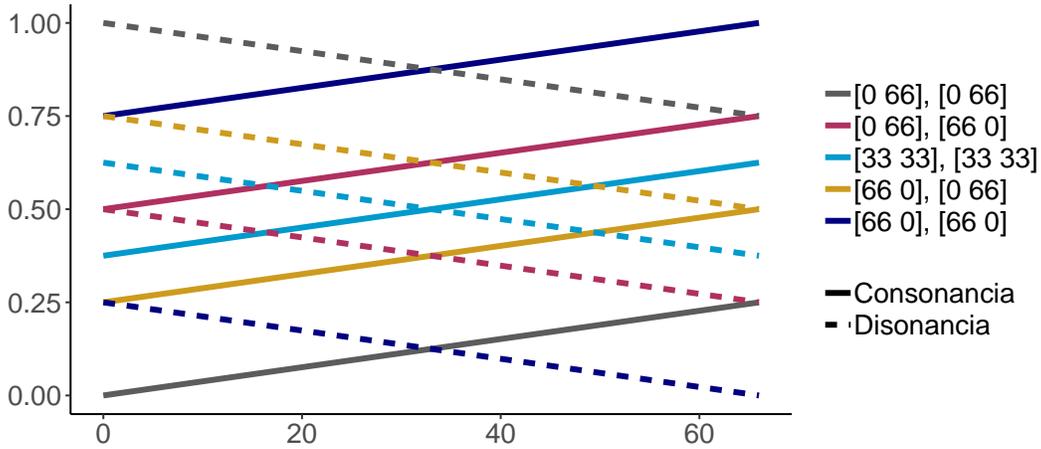


Figura 5.3: Potenciales de activación (medidas de consonancia y disonancia para conjuntos tonales) asociados a una sucesión de  $k = 3$  acordes valuados para las secuencias  $\{[0 \ 66], [0 \ 66], I\}$ ,  $\{[0 \ 66], [66 \ 0], I\}$ ,  $\{[33 \ 33], [33 \ 33], I\}$ ,  $\{[66 \ 0], [0 \ 66], I\}$  y  $\{[66 \ 0], [66 \ 0], I\}$  respecto al valor  $C$  de la entrada asociada a la consonancia del último acorde, con  $I = [C \ 66 - C]$ . Las líneas punteadas corresponden a la evaluación de  $V_{\mathcal{D}}$  y el resto a la de  $V_{\mathcal{C}}$ .

Se concluye que las medidas de disonancia y consonancia planteadas son sensibles, tanto las cualidades de las primeras intervenciones de la secuencia, como a la variación que se presenta al pasar a la última; razón por la cual son adecuadas para la modelación que se pretende realizar. En específico, cabe destacar lo siguiente:

- Manteniendo los primeros dos acordes de la progresión constantes, la consonancia del vector de clase de intervalo asociado al tercer acorde mantiene una relación directamente proporcional a la medida de consonancia de la progresión, e inversamente proporcional a la medida de disonancia. Por ello es que las líneas continuas ( $V_{\mathcal{C}}$ ) tienen un comportamiento creciente en la Figura 5.3.
- En el mismo contexto del punto anterior, la disonancia del vector de clase de intervalo del último acorde de la sucesión es inversamente proporcional a la medida de consonancia de la progresión y directamente proporcional a la medida de disonancia de la sucesión. Consecuentemente, las líneas punteadas ( $V_{\mathcal{D}}$ ) son decrecientes respecto el valor de  $C$ .

- Las cualidades del acorde inmediato anterior al último en la sucesión tienen impacto en las medidas de consonancia y disonancia. Obsérvese las líneas amarilla y azul marino: en ambos casos el primer acorde de la progresión es uno totalmente consonante [66 0]. El segundo acorde de la progresión representada en amarillo es lo más disonante posible, mientras que el segundo acorde asociado a la línea azul es lo más consonante posible. Al mantener el tercer acorde de la progresión igual para ambas sucesiones, dados los acordes de en medio, la progresión azul es mucho más consonante que la amarilla de acuerdo a la medida propuesta. Asimismo, la línea punteada amarilla se encuentra por encima de la línea punteada azul, i.e., la progresión denotada por azul es menos disonante que la amarilla. La contraparte del razonamiento es representada por las líneas gris y magenta, siendo la progresión representada por el primer color menos consonante y más disonante que la otra.
- Adicionalmente, la medida propuesta toma en cuenta los primeros acordes de la sucesión, la acumulación de disonancia y consonancia en la progresión antes del acorde final, por así decirlo. Compárese ahora las líneas magenta y azul marino: ambas representan progresiones con el mismo acorde intermedio, [66 0], pero la primera tiene un primer acorde totalmente disonante y la segunda uno totalmente consonante. De esta manera, la línea continua magenta se sitúa por debajo de la azul, ya que la progresión dado su inicio, es menos consonante que la segunda. También, la línea punteada magenta se encuentra por arriba de la punteada azul puesto que corresponde a una progresión de inicio más disonante. La cualidad contrapuesta es ejemplificada por las líneas amarilla y gris, en que la progresión representada por amarillo es más consonante y menos disonante que la gris.
- La medida de consonancia toma su valor mínimo (nulo) en el caso en que la progresión es lo menos consonante posible, a saber, con la progresión [0 66], [0 66], [0 66], pues a lo largo de ella no se *gana* consonancia en absoluto; que es el caso en que la medida de disonancia toma su valor máximo.
- El punto de máxima consonancia (unitaria) de acuerdo a la medida propuesta está dado por aquella progresión más consonante, es decir, [66 0], [66 0], [66 0]. Ahora bien, esta progresión es la que causa que la medida de disonancia se anule.
- Nótese que

$$V_{\mathcal{C}}([C_1 D_1], [C_2 D_2], [33 33]) = V_{\mathcal{D}}([66 - C_1 66 - D_1], [66 - C_2 66 - D_2], [33 33]);$$

con  $C_i, D_i \in \{0, 66\}$ . Esto es debido a que con la primera progresión de acordes se pierde (gana) la misma cantidad de consonancia (disonancia) que se gana (pierde) con la segunda.

- Para los casos en que la progresión involucra acordes *medianamente* consonantes y *medianamente* disonantes –i.e., si la progresión está dada por acordes cuyos vectores de clase de intervalo reducidos son  $[33\ 33]$ ,  $[33\ 33]$ ,  $[33\ 33]$ –, las medidas de consonancia y disonancia se cruzan en el valor 0.5. Manteniendo los primeros dos acordes constantes, las medidas son meramente impactadas por la consonancia y disonancia del último acorde de la progresión.

Además, cabe destacar que, en general, para  $C, D \in \{0, \dots, 66\}$  y  $A_i, A'_i \in \mathcal{C}\mathcal{T}$ ,  $i \in \{2, \dots, k-2\}$  no necesariamente iguales,

$$V_{\mathcal{C}}([C\ D], \mathcal{I}_{n_2}^*(A_2), \dots, \mathcal{I}_{n_{k-2}}^*(A_{k-2}), [D\ C], [C\ D]) = \\ V_{\mathcal{D}}([D\ C], \mathcal{I}_{m_2}^*(A'_2), \dots, \mathcal{I}_{m_{k-2}}^*(A'_{k-2}), [C\ D], [D\ C]) = 0.5$$

puesto que la sucesión de ambas progresiones resultan en la misma fluctuación de consonancia y disonancia respectivamente, con magnitud media. Lo cual da lugar para justificar por qué, en general, las medidas propuestas no contienen en su expresión desde el segundo acorde al  $(k-2)$ -ésimo, para  $k \geq 4$ . Nótese que la medida de consonancia definida corresponde a la normalización de la siguiente expresión

$$\begin{aligned} C_k - D_{k-1} + C_{k-1} - D_1 &= \\ &= (C_k - D_{k-1}) + (C_{k-1} - D_{k-2}) + \dots + (C_2 - D_1) + \quad \left. \vphantom{(C_k - D_{k-1})} \right\} \text{Consonancia acumulada.} \\ &\quad - (C_{k-2} - D_{k-2}) - \dots - (C_2 - D_2) \quad \left. \vphantom{(C_{k-2} - D_{k-2})} \right\} \text{Consonancia ganada entre las progresiones.} \\ &= (C_k - D_{k-1}) + (C_{k-1} - D_{k-2}) + \dots + (C_2 - D_1) + \quad \left. \vphantom{(C_k - D_{k-1})} \right\} \text{Consonancia acumulada.} \\ &\quad + (D_{k-2} - C_{k-2}) + \dots + (D_2 - C_2). \quad \left. \vphantom{(D_{k-2} - C_{k-2})} \right\} \text{Disonancia perdida entre las progresiones.} \end{aligned}$$

Asimismo, para la medida de disonancia,

$$\begin{aligned}
D_k - C_{k-1} + D_{k-1} - C_1 &= \\
&= (D_k - C_{k-1}) + (D_{k-1} - C_{k-2}) + \dots + (D_2 - C_1) + \quad \left. \vphantom{(D_k - C_{k-1})} \right\} \text{Disonancia acumulada.} \\
&\quad - (D_{k-2} - C_{k-2}) - \dots - (D_2 - C_2) \quad \left. \vphantom{(D_{k-2} - C_{k-2})} \right\} \text{Disonancia ganada entre las progresiones.} \\
&= (D_k - C_{k-1}) + (D_{k-1} - C_{k-2}) + \dots + (D_2 - C_1) + \quad \left. \vphantom{(D_k - C_{k-1})} \right\} \text{Disonancia acumulada.} \\
&\quad + (C_{k-2} - D_{k-2}) + \dots + (C_2 - D_2). \quad \left. \vphantom{(C_{k-2} - D_{k-2})} \right\} \text{Consonancia perdida entre progresiones.}
\end{aligned}$$

De manera que las medidas propuestas pueden interpretarse como saldos de consonancia y disonancia, respectivamente, que se dan al ejecutar una progresión de acordes; los cuales son normalizados para tomar valores en  $[0, 1]$  a partir de los hechos de que  $C_k, D_k \in \{0, \dots, 66\}$ ,  $-C_1, -D_1 \in \{-66, \dots, 0\}$  y  $|C_{k-1} - D_{k-1}| \leq C_{k-1} + D_{k-1}$ . Nótese además que, dado que el Campo Aleatorio definido es dinámico, las valuaciones de consonancia y disonancia son llevadas a cabo para cada una de las intervenciones de una obra, pues *las recorre horizontalmente*.

En la Figura 5.4 se muestra la variación de la energía del campo respecto los parámetros  $\underline{\theta}$ , la coloración en cada caso corresponde al valor de la energía de acuerdo a la escala de colores en la leyenda. Obsérvese la Figura 5.4.a), la cual corresponde a la progresión con  $k = 3$  más disonante posible, en tal caso, la energía del campo es determinada meramente por el parámetro  $\theta_{\mathcal{D}}$  puesto que  $V_{\mathcal{C}}(\mathcal{I}^*) = 0$ ; ocurriendo su contrapuesta para el caso de la Figura 5.4.e), en el que la progresión es lo más consonante posible. En 5.4.c), la progresión es tan disonante, como consonante, por lo que el signo de la energía está determinado por el signo de  $\theta_{\mathcal{C}} - \theta_{\mathcal{D}}$ . Asimismo, con el resto de las subfiguras puede observarse que mientras más disonante sea la progresión, mayor impacto tiene el parámetro  $\theta_{\mathcal{D}}$  en la evaluación de la energía de la configuración; mientras que una sucesión de acordes más consonante incurre en mayor impacto de  $\theta_{\mathcal{C}}$  en  $\varepsilon(x)$ . Por lo tanto, puede interpretarse a los parámetros  $\underline{\theta}$  como la fuerza de incurrencia, factor o noción de consonancia y disonancia en una obra musical dada, o bien, un autor o corriente musical.

Finalmente, la medida de Gibbs del campo está dada por

$$\pi(x) = \frac{1}{Z(\theta_{\mathcal{C}}, \theta_{\mathcal{D}})} \cdot \exp\{-[\theta_{\mathcal{C}} \cdot V_{\mathcal{C}}(x) - \theta_{\mathcal{D}} \cdot V_{\mathcal{D}}(x)]\}.$$

Como los parámetros  $\underline{\theta}$  son no negativos, mayor energía en una configuración es equivalente a menor probabilidad de ocurrencia. Nótese que la energía del campo está dada por las diferencias entre sus medidas de consonancia y disonancia ponderadas por la *concepción* de consonancia y disonancia realizada en la pieza, por lo que menor energía en una configuración implica que la progresión de acordes fue casi tan consonante como disonante, i.e., equilibrada<sup>12</sup>, lo cual es concordante con las maneras musicales y el Principio Aristotélico de la imperfección a la perfección.

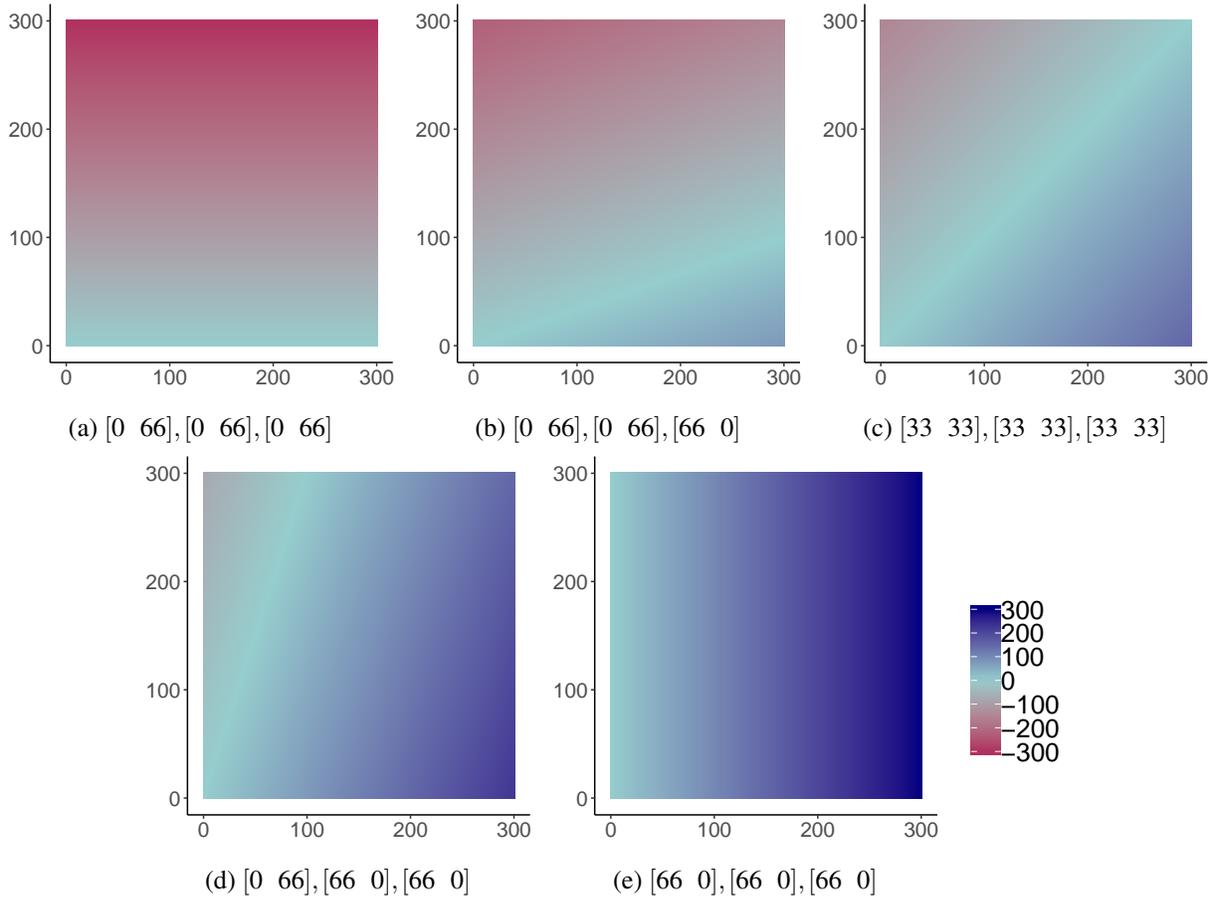


Figura 5.4: Energía del campo respecto las variaciones en los parámetros  $\theta_{\mathcal{C}}$  (eje x) y  $\theta_{\mathcal{D}}$  (eje y) para las progresiones especificadas en cada inciso ( $k = 3$ ).

Claramente el Campo Aleatorio definido es un Campo de Gibbs y un Campo de Markov. Primero, todos los Campos de Gibbs son Markovianos de acuerdo al Teorema 3.2. Luego, por construcción, el espacio de configuraciones  $\Lambda^S$  es finito (de hecho,  $\#(\Lambda^S) = 67^{k-2}$ ) y su sistema de

<sup>12</sup>Tanto en el contexto musical, como (por lo tanto) en la construcción de las medidas propuestas, una progresión de acordes no puede ser consonante y disonante simultáneamente, más que en valores cercanos a 0.5.

vecindades es trivial, i.e., es un Campo de Markov. Adicionalmente, nótese que esto no contradice el Teorema 3.3 ya que  $\pi$  satisface la condición de positividad, pues para todo  $x \in \Lambda^S$ ,  $\varepsilon(x) < \infty$  por construcción, de donde se sigue que para toda configuración  $x$ ,  $\pi(x)$  es estrictamente mayor que cero, cumpliéndose por vacuidad la condición de positividad. Además, por el Teorema 3.1, la medida de Gibbs y distribución del Campo,  $\pi$ , es única, por lo que el campo es ergódico.

La estimación de los parámetros  $\underline{\theta}$  fue llevada a cabo maximizando la función de log-Verosimilitud expresada analíticamente en la Sección 3.3.2. y mediante el Algoritmo Evolutivo implementado para vecindades tipo  $\beta$ , donde el conjunto de sitios corresponde al producto cruz de la partición de cardinalidad 3,000 del conjunto  $[0, 300]$  consigo misma. Se llevó a cabo la optimización con probabilidad de transición a estados vecinos de 0.7 y 1,000 iteraciones para cada obra individualmente; así como para corriente musical histórica y autor mencionados en la sección anterior, notando que al asumir que cada pieza es independiente de las otras, la función de log-Verosimilitud en cada caso corresponde a la suma de las funciones de log-Verosimilitud de cada obra. Cabe hacer notar que las medidas planteadas no están definidas para notas, ya que mediante ellas se pretende cuantificar la disonancia y consonancia causadas por sonidos simultáneos, por lo que para esta modelación se llevó a cabo la transformación de las piezas descritas al inicio de este capítulo, con la diferencia de que se excluyen de las intervenciones las notas *individuales*, así como los silencios. Una descripción detallada del procedimiento llevado a cabo se encuentra en el *Anexo II*.

A pesar de que la utilización del Algoritmo Evolutivo reduce considerablemente el tiempo requerido de compilación de la optimización en comparación a otros métodos, es necesario establecer la complejidad computacional que implica obtener dichas estimaciones: en cada iteración, es necesario evaluar la constante de normalización de la medida de Gibbs

$$Z(\theta_{\mathcal{C}}, \theta_{\mathcal{D}}) = \sum_{\lambda \in \Lambda^S} \exp\{-\varepsilon(\lambda)\};$$

donde  $\#(\Lambda^S) = 67^{k \cdot 2}$ . Es decir, escogiendo el valor mínimo posible  $k = 2$ , sólo para valuar  $Z(\theta_{\mathcal{C}}, \theta_{\mathcal{D}})$  para un par ordenado  $\underline{\theta}$ , es necesario calcular  $67^4 = 20,151,121$  multiplicaciones. Aún sin mencionar el resto de los cálculos para obtener la evaluación de la función de log-Verosimilitud, esto resulta en una cantidad exorbitante de tareas a realizar puesto que tal procedimiento debe realizarse una cantidad *suficientemente grande* de veces<sup>13</sup> para la optimización de una sola obra, lo cual crece aún más a razón de que el valor escogido para  $k$  lo hace.

<sup>13</sup>El parámetro utilizado para la estimación fue de 1,000 iteraciones.

La medida tomada para reducir el tiempo de compilación computacional de las estimaciones fue redefinir el espacio de configuraciones a considerar por únicamente las fases observadas en la base de datos *Classical Archives Corpus* completa [36] –la cual cuenta con una cantidad total de 8,980 obras–, bajo la justificación de que las entradas en conjunto de un vector de clase de intervalo reducido pueden tomar  $67^2 = 4,489$  diferentes combinaciones en teoría, pero su uso práctico real puede diferir considerablemente. De hecho, los vectores de clase de intervalo reducidos que verdaderamente se presentan en toda la base de datos son únicamente 43; entonces considerando el espacio de configuraciones

$$\Lambda^S = \{(7,3), (6,9), (15,13), (10,11), (3,0), (2,1), (13,8), (5,1), (3,7), (0,3), (12,9), (36,30), (1,2), (11,10), (3,3), (5,5), (17,11), (6,3), (1,5), (14,14), (12,3), (6,4), (8,2), (19,17), (13,15), (8,7), (4,2), (1,0), (30,25), (9,6), (16,12), (0,1), (10,5), (21,15), (4,6), (14,7), (24,21), (25,20), (18,18), (7,8), (2,4), (9,12), (20,16)\}^k,$$

tales cálculos por iteración por obra pasan de  $67^{k \cdot 2}$  a ser  $43^k$ .

Por último, debe destacarse que llevar a cabo la modelación de mejor manera significaría hacerlo a través de los vectores de clase de intervalo (no reducidos)<sup>14</sup>, puesto que hace distinción entre los distintos tipos de intervalo disonantes y consonantes. No obstante, ¡la compilación computacional análoga a la descrita anteriormente aumentaría a  $67^{6 \cdot 2} = 8,182,718,904,632,857,144,561$  cálculos por iteración y por pieza! Es decir, realizar estimaciones bajo esta modelación es prácticamente imposible en la actualidad, computacionalmente hablando. He aquí la necesidad de la simplificación del modelo mediante los vectores de clase de intervalo reducidos: un costo grande, puesto que claramente tal representación no es única para cada acorde, a diferencia del vector de clase de intervalo *tradicional* en la mayoría de los casos en vista de que las clases de conjunto Z-relativas son poco comunes (de hecho, no fueron observadas en absoluto mediante la base de datos).

---

<sup>14</sup>Definición 1.10.

### 5.3. Análisis exploratorio de las estimaciones

Para cada obra musical incluida en la base de datos fueron estimadas por Máxima Verosimilitud las probabilidades límite de las Cadenas de Markov formuladas, así como los parámetros  $\theta$  correspondientes a los Campos Aleatorios secuenciales definidos en la sección anterior. Esto fue llevado a cabo de manera *individual* para cada observación en la base, por corriente musical histórica y por autor en los casos en que fue posible dada la cardinalidad de las piezas en la base: Bach, Beethoven, Chopin, Handel, Haydn, Mozart, Palestrina y Vivaldi. Para una descripción detallada de los aspectos de la implementación sobre una obra musical en específico, consúltese el *Anexo II*.

Cabe destacar que las estimaciones de las probabilidades límite para las Cadenas de Markov llevadas a cabo son consistentes en el sentido de que, como sugieren las siguientes figuras, hay variaciones en las probabilidades obtenidas pero no difieren fundamentalmente al compararlas por autor. Esto no ocurre necesariamente por corriente musical histórica y, en realidad, a pesar de que dos autores pertenecieran al mismo momento histórico, desde luego que sus obras podrían ser marcadamente diferentes. Nótese que esta observación brinda información a priori de la clasificación por corrientes musicales, de manera que se espera que el clasificador por autor sea mucho más eficiente.

En la Figura 5.5 se muestran las estimaciones de las probabilidades límite correspondientes a las Cadenas de Markov asociadas a las duraciones de las intervenciones de cada obra musical individualmente, i.e., se presentan los 1,456 vectores estimados. En primera instancia, se hace clara la diferencia entre las duraciones utilizadas en el Renacimiento, respecto al resto de las corrientes históricas musicales. En concordancia con la Musicología, en tal periodo se tendía a utilizar duraciones mucho más largas; a tal grado que las entradas del vector de probabilidades límite correspondientes a duraciones menores a la de la corchea son prácticamente cero.

Luego, desde el Barroco se evidencia el uso fundamentalmente de las corcheas; siendo a su vez menos común el uso de largas duraciones. Cabe destacar que la evidencia cuantitativa encontrada sugiere un uso relativamente uniforme de las duraciones de las intervenciones por los autores: Chopin, Liszt, Tchaikovsky, Joplin y Rachmaninov. Mismos que, junto con Vivaldi y Bach, mostraron mayor tendencia que el resto a utilizar las duraciones menos usuales englobadas en la categoría *Otra*.

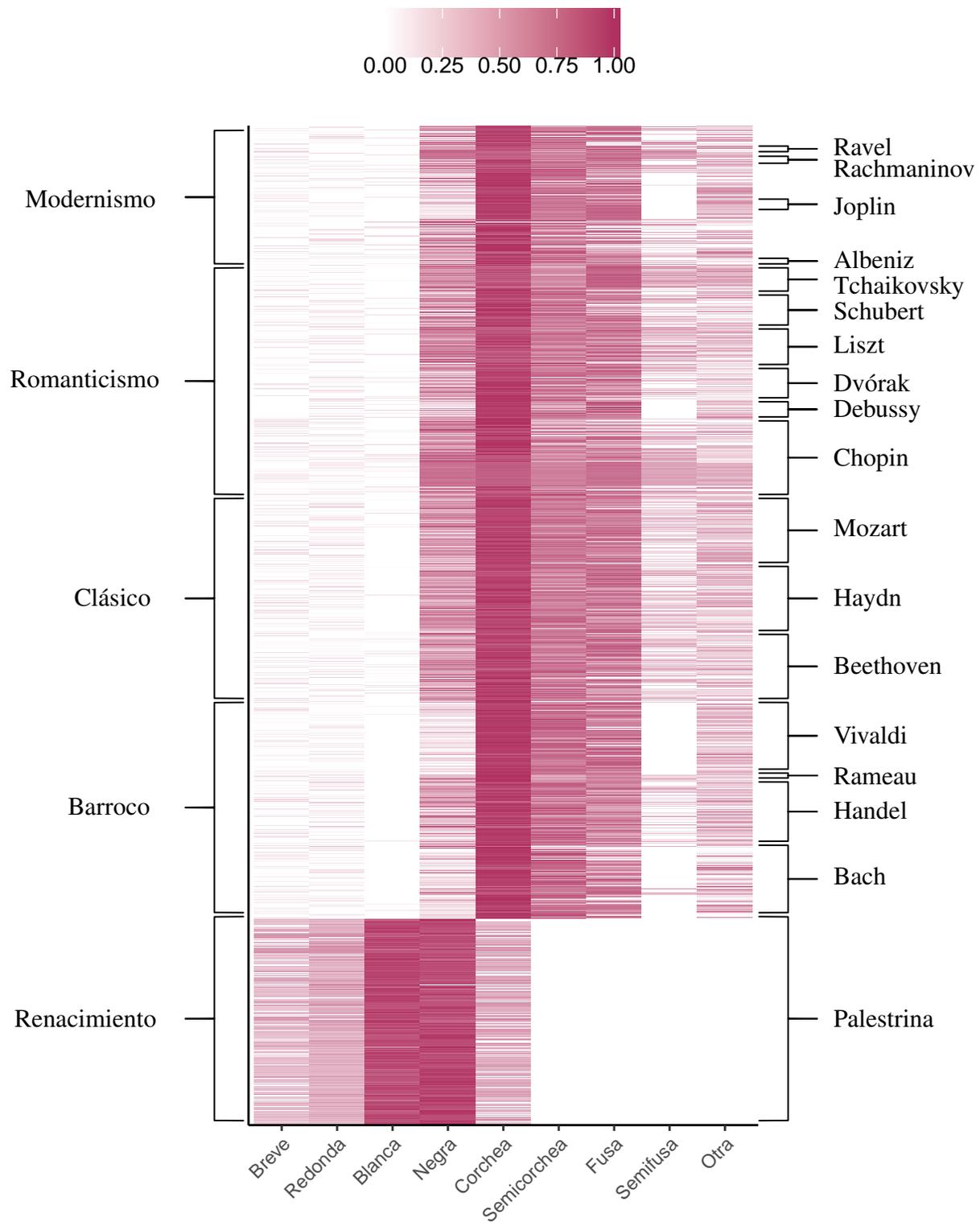


Figura 5.5: Estimaciones de las probabilidades límite de las Cadenas de Markov  $\{D_n\}_{n \in \mathbb{N}}$  asociadas a las duraciones de las intervenciones de las obras musicales en la base de datos.

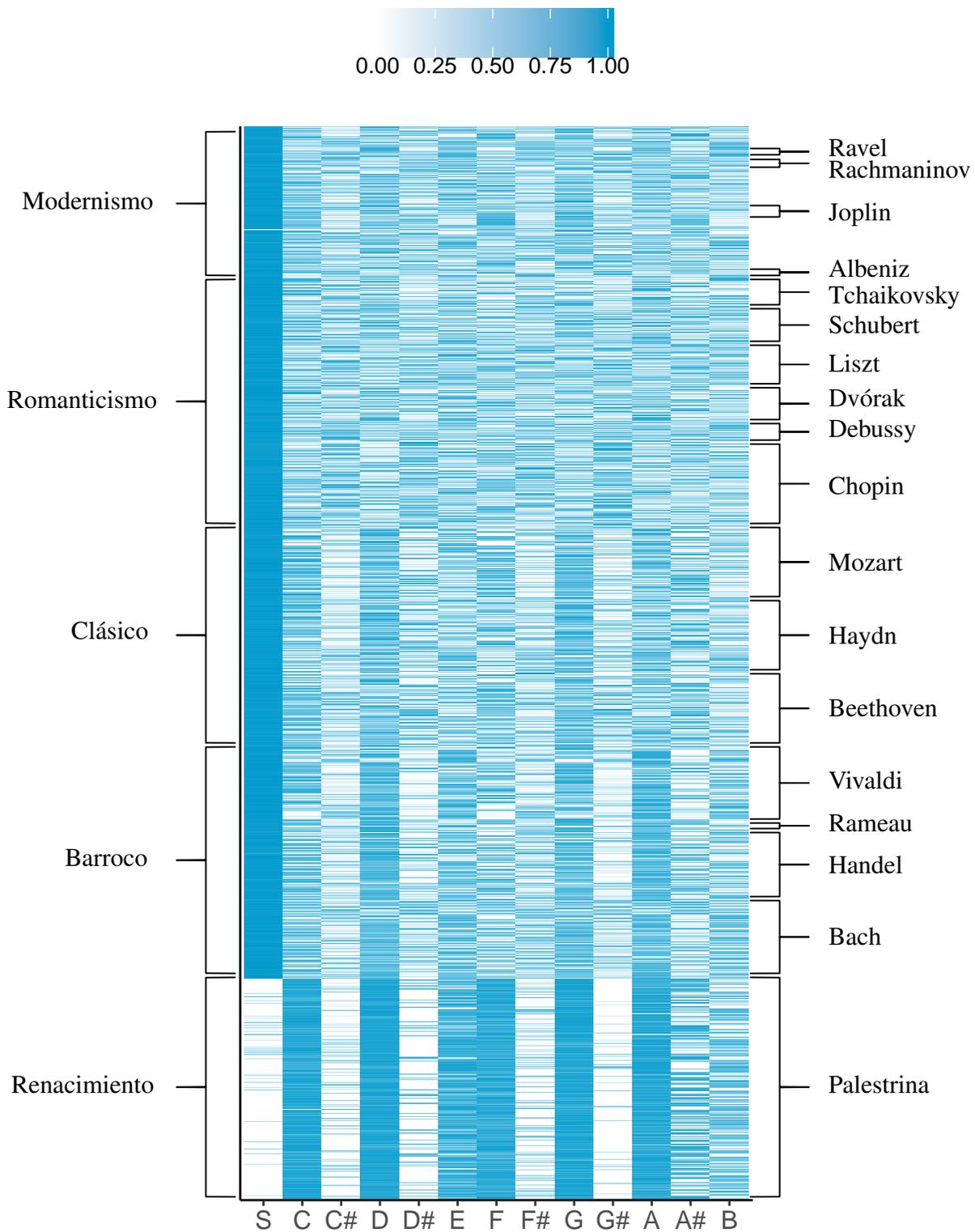


Figura 5.6: Estimaciones de las probabilidades límite de las Cadenas de Markov  $\{R_n\}_{n \in \mathbb{N}}$  asociadas a las raíces de las intervenciones de las obras musicales en la base de datos.

Por otro lado, en concordancia con lo anterior, puede apreciarse en las figuras 5.6 y 5.7 la diferencia abismal entre la construcción de acordes antes del Barroco, periodo en el que se gestó la Teoría de la Armonía. Antes que nada, nótese que el uso de los silencios en el Renacimiento era mucho menor, las entradas de las probabilidades límite asociadas a ello (denotadas en la Figura 5.6 por  $S$ , y en la Figura 5.7 por  $0 - 1$ ) son casi nulas, posteriormente se evidencia un uso mucho más frecuente hasta la llegada del Modernismo, punto a partir del cual decrece ligeramente.

Adicionalmente, en la música Renacentista se hacía uso principalmente de las notas naturales como raíces de los acordes, lo cual fue diluyéndose al pasar el tiempo a razón de que se vuelven relativamente más uniformes las entradas de los vectores de las probabilidades límite, comparando del Barroco en adelante. Sin embargo, se refleja una clara preferencia por las notas naturales, excepto tal vez en el Romanticismo, periodo en el que, de hecho, las entradas asociadas a las accidentales son mayores que las naturales en la mayoría de los casos<sup>15</sup>. Asimismo, nótese que Mozart, Haydn, Vivaldi, Rameau, Handel y Bach muestran una tendencia al uso de las notas *Sol*, *La* y *Re* como notas fundamentales.

Como se puede apreciar en la Figura 5.7, las probabilidades límite asociadas a la Clase Forte de las intervenciones en las obras musicales reflejan lo atípico que es el uso de acordes con más de 5 notas, debido a que son más disonantes en general. De acuerdo con los resultados obtenidos, las Clases Forte más utilizadas a largo plazo en una pieza posteriormente al Renacimiento son: el silencio ( $0-1$ ); una sola nota ( $1-1$ ); pares de notas con intervalos de tres, cuatro y cinco semitonos<sup>16</sup> ( $2-3$ ,  $2-4$ ,  $2-5$ ); desde luego las triadas mayores y menores ( $3-11$ ), disminuidas ( $3-10$ ) y séptimas incompletas –mayores ( $3-4$ ), menores ( $3-7$ ) y dominantes ( $3-8$ )–.

En el Renacimiento se acentúa la aparición de las clases mencionadas anteriormente, a las que se suman el tritono<sup>17</sup> ( $3-9$ ), tetracordes ( $4-14$ ,  $4-22$ ) y séptimas mayores ( $4-20$ ), menores ( $4-26$ ) y semidisminuidas ( $4-27$ ). Además, es claro que al pasar el tiempo se fue optando por disminuir el uso de intervenciones formadas por dos notas únicamente.

---

<sup>15</sup>Lo cual es totalmente concordante en el contexto debido a que tal corriente musical se caracterizó, entre otros aspectos, por la contraposición a los usos musicales anteriores.

<sup>16</sup>Recuérdese que tales intervalos construyen relaciones consonantes, desde el enfoque contrapuntístico, entonces es natural que sean los de mayor uso.

<sup>17</sup>Probablemente como respuesta social a la cualidad *diabólica* (*Diabolus in Musica*) atribuida a este tipo de intervalos en la Edad Media e inicio del Renacimiento. Recuérdese que las obras musicales incluidas en la base de datos correspondientes al periodo Renacentista fueron únicamente compuestas por Giovanni Pierluigi da Palestrina, quien vivió hacia finales del Renacimiento.

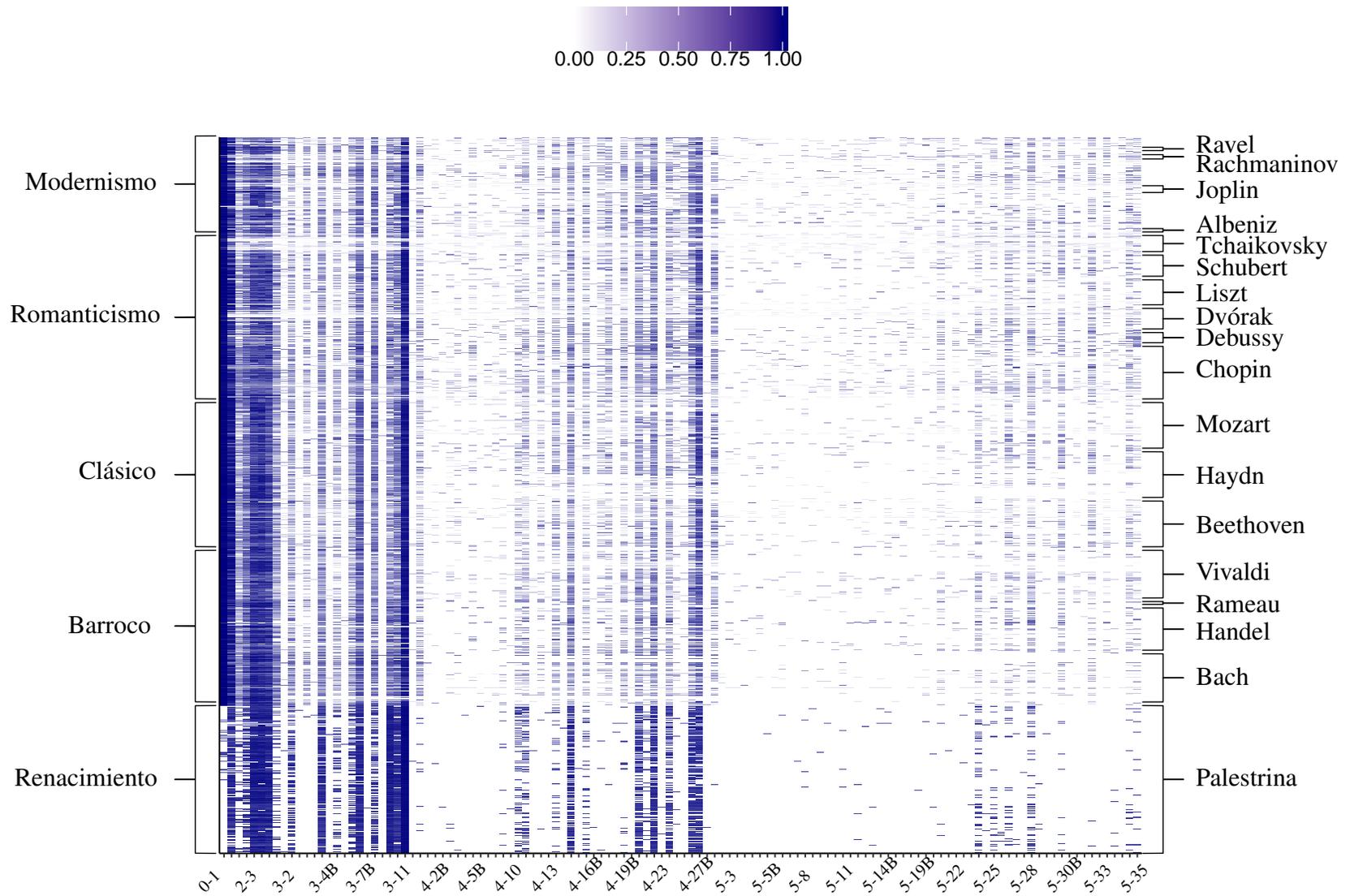


Figura 5.7: Estimaciones de las probabilidades límite de las Cadenas de Markov  $\{C_n\}_{n \in \mathbb{N}}$  asociadas a la Clase Forte (con menos de seis notas) de las intervenciones de las obras musicales en la base de datos.

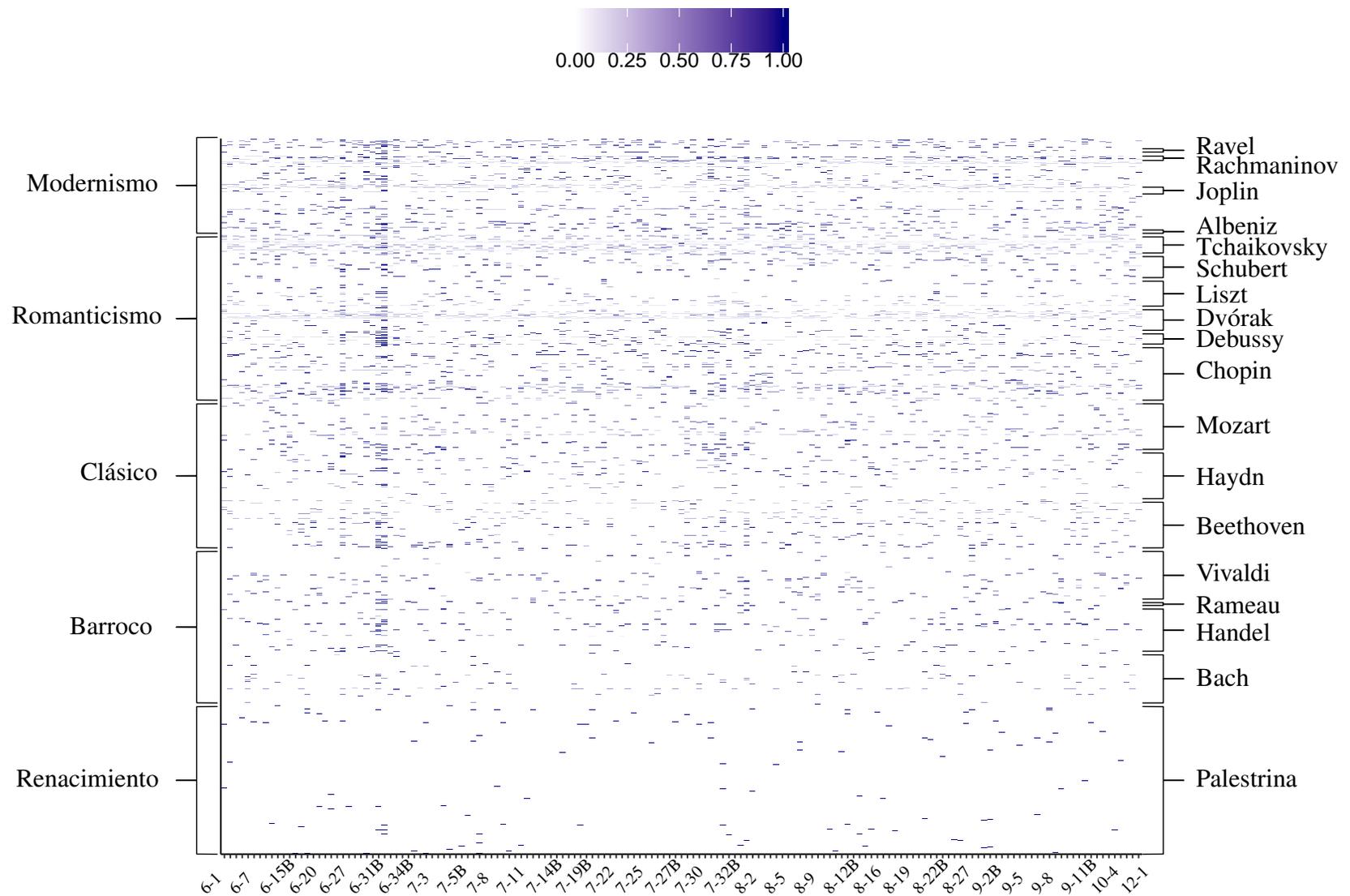


Figura 5.8: Estimaciones de las probabilidades límite de las Cadenas de Markov  $\{C_n\}_{n \in \mathbb{N}}$  asociadas a la Clase Forte (con más de seis notas) de las intervenciones de las obras musicales en la base de datos.

Hacia el Romanticismo aumenta la aparición de acordes más *complejos*; no obstante, sigue siendo poco habitual su uso (por su disonancia). En específico, puede observarse en la Figura 5.8 que Dvorak, Tchaikovsky, Joplin, Rachmaninov y Ravel utilizan en sus composiciones intervenciones con seis o más sonidos simultáneos más frecuentemente que el resto de los autores contemplados.

Por otro lado, se llevó a cabo un Análisis de Conglomerados Jerárquico sobre las variables formadas por las entradas de las probabilidades límite con el fin de determinar cómo impacta a largo plazo el uso de una duración, raíz o Clase Forte sobre el uso del resto. Para ello fue utilizada como función de enlace *Complete Linkage* y como medida de disimilitud la siguiente métrica

$$d(X_i, X_j) = 1 - |\hat{\rho}_{X_i, X_j}|;$$

donde  $\hat{\rho}$  corresponde a la estadística *Rho de Spearman*.

Los hallazgos al aplicar tal método son representados en la Figura 5.9. Primero, sugiere la separación de las entradas de las probabilidades límite asociadas a las duraciones fusa y semifusa del resto casi inmediatamente, a la altura de 0.99. Al llegar a la altura 0.86, la jerarquía se bifurca una vez más, separando las duraciones englobadas por *Otra*. Y, finalmente, las duraciones restantes se separan de las semicorcheas en 0.58. Esto implica que, el uso de las fusas impacta en gran medida el uso de las semifusas en una pieza musical, e inversamente; i.e., el mayor uso de fusas causa un aumento en la probabilidad del uso de semifusas a largo plazo en la obra. Los mismos efectos sobre el resto de las duraciones son obtenidos al aplicar el razonamiento análogo a los cortes en el dendrograma a las alturas mencionadas anteriormente. En general, puede concluirse que el uso de las notas, acordes y silencios largos es equivalente a que en la obra se tienda a no utilizar duraciones cortas; puesto que mientras más probable sea la aparición de grandes duraciones habiendo alcanzado el punto estacionario en la armonía de la pieza, menor es la probabilidad del uso de las cortas y viceversa.

Luego, nótese que el dendrograma resultante de la aplicación del Análisis de Conglomerados a las cadenas asociadas a las notas fundamentales de los acordes agrupa a las entradas de los vectores de las probabilidades límite correspondientes a: Do y Sol, Do#-Re♭ y Sol, Do#-Re♭ y Sol#-La♭, Re y La, Re#-Mi♭ y La, Mi y La, Fa y La#-Si♭, Fa y Si, Fa#-Sol♭ y La#-Si♭, Fa#-Sol♭ y Si; aquellas notas cuyos intervalos corresponden a cuartas y quintas justas, y tritonos. Esto significa que el uso en específico de una nota como tónica a lo largo de la pieza hace más probable el uso de los grados

subdominante (IV) y dominante (V) relativos a ellas, por lo cual se sostiene que la propuesta hecha modela adecuadamente los grados de los acordes utilizados en una obra musical, a razón de que refleja la progresión armónica paradigmática: I-IV-V-I.

En el inciso c) de la figura se muestra el dendrograma correspondiente a la cadena que denota las Clases Forte. La coloración corresponde a los cinco más grandes agrupaciones, las cuales ocurren respectivamente a las alturas de corte: 0.99998 (separa los grupos mostaza y gris del resto), 0.99995 (separa el violeta de la unión de los grupos magenta y turquesa), 0.9997 (causa la bifurcación entre los grupos magenta y turquesa) y 0.998.

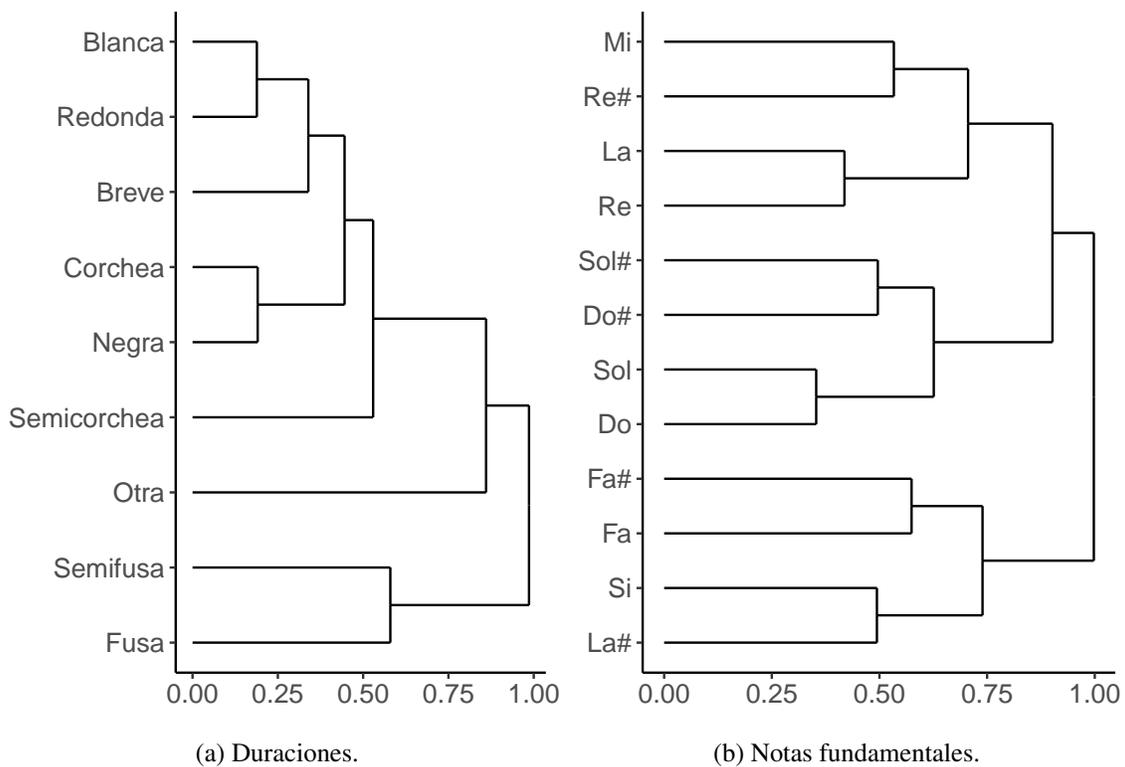
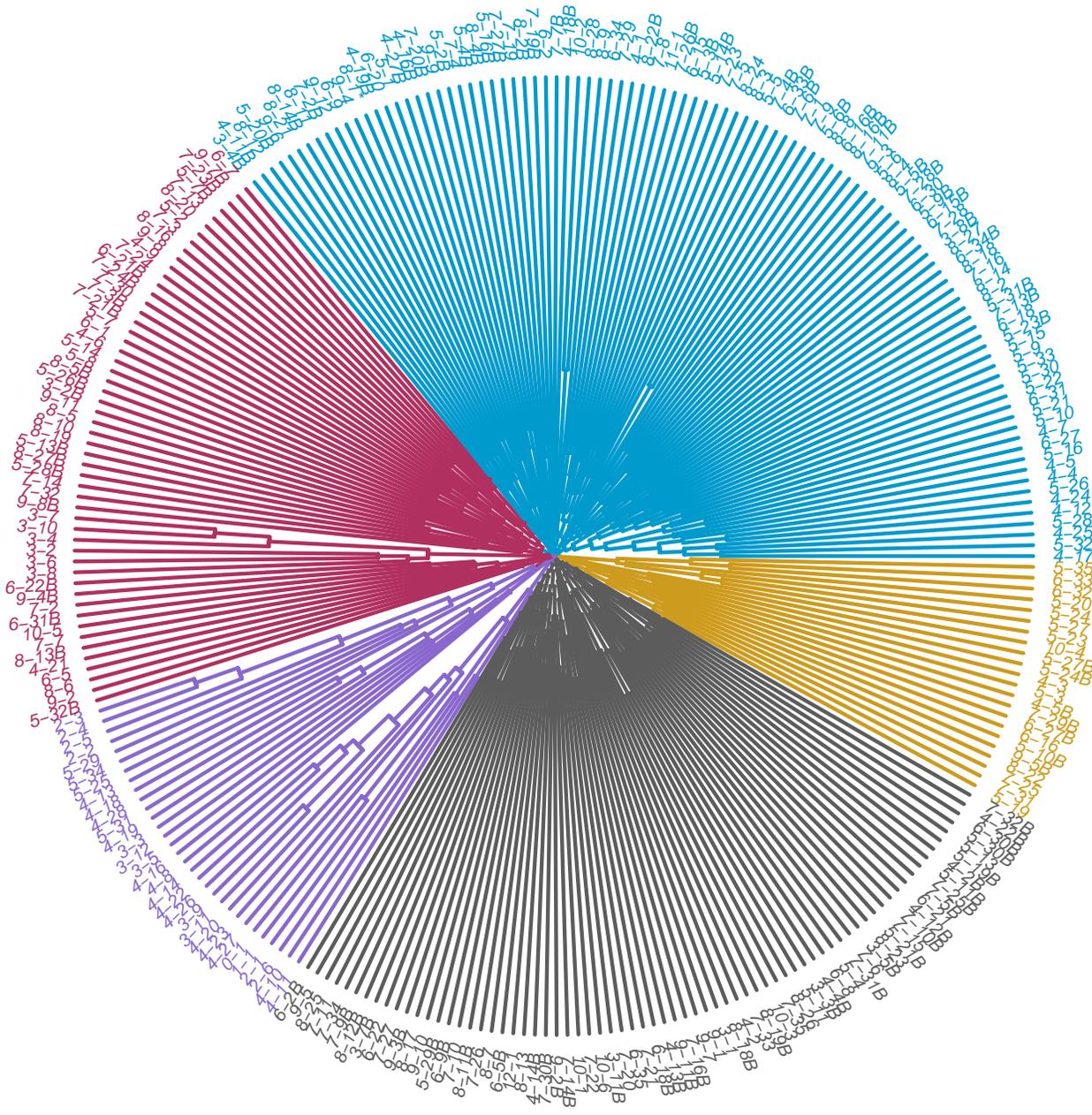


Figura 5.9: Dendrogramas resultantes del Análisis de Conglomerados jerárquico de las probabilidades límite estimadas de las Cadenas de Markov asociadas a las duraciones, notas fundamentales y Clase Forte.



(c) Clases Forte.

Figura 5.9: Dendrogramas resultantes del Análisis de Conglomerados jerárquico de las probabilidades límite estimadas de las Cadenas de Markov asociadas a las duraciones, notas fundamentales y Clase Forte.

La evaluación de las funciones de Verosimilitud para la modelación mediante Campos Aleatorios, aún maximizadas, resulta en cifras relativamente pequeñas debido a que involucra la multiplicación en serie de una cantidad considerable de probabilidades (tantas como subsecuencias de longitud  $k$  tenga una pieza correspondientes a sus intervenciones sucesivas). Razón por la cual se presentan los gráficos de caja y de violín de las estimaciones de las funciones de log-Verosimilitud,

donde se sugiere que se cuenta con 15 datos atípicos<sup>18</sup>: una obra de Beethoven, dos de Chopin, dos de Debussy, cuatro de Liszt, tres de Ravel, una de Schubert y una de Tchaikovsky.

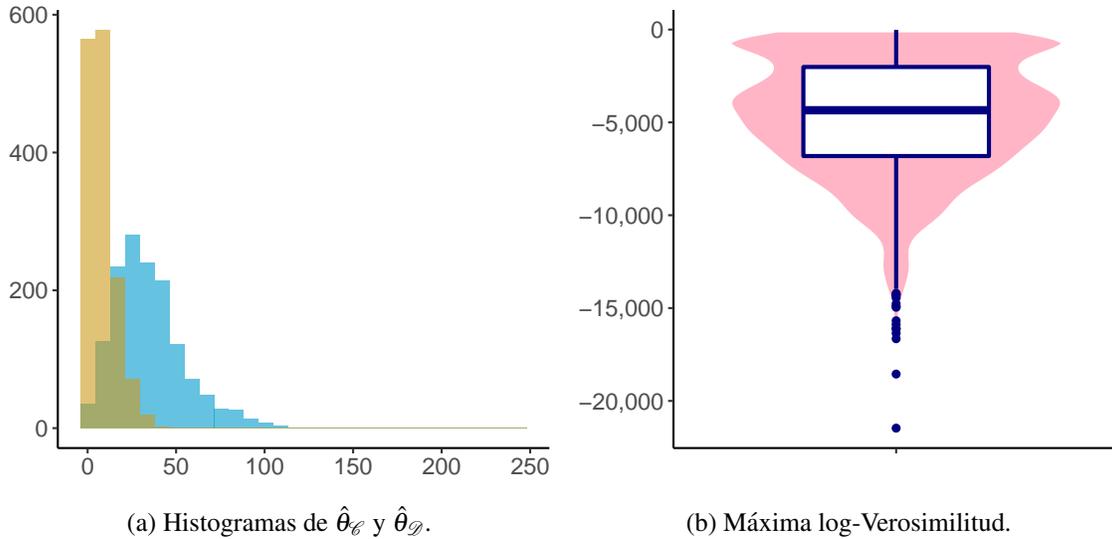


Figura 5.10: Histogramas de los valores estimados  $\hat{\theta}_\epsilon$  (turquesa) y  $\hat{\theta}_\phi$  (mostaza); y gráfico de caja y violín de las estimaciones de la función de log-Verosimilitud de las obras musicales de la base de datos, para  $k = 2$ .

En general, los parámetros de noción de disonancia estimados resultaron ser menores que aquellos asociados a la consonancia, como se da a notar en la Figura 5.10.a. La media empírica de  $\hat{\theta}_\epsilon$  resultó ser de 34.56 y su valor máximo observado de 244.01. Asimismo, la media obtenida de los parámetros  $\hat{\theta}_\phi$  es de 7.99 y su máximo 47.38. El valor mínimo observado en ambos casos fue de cero:

- En el *Concerto 532 I en Sol Mayor* de Vivaldi para  $\hat{\theta}_\epsilon$ .
- En el caso de  $\hat{\theta}_\phi$ , se anularon las estimaciones para *French Overture BWV831 Passapied B minor* de Bach, *Bagatelles Op 119 I G minor* de Beethoven, *Poetic Tone Pictures 5 Peasant Ballad Op85 Bb major* de Dvórák, *String Quartet No11 K171 ii Eb major* de Mozart, así como en *Gloria 19 C* y *Gloria 68 C* de Palestrina.

La correlación lineal muestral entre las estimaciones de los parámetros  $\underline{\theta}$  resultó ser de  $-0.01$ , lo que puede interpretarse como que la noción de consonancia y disonancia de cada autor son determinadas independientemente una de la otra. En la figura siguiente se muestran las estimaciones de

<sup>18</sup>Los cuales, cabe hacer notar, no corresponden a los casos en que los parámetros  $\underline{\theta}$  estimados se anulan.

$\hat{\theta}_{\mathcal{C}}$  contra  $\hat{\theta}_{\mathcal{D}}$  para cada obra, autor y corriente, con coloración respectiva a cada obra o corriente, según sea el caso<sup>19</sup>. En primera instancia, los parámetros estimados para cada autor y corriente histórica mostrados parecen corresponder con las estimaciones individuales de los parámetros. Luego, el grupo que se nota claramente distinto a los otros, una vez más, es el correspondiente al Renacimiento (Palestrina), cuyas nociones de disonancia y consonancia son respectivamente menores y mayores en comparación a los demás. Respecto el resto de las clases, se tienen grupos mucho más mezclados, no se perciben a simple vista hiperplanos que los separen.

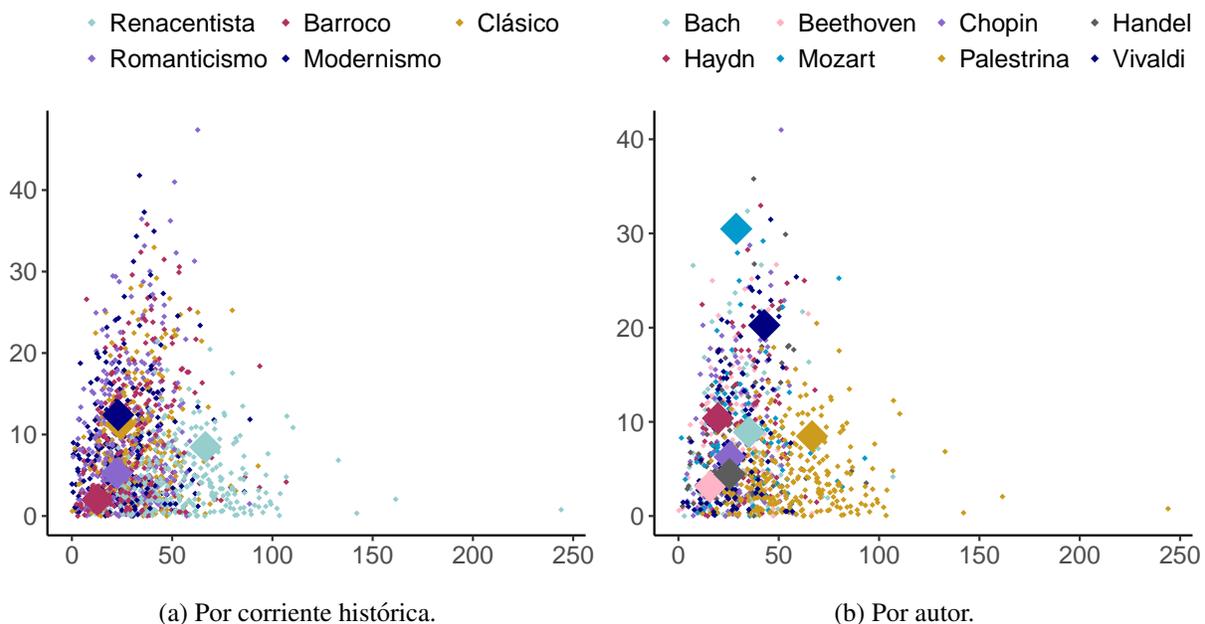


Figura 5.11: Gráficos de dispersión de los parámetros estimados  $\hat{\theta}_{\mathcal{C}}$  (eje x) y  $\hat{\theta}_{\mathcal{D}}$  (eje y) con  $k = 2$ , individualmente y por corriente histórica y por su autor. Los puntos de mayor tamaño corresponden a las estimaciones de los parámetros dado que se pertenece a alguna corriente histórica en el inciso a, y dado que la obra fue escrita por determinado autor en el inciso b. Los puntos pequeños corresponden a las estimaciones llevadas a cabo *individualmente*.

Es momento de dar a saber que se presentarán y utilizarán únicamente las estimaciones obtenidas para  $k = 2$  en medida de que la obtención de resultados mediante este segundo enfoque fue consistente, lo cual se argumenta a continuación. En el siguiente cuadro se presentan las cinco obras musicales en la base de datos cuyas log-Verosimilitudes estimadas<sup>20</sup> cambiaron más al variar el parámetro del campo secuencial  $k \in \{2, 3, 4\}$ . El valor absoluto de la diferencia más grande es

<sup>19</sup>Los valores específicos de los parámetros por autor y corriente se presentan en el Cuadro 5.4 y el Cuadro 5.5.

<sup>20</sup>Inicialmente se llevaron a cabo las estimaciones correspondientes a 405 obras debido a alto costo computacional que deriva de la modelación.

de  $1.8 \cdot 10^{-14}$  y el 53.1 % de las obras contempladas presentaron diferencias tan pequeñas que se perdieron por la precisión de la máquina. Nótese que claramente los parámetros  $\hat{\theta}$  difieren *bastante* a distintos valores de  $k$ ; sin embargo, esto no implica lo contrario a lo sostenido, difieren debido a que corresponden a conceptualizaciones distintas en cuanto a la dependencia de las intervenciones de una pieza. Mediante el enfoque de Campos Aleatorios se está caracterizando a una obra musical a través de su medida de Gibbs, de la cual deriva por completo su función de Verosimilitud como fue descrito en el Capítulo 3. Además, como es evidenciado en la Figura 1.13, incisos d al f, a medida que aumenta el valor de  $k$ , parecería ser menor el cambio en las medidas de consonancia y disonancia.

$\hat{\theta}_{\mathcal{G}_2}$	$\hat{\theta}_{\mathcal{G}_3}$	$\hat{\theta}_{\mathcal{G}_4}$	$\hat{\theta}_{\mathcal{D}_2}$	$\hat{\theta}_{\mathcal{D}_3}$	$\hat{\theta}_{\mathcal{D}_4}$	$\mathcal{L}(\underline{X}; \hat{\theta})_2$	$\mathcal{L}(\underline{X}; \hat{\theta})_3$	$\mathcal{L}(\underline{X}; \hat{\theta})_4$
122.8	55.5	114.0	$1.5 \cdot 10^{-8}$	0.2	7.1	$1.8 \cdot 10^{-14}$	$2.7 \cdot 10^{-19}$	$2.7 \cdot 10^{-23}$
130.2	63.9	89.3	$4.5 \cdot 10^{-7}$	1.2	0.3	$8.1 \cdot 10^{-27}$	$1.8 \cdot 10^{-25}$	$2.9 \cdot 10^{-26}$
121.9	47.2	54.4	$1.9 \cdot 10^{-11}$	1.9	0.1	$2.7 \cdot 10^{-39}$	$1.3 \cdot 10^{-25}$	$4.8 \cdot 10^{-26}$
53.2	44.7	89.8	$1.1 \cdot 10^{-6}$	0.5	7.0	$2.7 \cdot 10^{-29}$	$1.7 \cdot 10^{-34}$	$1.1 \cdot 10^{-39}$
133.4	41.8	49.1	$5.6 \cdot 10^{-7}$	1.6	14.6	$1.7 \cdot 10^{-56}$	$1.3 \cdot 10^{-31}$	$1.0 \cdot 10^{-37}$

Cuadro 5.3: Estimaciones de  $\theta_{\mathcal{G}}$ ,  $\theta_{\mathcal{D}}$  y de la evaluación de la función de log-Verosimilitud para las cinco obras musicales con mayores diferencias entre las valuaciones de su Verosimilitud estimadas al variar el orden del modelo.

Finalmente, se llevó a cabo un Análisis de Componentes Principales incluyendo como variables todas las estimaciones aquí presentadas, la proporción de varianza explicada y su acumulación se muestran en la Figura 5.12. Con el octavo Componente Principal se da un decremento marcado de la proporción de varianza explicada, sin embargo, con éste se acumula apenas el 21 % de varianza explicada. Con los primeros 204 Componentes Principales (de 302) se explica el 90 % de la varianza, con los primeros 239 se alcanza el 95 % y hasta el 280 se obtiene el 99 %.

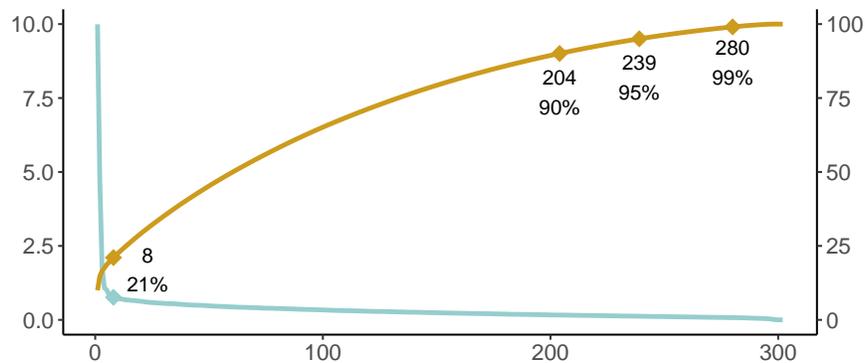


Figura 5.12: Porcentaje de varianza explicada y su acumulación respecto al número de Componentes Principales. El eje izquierdo corresponde a la varianza explicada (línea turquesa) y el derecho a su acumulación.

Nótese que tiene sentido que pueda reducirse la dimensionalidad del espacio de variables por su naturaleza. Para empezar, las entradas estimadas de los vectores de las probabilidades límite deben sumar cero, i.e., cada una es combinación lineal de las otras para que se cumpla la condición  $\pi_i = 1 - \sum_{j \in \mathcal{S} \setminus i} \pi_j$ . Adicionalmente, por construcción de la medida de Gibbs, es natural que las estimaciones de los parámetros  $\underline{\theta}$  estén correlacionadas con la evaluación de la función de log-Verosimilitud de cada obra.

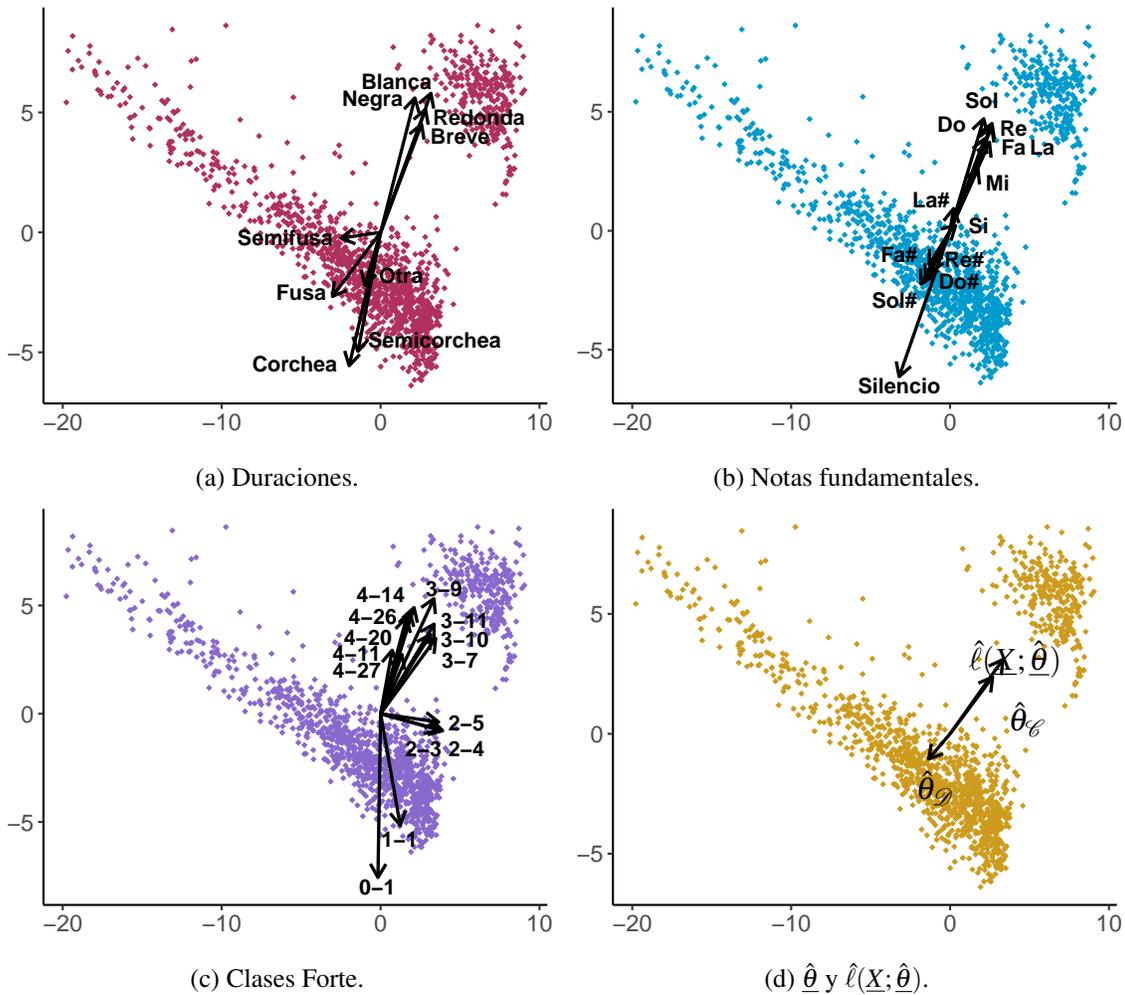


Figura 5.13: *Biplots* para los primeros dos Componentes Principales por estimaciones: parámetros  $\theta_{\mathcal{C}}$  y  $\theta_{\mathcal{D}}$ , evaluación de la función de log-Verosimilitud  $\hat{\ell}(\underline{X}; \hat{\theta}_{\mathcal{C}}, \hat{\theta}_{\mathcal{D}})$  y probabilidades límite asociadas a duraciones, notas fundamentales y Clases Forte (variables cuyos valores máximos son no menores a 0.1). Por presentación, los vectores de cargas fueron reescalados a 30 veces su magnitud.

La Figura 5.13 muestra resultados afines al Análisis de Conglomerados jerárquico aplicado a las variables<sup>21</sup>. Se tiene evidencia de que el uso de duraciones cortas en las intervenciones de una obra es equivalente a la aversión a las duraciones largas. También, en 5.13.b se nota el comportamiento análogo para las notas accidentales y naturales como raíces de los acordes en las intervenciones. Las Clases Forte proyectadas a los primeros dos Componentes Principales muestran que el número de notas que conforman las intervenciones están correlacionadas entre sí mientras sean iguales, a pesar de los intervalos que las componen. Y, finalmente, se muestra la clara dependencia lineal entre los parámetros asociados a las nociones de consonancia y disonancia y la evaluación de la función de log-Verosimilitud de cada pieza, siendo claramente el factor de disonancia inversamente proporcional a la misma.

### 5.3.1. Análisis de corrientes históricas

Como fue mencionado anteriormente, fueron estimadas las probabilidades límite asociadas a las Cadenas de Markov propuestas por corriente histórica a través de la igualdad

$$\mathcal{L}(\underline{X}; P_j) = \prod_{i \in I_j} \mathcal{L}(X_i; P_j);$$

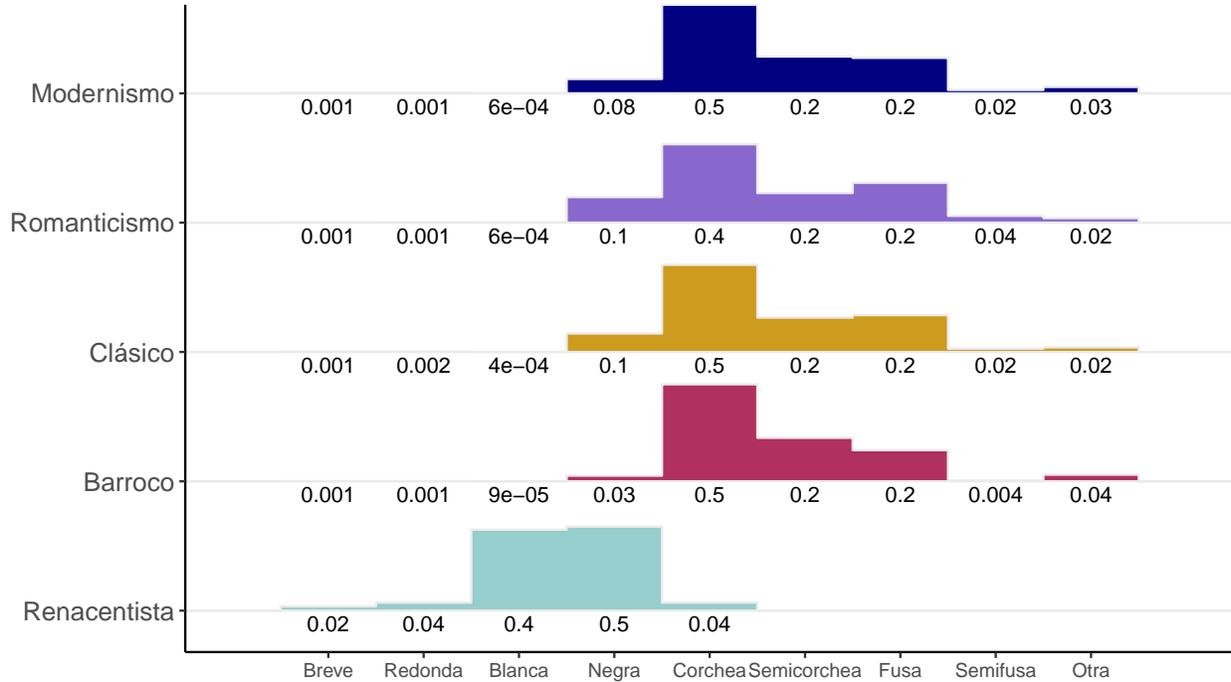
donde  $\underline{X}$  corresponde al conjunto de observaciones (obras musicales),  $P_j$  a la matriz de transición asociada a la  $j$ -ésima corriente histórica,  $I_j$  al conjunto de índices de las obras musicales pertenecientes a tal corriente en la muestra y  $j \in \{\text{Renacentista}, \text{Barroco}, \text{Clásico}, \text{Romanticismo}, \text{Modernismo}\}$ .

Los valores obtenidos para cada Cadena de Markov propuesta se muestran en la Figura 5.14 anterior y a continuación se resaltan algunas observaciones a rescatar:

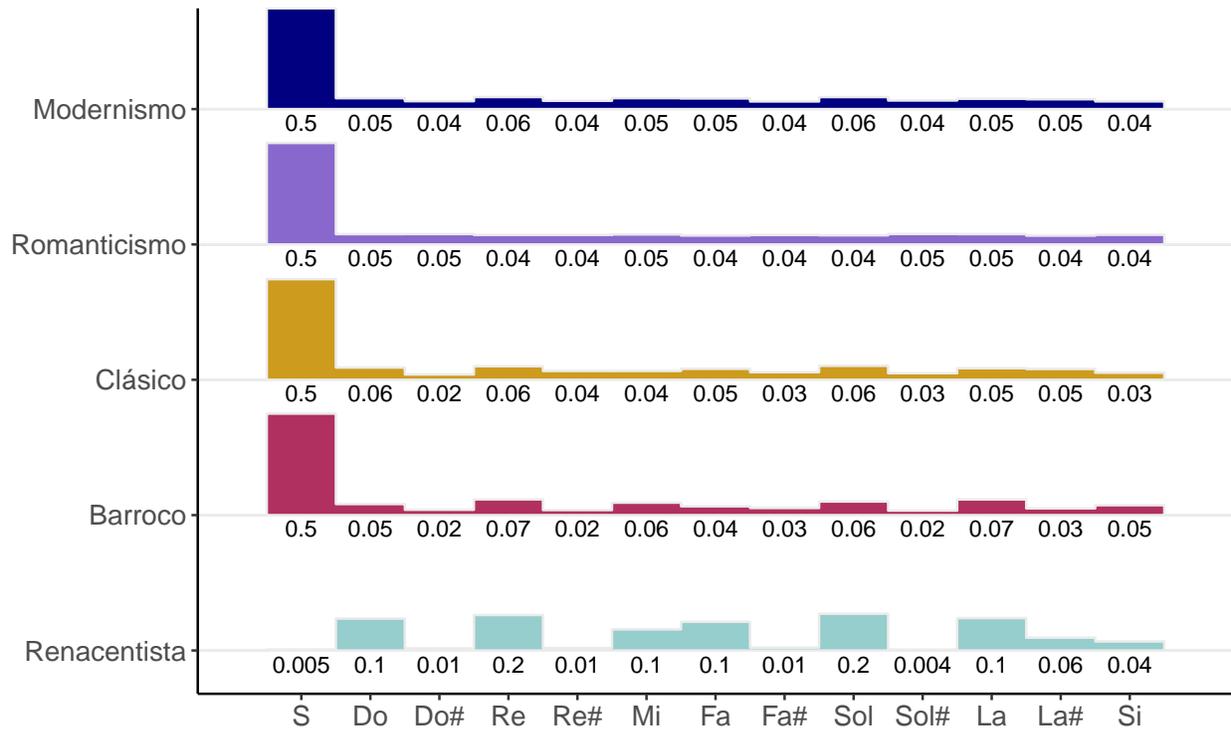
- En el Renacimiento se utilizaron sobre todo duraciones correspondientes a notas blancas y negras, y en absoluto, duraciones menores a la corchea. La duración a largo plazo predominante en las intervenciones de las obras musicales del resto de las corrientes es la corchea, a la que corresponden probabilidades de ocurrencia de entre 0.4 y 0.5. También se evidencia la caída en desuso de notas breves, valuando su probabilidad de ocurrencia en valores menores a 0.001.

---

<sup>21</sup>Véase Figura 5.9.

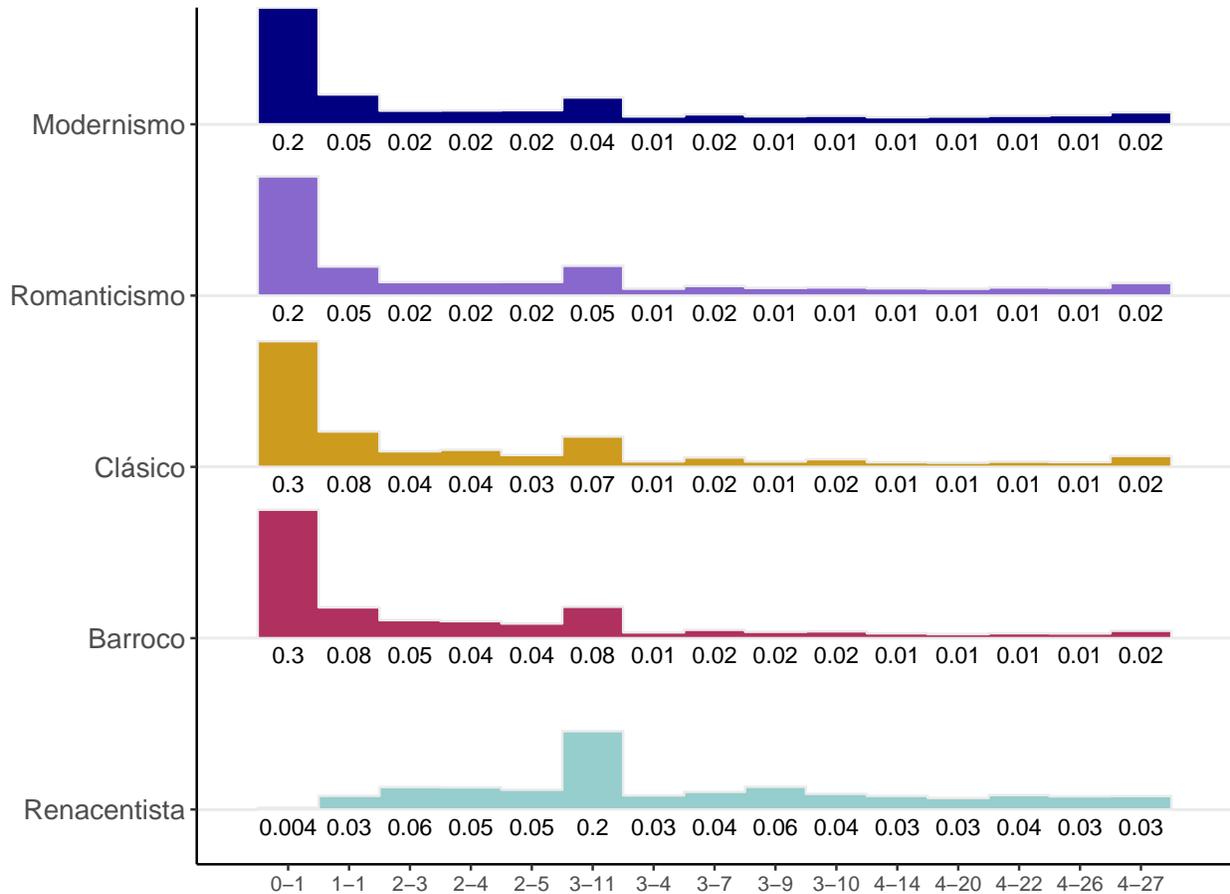


(a) Duraciones:  $\{D_n\}_{n \in \mathbb{N}}$ .



(b) Notas fundamentales:  $\{R_n\}_{n \in \mathbb{N}}$ .

Figura 5.14: Estimaciones de las probabilidades límite asociadas a las cadenas  $\{D_n\}_{n \in \mathbb{N}}$ ,  $\{R_n\}_{n \in \mathbb{N}}$  y  $\{C_n\}_{n \in \mathbb{N}}$  por corriente histórica.



(c) Clase Forte (entradas del vector de probabilidades con valores mínimos de 0.1):  $\{C_n\}_{n \in \mathbb{N}}$ .

Figura 5.14: Estimaciones de las probabilidades límite asociadas a las cadenas  $\{D_n\}_{n \in \mathbb{N}}$ ,  $\{R_n\}_{n \in \mathbb{N}}$  y  $\{C_n\}_{n \in \mathbb{N}}$  por corriente histórica.

- Por otro lado, a partir del Barroco se hizo mucho más frecuente el uso de los silencios, con prácticamente el 0.5 de probabilidad de ocurrencia en las obras a largo plazo; en contraste con las obras Renacentistas, en que la probabilidad del uso de los silencios es apenas 0.005.
- En el mismo periodo fue sumamente improbable la utilización de notas no naturales como raíces de los acordes, ya que sus probabilidades asociadas resultaron menores a 0.01; factor que fue diversificándose al paso del tiempo puesto que las probabilidades asociadas a notas no naturales son mayores en los periodos siguientes.
- De manera posterior al Renacimiento, aumenta el uso de notas individuales al menos al doble, no sólo acordes (típico de los cantos gregorianos) y a la reducción del uso de sonidos simultáneos de dos notas.

- Asimismo, se evidencia la diversificación en el uso de triadas de cualidades más allá de menor y mayor (3-11), naturalmente ante la sofisticación derivada del establecimiento de la Teoría de la Armonía en el Barroco.

De manera análoga a la estimación por corriente de las probabilidades límite de las cadenas planteadas, fueron obtenidas estimaciones de los factores de consonancia y disonancia  $\theta$  por periodo histórico, las cuales son presentadas en el siguiente cuadro. Como es de esperarse, el periodo con mayor noción de consonancia en la Música es el Renacimiento, el cual es seguido por el Clasicismo, periodo en que se exaltaba la belleza. Por otro lado, la corriente histórica en que se muestra una tendencia mayor a la disonancia es el Modernismo, lo cual es congruente con el uso de acordes constituidos por un número mayor de notas. Asimismo, el periodo en que se estima una noción menor de disonancia es el Barroco, mismo cuya noción de consonancia es menor en comparación al resto: esto puede implicar que en tal periodo las oscilaciones en el cambio de consonancia y disonancia en las composiciones eran *menos abruptas* que en el resto.

Corriente histórica	$\hat{\theta}_c$	$\hat{\theta}_d$
Renacimiento	66.59	8.47
Barroco	12.7	1.99
Clasicismo	24.19	11.28
Romanticismo	22.38	5.25
Modernismo	23.1	12.39

Cuadro 5.4: Estimaciones de los factores de consonancia y disonancia por corriente histórica.

De la Figura 5.15 cabe destacar que existe una gran separación entre las obras musicales renacentistas y el resto. La separación entre los otros grupos no es tan evidente; si acaso pueden separarse otros dos grandes grupos, los conformados por las obras del Clásico (mostaza) y del Barroco (magenta), y las del Romanticismo (violeta) y el Modernismo (azul). Por último, cabe hacer notar que el último Componente Principal, separado por corriente, sugiere que la base de datos cuenta con 28 datos atípicos<sup>22</sup>: seis obras pertenecientes al Renacimiento, once al Barroco, cinco al Clásico, tres al Romanticismo y tres al Modernismo.

<sup>22</sup>Véase Anexo III. Observaciones atípicas.

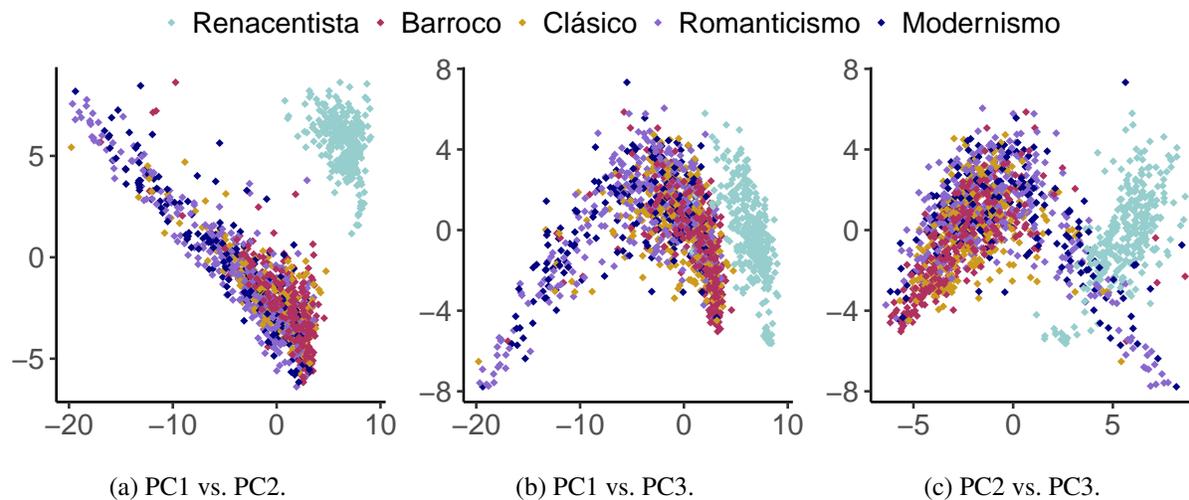
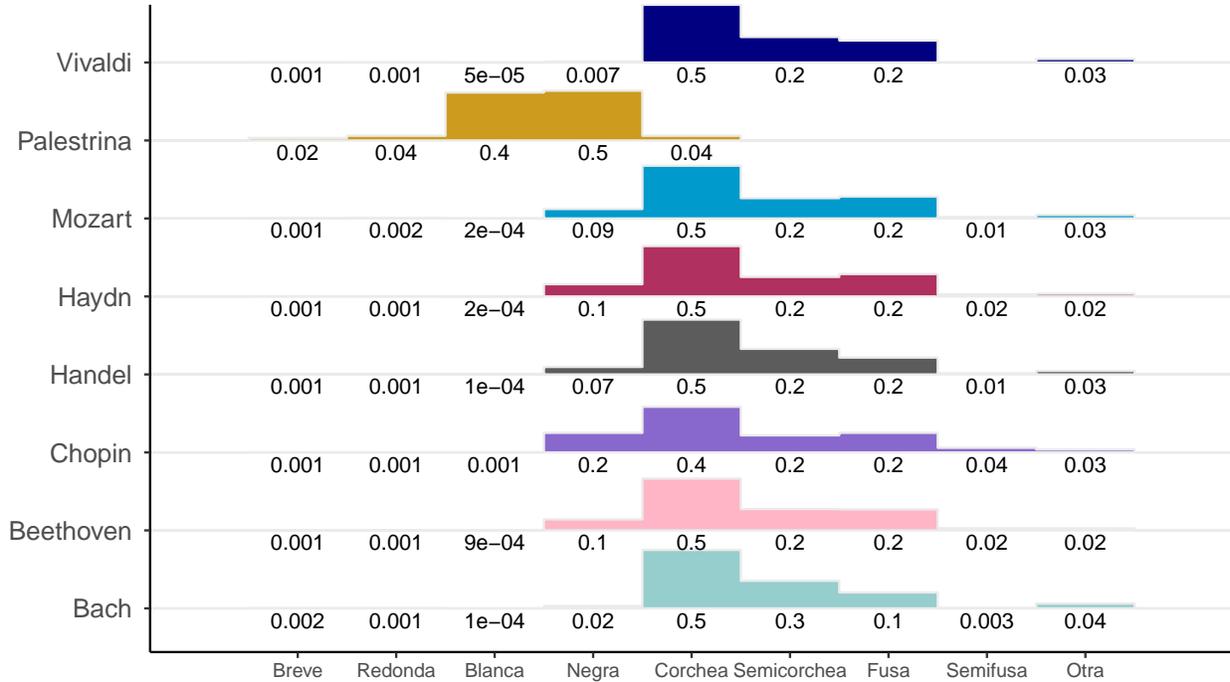


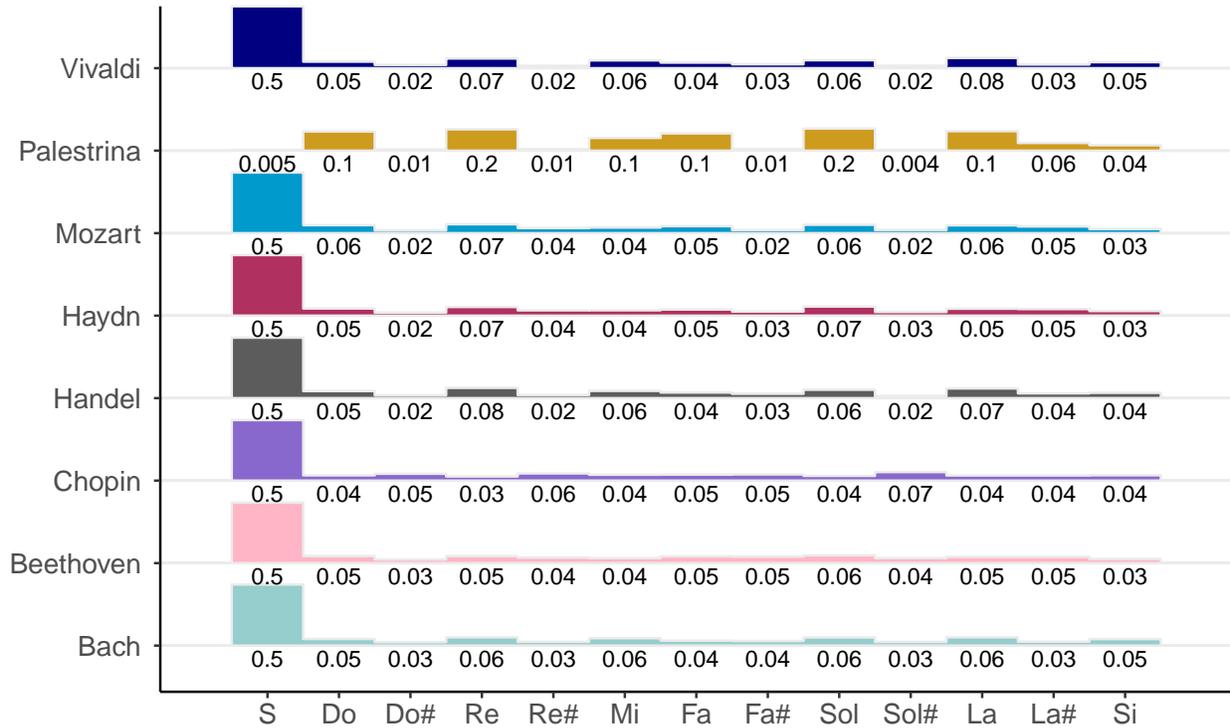
Figura 5.15: Proyección de las observaciones sobre los tres primeros Componentes Principales dos a dos por corriente histórica a la que pertenecen.

### 5.3.2. Análisis de autores

Las estimaciones de las variables en cuestión fueron hechas de manera análoga a la descrita en la sección anterior, respecto los autores Bach, Beethoven, Chopin, Handel, Haydn, Mozart, Palestrina y Vivaldi. Las probabilidades límite estimadas obtenidas para cada uno se muestran en la Figura 5.16; las cuales sugieren influencia de las tendencias de las épocas que vivieron tales personajes en sus obras musicales, debido a que las cifras obtenidas son sumamente *parecidas* entre sí para autores que convivieron en el mismo periodo, por ejemplo, Beethoven y Mozart. Cabe destacar que los autores analizados fueron los principales exponentes de la corriente a la que pertenecieron, por lo que tiene sentido que se hayan obtenido resultados parecidos al análisis por corrientes históricas, pues fundamentalmente ellos marcaron las tendencias musicales de su época.

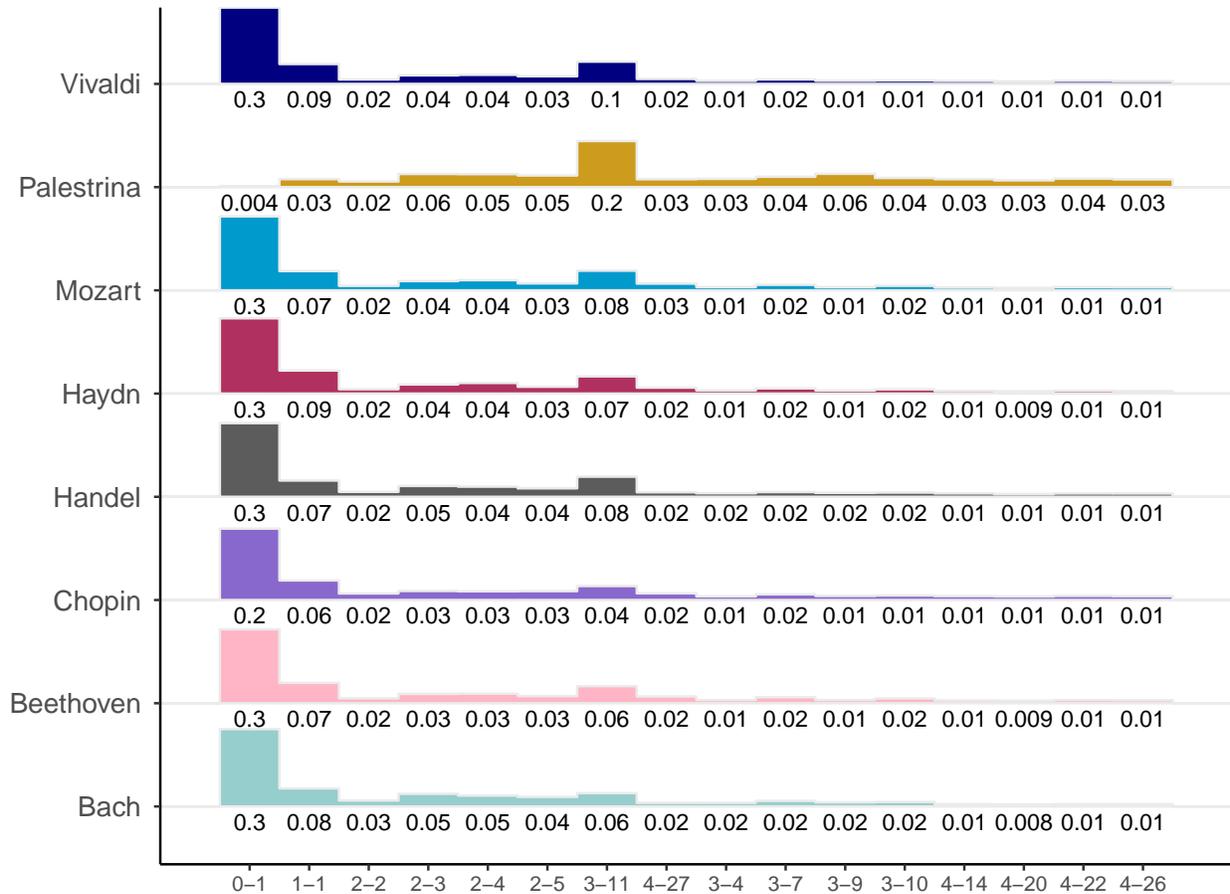


(a) Duraciones:  $\{D_n\}_{n \in \mathbb{N}}$ .



(b) Notas fundamentales:  $\{R_n\}_{n \in \mathbb{N}}$ .

Figura 5.16: Estimaciones de las probabilidades límite asociadas a las cadenas  $\{D_n\}_{n \in \mathbb{N}}$ ,  $\{R_n\}_{n \in \mathbb{N}}$  y  $\{C_n\}_{n \in \mathbb{N}}$  por autor.



(c) Clase Forte (entradas del vector de probabilidades con valores mínimos de 0.1):  $\{C_n\}_{n \in \mathbb{N}}$ .

Figura 5.16: Estimaciones de las probabilidades límite asociadas a las cadenas  $\{D_n\}_{n \in \mathbb{N}}$ ,  $\{R_n\}_{n \in \mathbb{N}}$  y  $\{C_n\}_{n \in \mathbb{N}}$  por autor.

Ahora bien, pese a lo anterior, desde luego que pertenecer a la misma corriente musical no implica que las piezas de los autores mencionados sean parecidas. Es claro que cada uno poseía estilos diferentes, lo cual es evidenciado, no a partir de la modelación por Cadenas de Markov, sino a partir de los factores de consonancia y disonancia reflejados en sus obras musicales y modelados con Campos Aleatorios:

- Palestrina resultó ser el autor con mayor noción de la consonancia.
- Vivaldi es aquel con mayores factores de consonancia y disonancia respecto su época (en comparación a Bach y Handel), i.e., las fluctuaciones en sus obras entre consonancia y disonancia son más marcadas que el resto. En cambio, Handel resulta estar asociado a menores factores de consonancia y disonancia (con un valor de apenas 3.15); y Bach se sitúa en un

punto medio entre los dos anteriores, siendo asociado a él un factor de disonancia de aproximadamente el doble que Handel y la mitad que Vivaldi.

- En comparación a Mozart y Haydn, Beethoven utilizó fluctuaciones más estables entre consonancia y disonancia en medida de que es quien reporta menores valores de tales factores. Cabe destacar que Mozart resultó estar asociado a factores de consonancia y disonancia de 28.8 y 30.49, siendo el único de los autores analizados cuya incurrencia en la disonancia es mayor que la de la consonancia, aunque son magnitudes afines.
- Finalmente, las magnitudes de los factores de consonancia y disonancia correspondientes a Chopin y Handel son sumamente cercanos, lo cual es sorprendente dado que se encuentran muy lejanos en la historia<sup>23</sup>.

Autor	$\hat{\theta}_c$	$\hat{\theta}_d$
Bach	35.18	8.9
Beethoven	15.82	3.15
Chopin	25.69	6.26
Handel	25.48	4.46
Haydn	19.71	10.38
Mozart	28.8	30.49
Palestrina	66.6	8.47
Vivaldi	42.78	20.27

Cuadro 5.5: Estimaciones de los factores de consonancia y disonancia por autor.

Nótese que la separación de las proyecciones sobre los primeros tres Componentes Principales de las observaciones por autor muestran que el grupo más distintivo son las obras musicales compuestas por Palestrina; aunque el resto de los grupos no muestran una separación muy marcada. Se presentan únicamente las proyecciones de los primeros tres Componentes Principales, mediante los cuales se explica apenas el 16.51 % de la varianza, lo cual muestra que la clasificación por autor no será labor sencilla.

<sup>23</sup>Chopin pertenece al Romanticismo, mientras que Handel al Barroco.

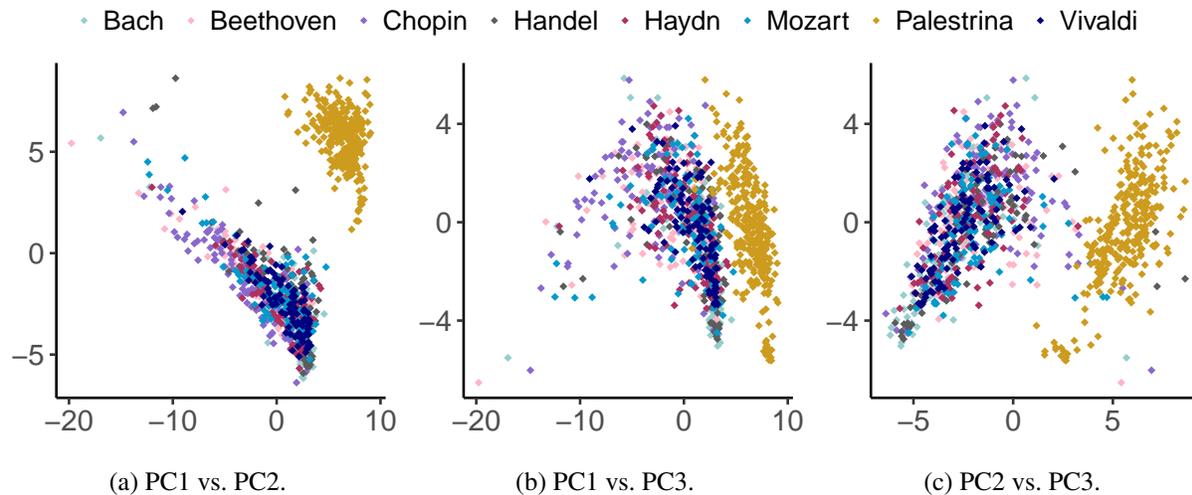


Figura 5.17: Proyección de las observaciones sobre los tres primeros Componentes Principales dos a dos por su autor.

Finalmente, bajo la separación por autor se sugiere la presencia de 22 datos atípicos en la muestra<sup>24</sup>: tres piezas de Bach, tres de Handel, dos de Haydn, una de Khachaturian, una de Lizst, seis de Palestrina, dos de Saint-Saens, dos de Schubert y dos de Vivaldi.

<sup>24</sup>Véase *Anexo III. Observaciones atípicas*.



## CAPÍTULO 6

# CLASIFICACIÓN DE OBRAS MUSICALES

En el presente capítulo se lleva a cabo la clasificación de obras musicales por su autor y la corriente histórica a la que pertenecen mediante los métodos de Aprendizaje Estadístico introducidos en el Capítulo 4 implementados en R con las paqueterías `h2o`<sup>1</sup>, `ordinal`<sup>2</sup>, `rpart`<sup>3</sup>, `class`<sup>4</sup> y `MASS`<sup>5</sup>. Para tal finalidad se utilizará como muestra las estimaciones obtenidas con los métodos presentados en el capítulo anterior para las obras musicales listadas en el *Anexo V* –disponible en [github.com/danielapdv/music-mcs-](https://github.com/danielapdv/music-mcs-), con lo cual se espera que la modelación propuesta brinde información representativa de la Música.

Específicamente, las variables predictoras que componen la base de datos son:

- Las probabilidades límite para las Cadenas de Markov asociadas a la duración, nota fundamental y Clase Forte,
- Los factores de disonancia y consonancia estimados, y
- La evaluación de la función de log-Verosimilitud de la obra en términos de su medida de Gibbs. Cabe destacar que inicialmente, en lugar de ésta, se incluyó en la muestra la evaluación de la medida de Gibbs relativa a cada obra para cada configuración posible del Campo

---

<sup>1</sup>Desarrollada por Erin LeDell, Navdeep Gill, Spencer Aiello, Anqi Fu, Arno Candel, Cliff Click, Tom Kraljevic, Tomar Nykodym, Patrick Aboyoun, Michal Kurka y Michal Malohlava.

<sup>2</sup>Desarrollada por Rune Haubo Bojesen Christensen.

<sup>3</sup>Desarrollada por Terry Therneau, Beth Atkinson y Brian Ripley.

<sup>4</sup>Desarrollada por Brian Ripley y William Venables.

<sup>5</sup>Desarrollada por Brian Ripley, Bill Venables, Douglas M. Bates, Kurt Hornik, Albrecht Gebhardt y David Firth.

Aleatorio. No obstante, se presentarán resultados para la alternativa mencionada debido a que de esta manera se alcanzan tasas de clasificación errónea menores.

Es momento de hacer notar que en los diversos métodos a aplicar se considerará la base de datos antes mencionada, los Componentes Principales presentados en el capítulo anterior y los Componentes Principales obtenidos al realizar el PCA sin reescalar las variables; lo último puesto que en algunos casos se tuvo mejores resultados (tasas de error de clasificación validadas menores) de esta manera<sup>6</sup>. Se incluirán para la clasificación los Componentes Principales por múltiples razones: para obtener una mejor representación de la información, para reducir la dimensión del espacio de predictores (se cuenta con 302 variables explicativas) y como medida de solución a la multicolinealidad que presentan entre sí, dada su naturaleza<sup>7</sup>.

De antemano se establece que a lo largo del presente capítulo se presentarán estadísticas validadas mediante *5-Fold Cross-Validation*, i.e., para ello se considerarán cinco particiones de la base de datos; las cuales serán exactamente iguales para cada método aplicado con el fin de poder compararlos y concluir. De tal manera, se ajustarán los modelos con el 80% de las observaciones y se evaluarán con el 20% restante en cada iteración de la validación de los modelos.

## 6.1. Clasificación por corriente musical

En esta sección se considerarán las tres bases de datos enunciadas, las cuales cuentan con 1,456 observaciones y 302 predictores cada una. La variable respuesta  $Y$  toma valores en el conjunto de clases  $\{\textit{Renacentista}, \textit{Barroco}, \textit{Clásico}, \textit{Romanticismo}, \textit{Modernismo}\}$  y está distribuida de manera relativamente uniforme sobre la muestra, como se aprecia en el siguiente cuadro. Nótese que la verdadera cualidad de  $Y$  en este contexto es ordinal, por lo que se incluirá clasificación por Regresión Ordinal; aunque la aplicación del resto de los métodos de clasificación implica que se le dé tratamiento correspondiente a variable categórica nominal.

---

<sup>6</sup>El respectivo análisis de encuentra en el *Anexo IV*.

<sup>7</sup>Evidenciada por el Análisis de Conglomerados Jerárquico para las variables y la Figura A.1 del *Anexo IV*.

Corriente (Y)	Obras musicales	Porcentaje
Renacentista	300	20.6
Barroco	315	21.63
Clásico	300	20.6
Romanticismo	339	23.28
Modernismo	202	13.87
Total	1,456	100

Cuadro 6.1: Piezas musicales por clase en las bases de datos para la clasificación por corriente histórica.

### 6.1.1. Regresión Multinomial

En primera instancia, se ajustó un modelo de Regresión Multinomial incluyendo todos los predictores de la base de datos *original*; exceptuando aquéllas que representan las entradas de los vectores de las probabilidades límite *Otra*, *Si* y *12-1* estimadas para las Cadenas de Markov asociadas a las duraciones, notas fundamentales y Clases Forte de la obra musical, respectivamente. Esta medida fue tomada debido a que incluirlas implica multicolinealidad entre los predictores, lo cual viola los supuestos de los GLM. De hecho, es muy probable que este modelo muestre poca bondad de ajuste justo por la misma razón<sup>8</sup>; de manera que, en realidad, el ajuste de este modelo es con fines exclusivamente comparativos respecto al resto. La tasa de error de clasificación global validada resultante del ajuste de tal modelo corresponde a 47.18 %, otras estadísticas respecto al mismo son consultables en el Cuadro 6.2.

Por otro lado, se ajustaron modelos de Regresión Multinomial sobre las observaciones proyectadas a los Componentes Principales obtenidos del Análisis, con las variables reescaladas y sin reescalar, así como variando el número de Componentes Principales incluidos para la modelación. En la siguiente figura se muestra la variación de las tasas de clasificación errónea globales y por grupo validadas en función de la cantidad de PCs contemplados en cada modelo, junto con las tasas globales aparentes y la validada para el modelo con los datos originales.

<sup>8</sup>Véase la matriz de correlaciones de las variables en la Figura A.1 del *Anexo IV*.

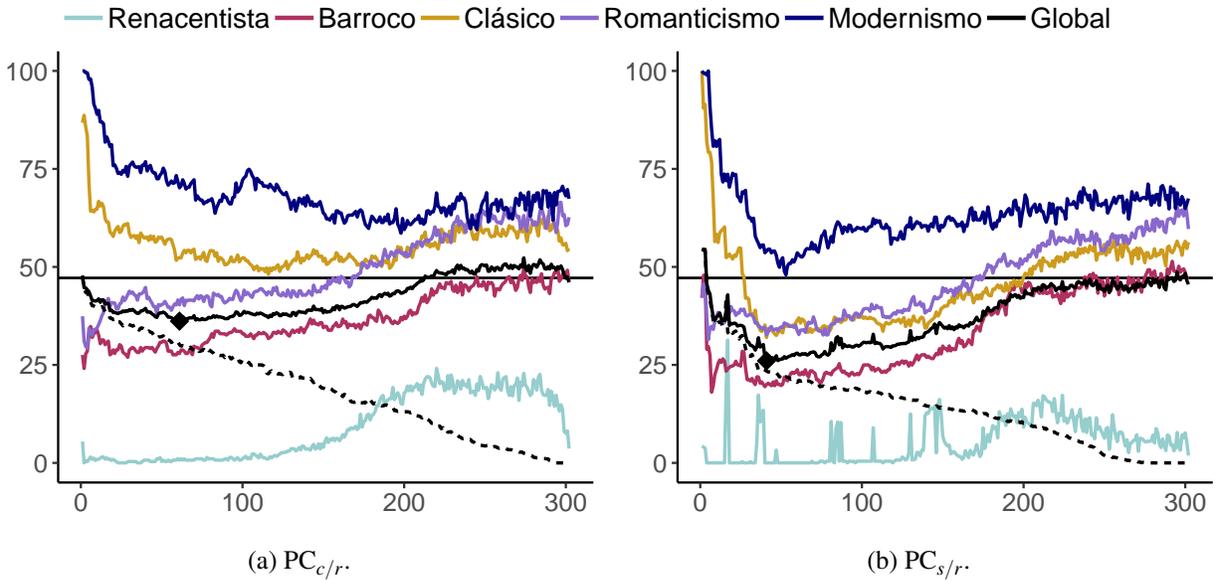


Figura 6.1: Tasas de error de clasificación (%) por corriente musical, globales y por grupo, validadas y aparente global estimadas mediante Regresión Multinomial respecto al número de Componentes Principales considerados para el ajuste del modelo, obtenidos al reescalar las variables ( $PC_{c/r}$ ) y sin reescalarlas ( $PC_{s/r}$ ). El punto indica la tasa de clasificación errónea global validada mínima; la recta constante, la tasa de error global del modelo de Regresión Multinomial con base en las observaciones originales; y la línea punteada, la tasa de error de clasificación global aparente.

A pesar de que la proporción de varianza explicada acumulada por los PCs en el caso en que se reescalaron las variables originales alcanza el valor de 95% hasta considerar los primeros 239, a partir de la inclusión de más de 120 Componentes Principales aproximadamente, el modelo comienza a saturarse (de igual manera para el caso en que los PCs fueron obtenidos sin reescalamiento de las variables). Esto es evidenciado por el hecho de que a partir de ese punto las tasas de error de clasificación validadas muestran un comportamiento a la alza, mientras que la global aparente sigue a la baja. Es decir, a medida que se aumenta la cantidad de PCs en los modelos, se tiene mayor bondad de ajuste a los datos del conjunto de entrenamiento, pero se reduce sustancialmente su poder predictivo *real*. Este hecho, en realidad, era de esperarse debido a la composición de la base de datos: con 302 predictores y aproximadamente 1,165 observaciones en el conjunto de entrenamiento, se tiene una relación de menos de cuatro observaciones por variable. De hecho, es una problemática común al ajustar este tipo de modelo en altas dimensiones. La solución, desde luego, radica en validar las estadísticas asociadas a su poder predictivo para tener una noción clara del sobreajuste y, por lo general, resulta poco conveniente incluir una cantidad *grande* de predictores.

Por otro lado, la tasas de clasificación errónea globales validadas se minimizan con los pri-

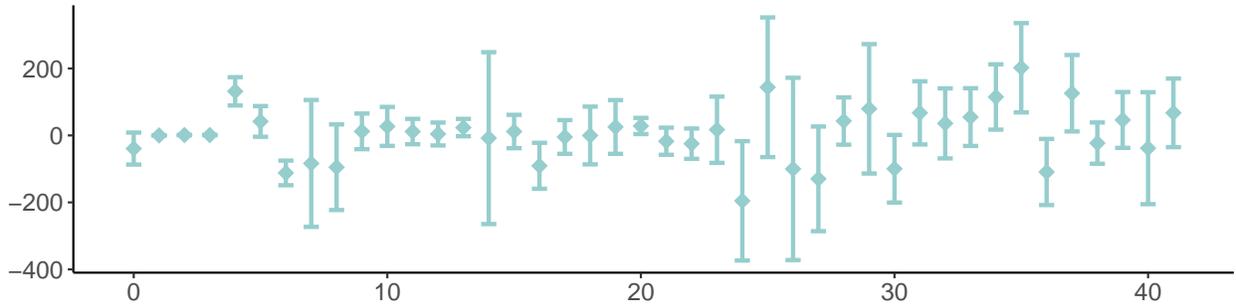
meros 61 y 41 Componentes Principales obtenidos al reescalar y sin reescalar las variables, respectivamente, resultando esos casos en tasas de 36.16 % y 26.04 %. Con esta última se logra una reducción del error a casi la mitad respecto a la tasa obtenida con la base de datos original.

	Modelo	Devianza	AIC	BIC	TCEG (%)
Aparentes	Variables originales	4,645.82	5,245.82	6,830.86	0
	PC <sub>c/r</sub> <sup>61</sup>	2,530.75	2,654.75	2,982.33	30.15
	PC <sub>s/r</sub> <sup>41</sup>	2,849.76	2,933.76	3,155.66	23.42
Validadas	Variables originales	3,716.41	4,316.41	5,901.45	47.18
	PC <sub>c/r</sub> <sup>61</sup>	<b>2,074.43</b>	<b>2,198.43</b>	<b>2,526</b>	36.16
	PC <sub>s/r</sub> <sup>41</sup>	2,309.41	2,393.41	2,615.31	<b>26.04</b>

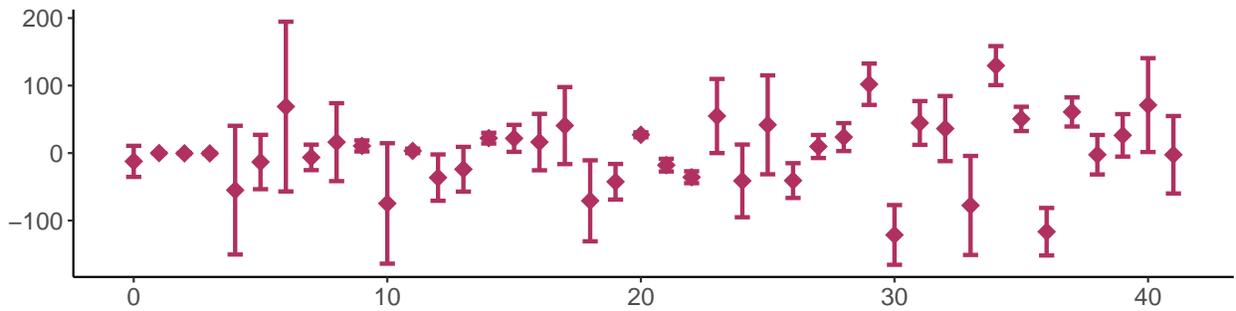
Cuadro 6.2: Devianza, Criterio de Información de Akaike, Criterio de Información Bayesiano y tasa de clasificación errónea global, aparentes y validadas, asociados a los modelos de Regresión Multinomial ajustados para clasificación de obras musicales por corriente histórica con las variables originales, 61 PCs obtenidos con reescalamiento en las variables (PC<sub>c/r</sub><sup>61</sup>) y 41 PCs obtenidos sin reescalarlas (PC<sub>s/r</sub><sup>41</sup>).

En el Cuadro 6.2 se evidencia la saturación del modelo ajustado con base en las variables originales: la tasa de clasificación errónea global es nula, pero las tasas de error de clasificación se disparan a la alza al validarlas, lo que implica que tal modelo carece de poder predictivo, pues se estima un error de clasificación casi del 50 %.

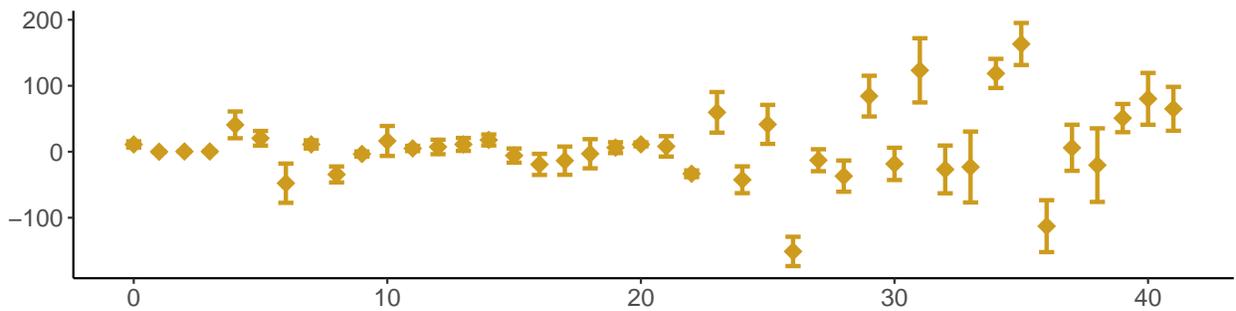
Ambos modelos construidos con base en los Componentes Principales son mucho más estables, en el sentido de que al validar sus estadísticas, no empeoran de manera sustancial; además, no se tiene indicios de sobreajuste. Aquel ajustado con los primeros 61 Componentes Principales obtenidos al reescalar las variables muestra mayor bondad de ajuste a los datos en cuanto a criterios de información, siendo entre 234.98 y 1,641.98 unidades menores los asociados a tal modelo en todos los casos, respecto los del modelo restante. No obstante, el modelo relativo a los primeros 41 PCs obtenidos sin reescalamiento de las variables reporta una tasa de clasificación errónea global validada menor que la del anterior por 10.12 %, por lo que se concluye que es el mejor modelo de la presente sección y se analizará con más detalle a continuación.



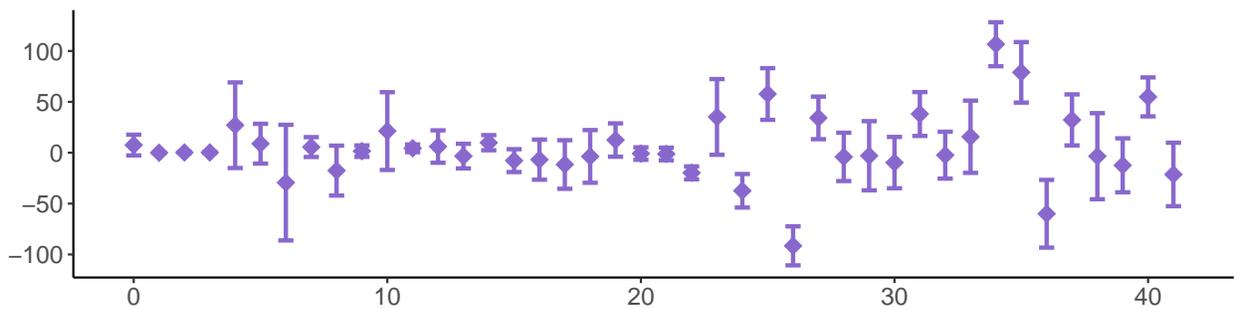
(a) Renacentista.



(b) Barroco.



(c) Clásico.



(d) Romanticismo.

Figura 6.2: Coeficientes de regresión estimados y sus intervalos al 0.95 de confianza por *Cross-Validation* del modelo de Regresión Multinomial para la clasificación por corriente musical ajustado con los primeros 41 PCs obtenidos sin reescalar las variables originales. El eje x indica el número del Componente Principal, el cero corresponde al intercepto.

La devianza *aparente* del modelo en cuestión se valuó en 2,849.76, mientras que la devianza del modelo *nulo* –ajustado únicamente con interceptos– en 4,646.17. De aquí, con el fin incluir un breve Análisis de Devianza, puede hacerse la prueba  $\chi^2$  al restar las devianzas mencionadas  $D_{nulo} - D_{PC_{s/r}^{41}} = 1,796.42$ , cuya hipótesis nula corresponde a que los efectos de las variables incluidas en el modelo no son significativos, i.e., el modelo nulo se ajusta mejor a los datos que el modelo con predictores. El *p-value* correspondiente resulta ser menor a  $2e-16$ , por lo que la prueba se rechaza, es decir, no se cuenta con evidencia suficiente para asumir que los efectos de las variables incluidas en el modelo con 41 Componentes Principales no son significativos (hay al menos uno que lo es)<sup>9</sup>.

Los coeficientes del modelo estimados por *Cross-Validation*, así como su desviación estándar, intervalos de confianza, sesgo, estadístico y *p-value* correspondientes a la prueba de Wald para evaluar su significancia se encuentran en los cuadros A.11 al A.14 del *Anexo VI* –disponible en [github.com/danielapdv/music-mcs](https://github.com/danielapdv/music-mcs)– y su representación gráfica se muestra en la Figura 6.2. Cabe destacar que mediante la prueba de Wald, resultan no ser significativos once coeficientes de regresión asociados a la clase *Renacentista*, cinco al *Barroco*, dos al *Clásico* y nueve al *Romanticismo*. Sin embargo, dado que con este modelo se logra la menor TEGG (tasa de error de clasificación global) validada del apartado y posteriormente se hará Selección de Modelos, no se modificará.

En las gráficas siguientes se muestran las probabilidades estimadas  $\hat{\mathbb{P}}(Y = k | \underline{X})$  validadas para cada clase  $k \in \{\textit{Renacentista}, \textit{Barroco}, \textit{Clásico}, \textit{Romanticismo}, \textit{Modernismo}\}$  proyectadas sobre el primer Componente Principal y sobre las variables (originales) asociadas a los factores de consonancia y disonancia y la evaluación de la función de log-Verosimilitud en términos de las medidas de Gibbs para cada obra musical. Es decir, con base en los coeficientes de regresión validados se hicieron predicciones anulando todos los predictores excepto por el primer Componente Principal. Para las variables mencionadas restantes, se realizó el procedimiento análogo, y se llevó a cabo la proyección mediante la matriz de rotación obtenida por el PCA.

En primera instancia, cabe destacar que con los Componentes Principales obtenidos sin reescalar las variables originales, el 99.996% de la proporción de varianza se acumula con el primer PC, por lo que es *interesante* observar la variación de las probabilidades mencionadas a través de la variación del mismo, como se muestra en la Figura 6.2a. Obsérvese que, respecto tal predictor, los intervalos en que lo más probable es pertenecer a las clases *Renacentista* y *Barroco* están bien

---

<sup>9</sup>Se obtienen los mismos resultados al realizar la prueba sobre los valores validados por *Cross-Validation*.

delimitados, lo cual refleja el hecho de que éstas son las dos clases con menores tasas de error de clasificación. Ahora bien, la probabilidad relativa a pertenecer al Modernismo está acotada por la del resto de las clases, i.e., el primer PC no sirve de mucho para asignar tal clase certeramente.

Luego, se muestra la variación de las probabilidades estimadas por clase respecto la evaluación de la función de log-Verosimilitud, y los factores de consonancia y disonancia debido a que son aquellas variables que mayor peso tienen en el primer PC, en cuanto a la magnitud de las ponderaciones respectivas (0.999996, 0.003 y  $-0.0006$  –véase *Anexo IV*–). En la Figura 6.2c puede apreciarse que el factor de consonancia aumenta la posibilidad de que el clasificador asigne la categoría Modernismo. En cuanto a la evaluación de su función de log-Verosimilitud y el factor de disonancia, el problema de acotamiento de la probabilidad de asignar tal clase prevalece y, de hecho, se intensifica. Sin embargo, nótese que claramente la log-Verosimilitud no puede tomar valores mayores o iguales a cero, ni los factores menores o iguales a cero.

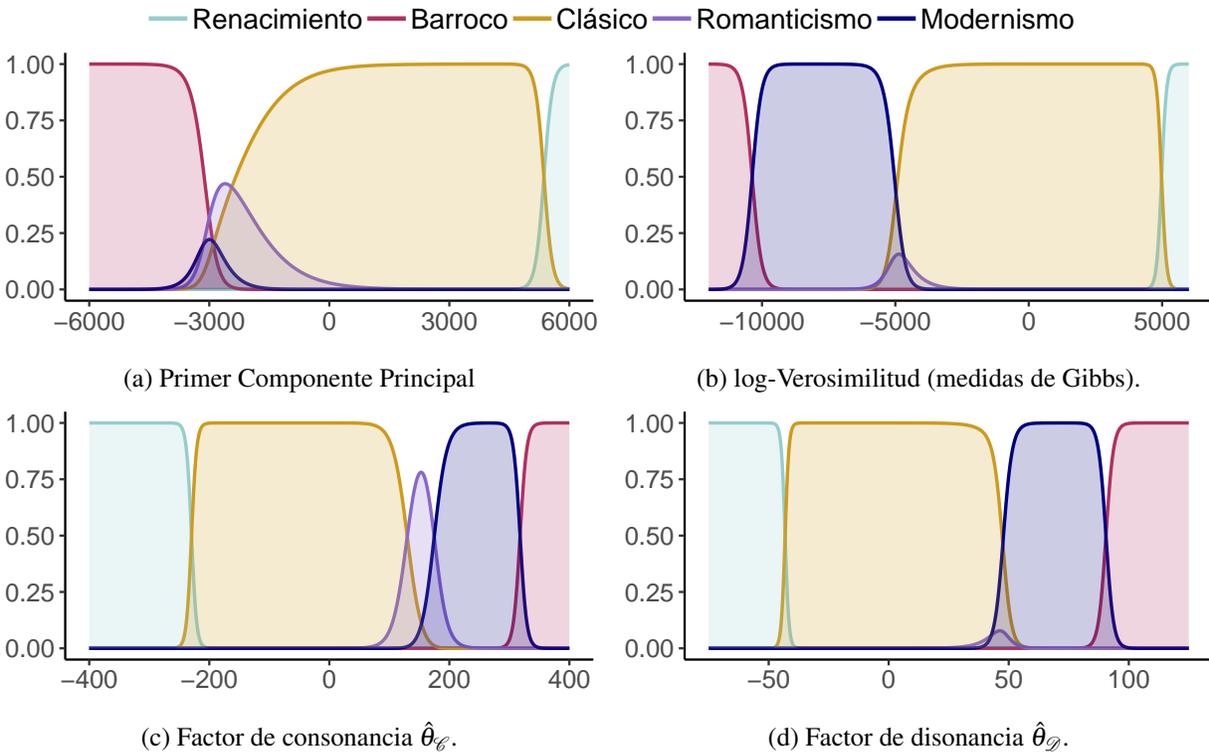


Figura 6.3: Proyección de las probabilidades estimadas por corriente musical  $\hat{\mathbb{P}}(Y | \underline{X})$  obtenidas mediante Regresión Multinomial validadas sobre el primer Componente Principal, la evaluación de la función de log-Verosimilitud maximizada en términos de las medidas de Gibbs y los factores de consonancia y disonancia.

## 6.1.2. Regresión Multinomial Regularizada

En el apartado anterior se evidencia la necesidad de ajustar modelos de regresión con regularización debido a la tendencia a la saturación que muestran. Por lo que, aunado a la correlación de los predictores que conforman la base de datos original, ésta sección está destinada al ajuste de modelos de Regresión Multinomial regularizada por los métodos Ridge, Lasso y *Elastic Net*.

En la Figura 6.4 se muestran las tasas de error validadas variando el número de Componentes Principales incluidos para el ajuste de los modelos utilizando como penalización la norma  $\ell_1$  (Lasso),  $\ell_2$  (Ridge) y una ponderación de ambas con pesos 0.5 (*Elastic Net*<sup>10</sup> con  $\alpha = 0.5$ ) y su comparación con las respectivas tasas de error de clasificación globales aparentes y la tasa de error global correspondiente a cada tipo de modelo ajustado con las variables originales. Los puntos a partir de los cuales los modelos se saturan son muy marcados en cuanto a Regresión por Lasso y *Elastic Net*. Específicamente, para ambos conjuntos de PCs (correspondientes a las variables originales reescaladas y no reescaladas), el sobreajuste se alcanza con los modelos construidos con los primeros 91 PCs o más. En cuanto a la Regresión Ridge, no es tan evidente el punto en que comienza su saturación, pero nótese que también sucede alrededor de los 90 PCs.

En todos los casos expuestos, se logran tasas de error de clasificación validadas globales menores a las relativas a los modelos ajustados con base en las variables originales. Esto, con los primeros 71, 171 y 61 Componentes Principales basados en las variables reescaladas y con los primeros 44, 49 y 44 PCs obtenidos sin reescalar las variables, respectivamente a los modelos regularizados por Lasso, Ridge y *Elastic Net*. Las tasas de error específicas obtenidas se muestran en el Cuadro 6.3.

---

<sup>10</sup>Mediante este método se considera la función objetivo a optimizar como sigue:

$$\frac{1}{2 \cdot n} \cdot \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j \cdot x_{ij} \right)^2 + \lambda \cdot \left( \alpha \cdot \|\underline{\beta}\|_1 + (1 - \alpha) \cdot \|\underline{\beta}\|_2^2 \right); \text{ donde } \alpha \in [0, 1].$$

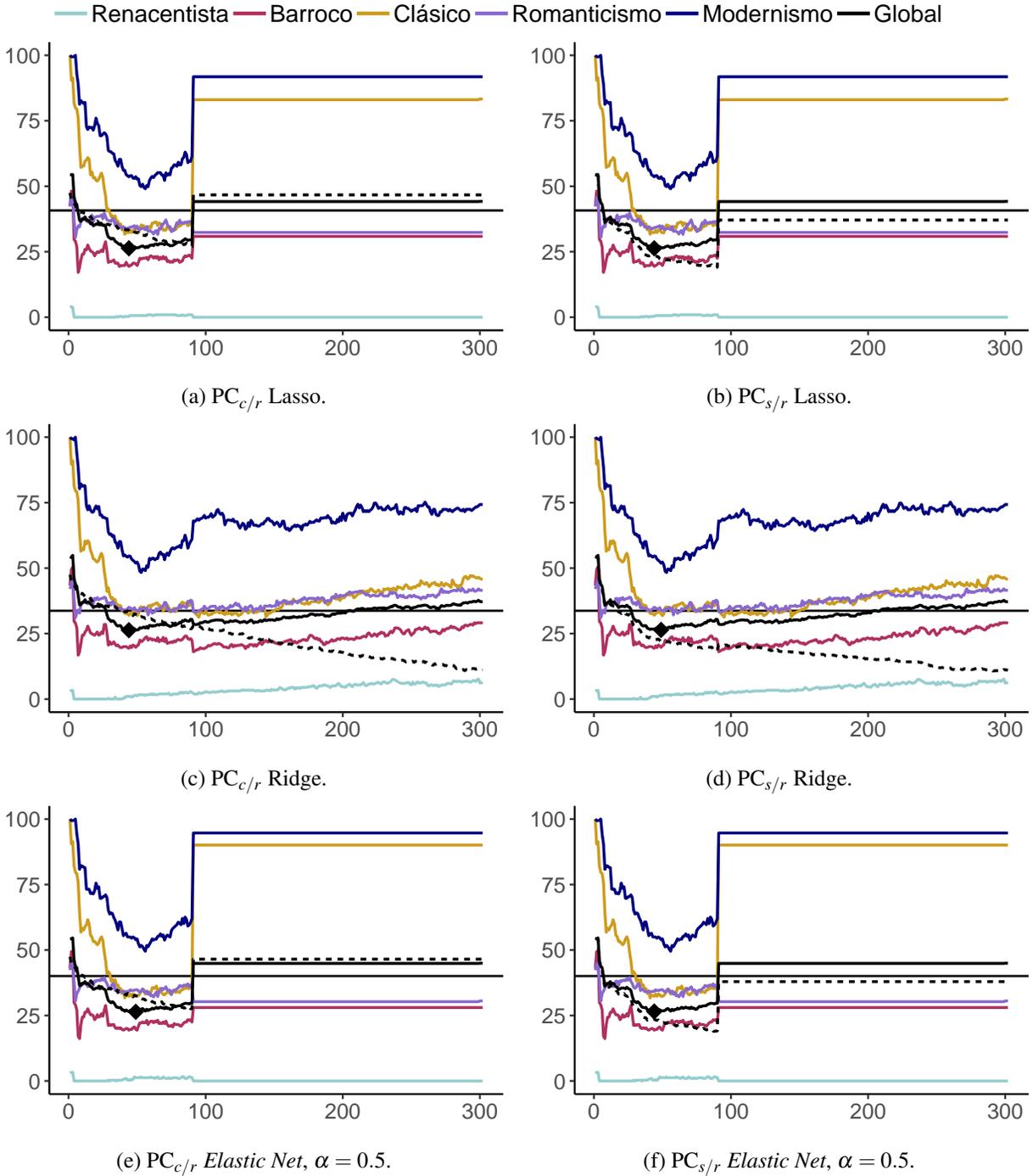


Figura 6.4: Tasas de error de clasificación (%) por corriente musical, globales y por grupo, validadas y aparente global estimadas mediante Regresión Multinomial Regularizada respecto al número de PCs considerados para el ajuste del modelo, obtenidos con ( $PC_{c/r}$ ) y sin reescalar las variables ( $PC_{s/r}$ ). El punto indica la tasa global validada mínima; la recta constante, la tasa global del modelo con las observaciones originales; y la línea punteada, el error global aparente.

Nótese que los datos con que se obtiene la tasa de error global validada menor es con los primeros 44 Componentes Principales obtenidos sin reescalar las variables originales, ajustando un modelo de Regresión Multinomial por Lasso. Tal tasa es menor que las obtenidas con los mismos datos pero por Ridge y *Elastic Net* por 0.1% y por 0.2%, respectivamente. Asimismo, es menor que las resultantes de aplicar el mismo método pero con los otros conjuntos de datos por más de 10%. Finalmente, cabe destacar que el valor óptimo de la penalización para el ajuste de ese modelo es de  $\lambda = 0.00027$ .

Modelo	Lasso			Ridge			<i>Elastic Net</i>		
	V. O.	PC <sub>c/r</sub> <sup>71</sup>	PC <sub>s/r</sub> <sup>44</sup>	V. O.	PC <sub>c/r</sub> <sup>171</sup>	PC <sub>s/r</sub> <sup>49</sup>	V. O.	PC <sub>c/r</sub> <sup>61</sup>	PC <sub>s/r</sub> <sup>44</sup>
Aparente	37.57	28.77	<b>23.35</b>	<b>13.26</b>	19.5	21.91	37.23	30.22	<b>23.56</b>
Validada	40.75	36.03	<b>26.31</b>	33.73	35.45	<b>26.41</b>	40.06	35.95	<b>26.51</b>

Cuadro 6.3: Tasas de clasificación errónea (%) globales, aparentes y validadas, asociadas a los modelos de Regresión Multinomial Regularizada (Lasso, Ridge y *Elastic Net*) ajustados para clasificación de obras musicales por corriente histórica con las variables originales (V. O.) y los primeros  $x$  PCs con que se minimiza el error validado obtenidos con reescalamiento (PC<sub>c/r</sub> <sup>$x$</sup> ) de las variables y sin reescalarlas (PC<sub>s/r</sub> <sup>$x$</sup> ).

### 6.1.3. Regresión Ordinal

A pesar de haber obtenido *buenos* resultados con la modelación mediante Regresión Multinomial, se llevará a cabo el ajuste de modelos de Regresión Ordinal con el objetivo de mejorarlos, ya que al ser  $Y$  una variable ordinal en realidad, podría estarse perdiendo información al tratarla como nominal. De manera análoga a las secciones anteriores, se ajustó un modelo con base en todas las variables originales (excepto los complementos de las probabilidades límite) y otros variando el número de Componentes Principales para cada conjunto de PCs (obtenidos con reescalamiento y sin reescalar las variables originales).

Como puede apreciarse en la Figura 6.5, ajustar este tipo de modelo resulta en mayor estabilidad en cuanto a su saturación, en medida de que, al variar el número de PCs en el modelo, no se eleva de sobremanera la tasa de clasificación errónea validada. Aunque cabe destacar que con el mejor modelo ajustado con Componentes Principales obtenidos al reescalar las variables apenas de mejora el error de clasificación del modelo con las variables originales. Adicionalmente, como se muestra en el siguiente cuadro, las tasas globales validadas de minimizan con los primeros 260 y 61 PCs obtenidos reescalando y sin reescalar las variables originales, respectivamente.

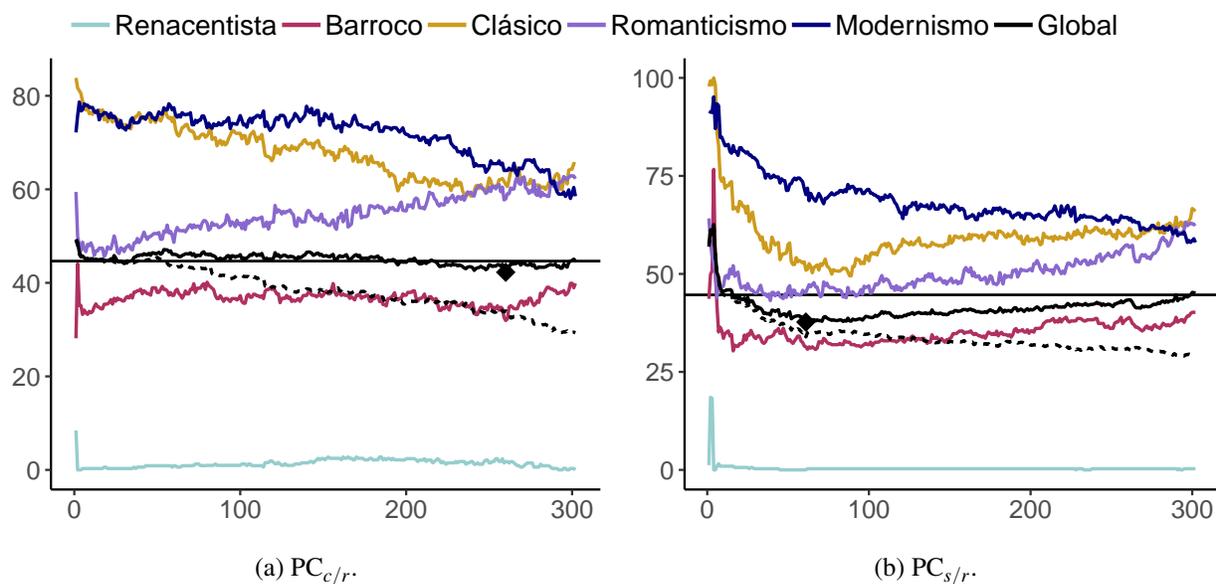


Figura 6.5: Tasas de error de clasificación (%) por corriente musical globales y por grupo validadas y aparente global estimadas mediante Regresión Ordinal respecto al número de Componentes Principales considerados para el ajuste del modelo, obtenidos al reescalar las variables ( $PC_{c/r}$ ) y sin reescalarlas ( $PC_{s/r}$ ). El punto indica la tasa de clasificación errónea global validada mínima; la recta constante, la tasa global del modelo con base en las observaciones originales; y la línea punteada, la tasa de error global aparente.

En cuanto a los criterios de información sin validar, el modelo con mayor bondad de ajuste es aquel construido con los primeros 61 PCs sin escalar, lo cual también indican el AIC, BIC y la tasa de clasificación errónea global validados. De hecho, la tasa de error de clasificación asociada a tal modelo es casi 5% menor que el resto, de manera que tal modelo corresponde al mejor de la sección, por lo que se procede a analizar con detenimiento.

	Modelo	Devianza	AIC	BIC	TCEG (%)
Aparentes	Variables originales	<b>2,057.09</b>	2,663.09	4,263.97	<b>29.4</b>
	$PC_{c/r}^{260}$	2,260.47	2,788.47	4,183.3	34
	$PC_{s/r}^{61}$	2,430.26	<b>2,560.26</b>	<b>2,903.68</b>	33.93
Validadas	Variables originales	<b>1,552.17</b>	2,158.17	3,691.43	44.65
	$PC_{c/r}^{260}$	1,734.17	2,262.17	3,598.08	42.26
	$PC_{s/r}^{61}$	1,929.56	<b>2,059.56</b>	<b>2,388.48</b>	<b>37.71</b>

Cuadro 6.4: Devianza, Criterio de Información de Akaike, Criterio de Información Bayesiano y tasa de clasificación errónea global, aparentes y validados, asociados a los modelos de Regresión Ordinal ajustados para clasificación de obras musicales por corriente histórica con las variables originales, 260 PCs obtenidos con reescalamiento en las variables ( $PC_{c/r}^{260}$ ) y 61 PCs obtenidos sin reescalarlas ( $PC_{s/r}^{61}$ ).

En primer lugar, el contraste de significancia del modelo sobre el modelo nulo se rechaza con un  $p$ -value menor a  $2 \cdot 10^{-16}$ , es decir, en efecto, hay al menos un coeficiente de regresión distinto de cero. Adicionalmente, los cuatro puntos de corte obtenidos mediante el ajuste del modelo resultaron ser significativos y están dados como se muestra en el siguiente cuadro. La interpretación en los puntos de corte subyace en la consideración de una variable latente asociada a la variable respuesta, esto es, una variable que no puede observarse y cuya información es percibida mediante la variable categórica ordinal. Así, si la variable latente toma valores menores a  $-4.33$ , el primer punto de corte, se observa que la variable categórica respuesta toma el valor Renacentista. En los casos en que la variable latente tomase valores entre  $-4.33$  y, el segundo punto de corte,  $4$ , se observarían que la variable respuesta toma el valor Barroco; y así sucesivamente.

Punto de corte	Estimación	Error Std.	IC Inf.	IC Sup.	Sesgo	Wald	$p$ -value
1   2	-4.33	0.46	-5.24	-3.43	2.7e-16	-9.4	$< 2 \cdot 10^{-16}$
2   3	4	0.34	3.34	4.67	5.3e-16	11.82	$< 2 \cdot 10^{-16}$
3   4	5.9	0.32	5.28	6.52	$< 2 \cdot 10^{-16}$	18.54	$< 2 \cdot 10^{-16}$
4   5	8.1	0.36	7.39	8.81	$< 2 \cdot 10^{-16}$	22.41	$< 2 \cdot 10^{-16}$

Cuadro 6.5: Estimaciones validadas, error estándar, intervalos de confianza, sesgo, estadístico del contraste de Wald y  $p$ -value de los puntos de corte obtenidos mediante Regresión Ordinal para clasificación por corriente histórica con base en los primeros 61 Componentes Principales obtenidos sin reescalar las variables originales. Por presentación, las categorías de la variable respuesta están codificadas del 1 al 5.

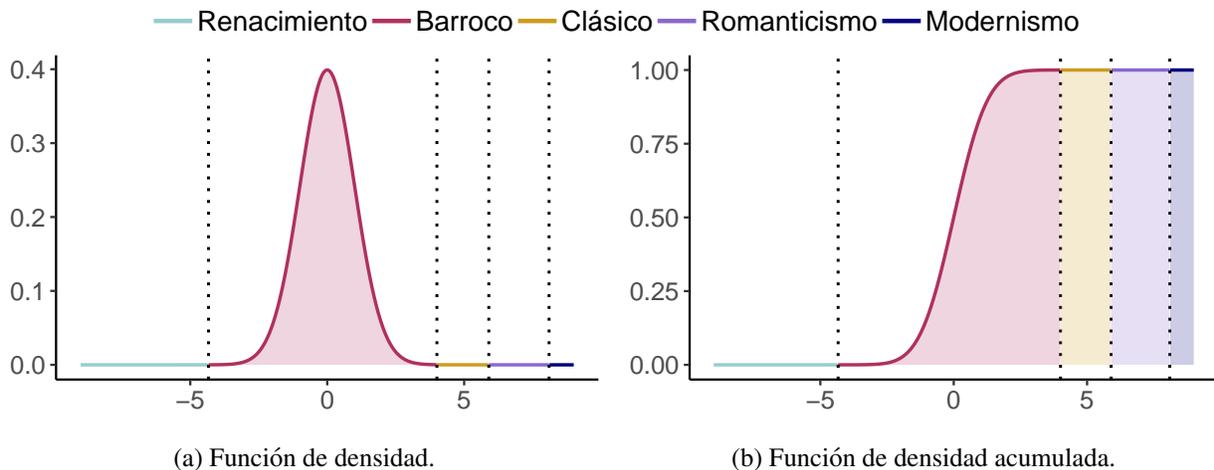


Figura 6.6: Funciones de densidad y densidad acumulada de la variable latente asociada a la variable respuesta del modelo Regresión Ordinal para clasificación por corriente histórica ajustado con los primeros 61 PCs obtenidos sin reescalar las variables. Las líneas punteadas indican los puntos de corte validados del modelo.

En la Figura 6.6 se muestran las funciones de densidad y densidad acumulada<sup>11</sup> asociadas a la variable latente del modelo, fijando como  $\underline{X}$  a los valores medios observados de cada uno de los predictores. Nótese que, por las estimaciones obtenidas, la media de la variable latente dada la respuesta media es cercana a cero, y toma mayormente los valores que remiten al periodo Barroco.

Por otro lado, se muestran las probabilidades estimadas validadas correspondientes a las probabilidades  $\hat{\mathbb{P}}(Y \leq i | \underline{X})$  y  $\hat{\mathbb{P}}(Y = i | \underline{X})$  para  $i \in \{\text{Renacentista}, \text{Barroco}, \text{Clásico}, \text{Romanticismo}, \text{Modernismo}\}$  proyectadas al primer Componente Principal. En la Figura 6.6a se evidencia la separación de la categoría correspondiente al Renacimiento respecto de las otras. Adicionalmente, en 6.6b se puede apreciar que el primer PC (con el cual se acumula casi el 100% de la varianza explicada por los componentes) aporta mucha información en el sentido de la asignación de las categorías Barroco y Renacentista. No obstante, las clases Romanticismo y Clásico parecen ser *poco claras* para el clasificador en cuanto al mismo PC, dado que los intervalos en que lo más probable es pertenecer a alguna tales clases son mucho menores en comparación a las demás y se superponen las curvas.

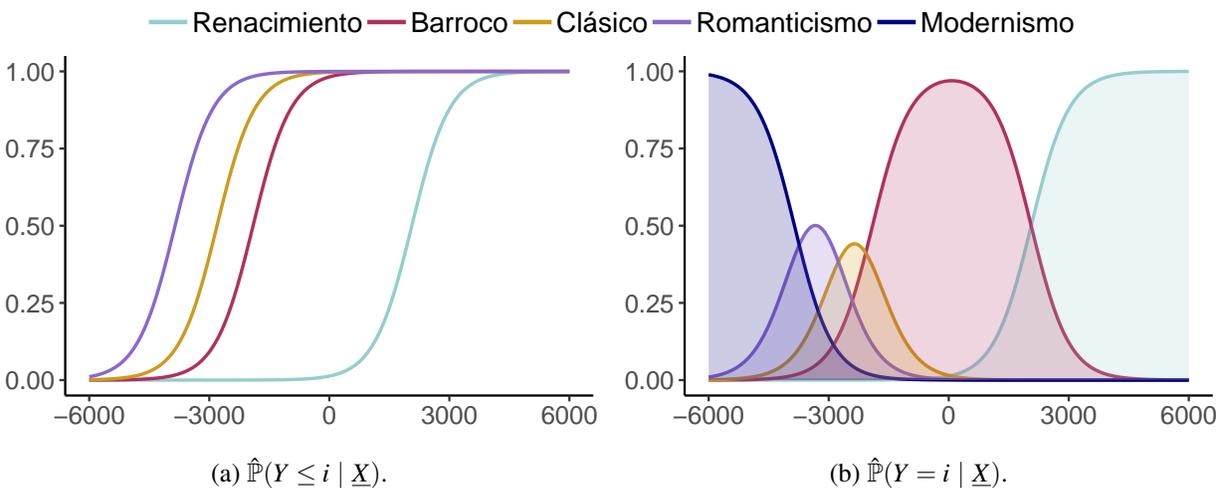


Figura 6.7: Proyección sobre el primer Componente Principal de las probabilidades estimadas por corriente musical resultantes del modelo de Regresión Ordinal ajustado con los primeros 61 PCs sin reescalamiento.

Finalmente, es debido hacer notar que los modelos de esta sección son mucho más estables que los correspondientes a Regresión Multinomial y la modelación es mucho más adecuada considerando la naturaleza de la variable respuesta. Sin embargo, el mejor modelo de Regresión Ordinal

<sup>11</sup>La cual corresponde a  $\Phi(\beta_{0j} - \underline{X}'\beta)$ , de acuerdo al planteamiento teórico del modelo.

no compite para nada con el mejor de tal apartado: ni en bondad de ajuste, ya que el BIC del primer modelo es 1,677.48 unidades menor, ni en poder predictivo, pues el segundo reporta una tasa de clasificación errónea global validada 11.67% mayor. La razón detrás de este hecho radica en el supuesto de proporcionalidad de momios. Para verificar tal supuesto –como es sugerido por Alan Agresti en [1, p. 204]–, se obtuvo la diferencia promedio sobre todas las observaciones del logaritmo de las razones de momios acumulados correspondientes la modelo de Regresión Multinomial ajustados exactamente con los mismos datos, i.e., el promedio de

$$\ln \left( \frac{\sum_{m=1}^{i+1} \mathbb{P}(Y = m | X) / \mathbb{P}(Y = 5 | X)}{\sum_{j=1}^i \mathbb{P}(Y = j | X) / \mathbb{P}(Y = 5 | X)} \right) = \ln \left( \sum_{m=1}^{i+1} \exp\{\beta_{0m} + x^t \beta_m\} \right) - \ln \left( \sum_{j=1}^i \exp\{\beta_{0j} + x^t \beta_j\} \right);$$

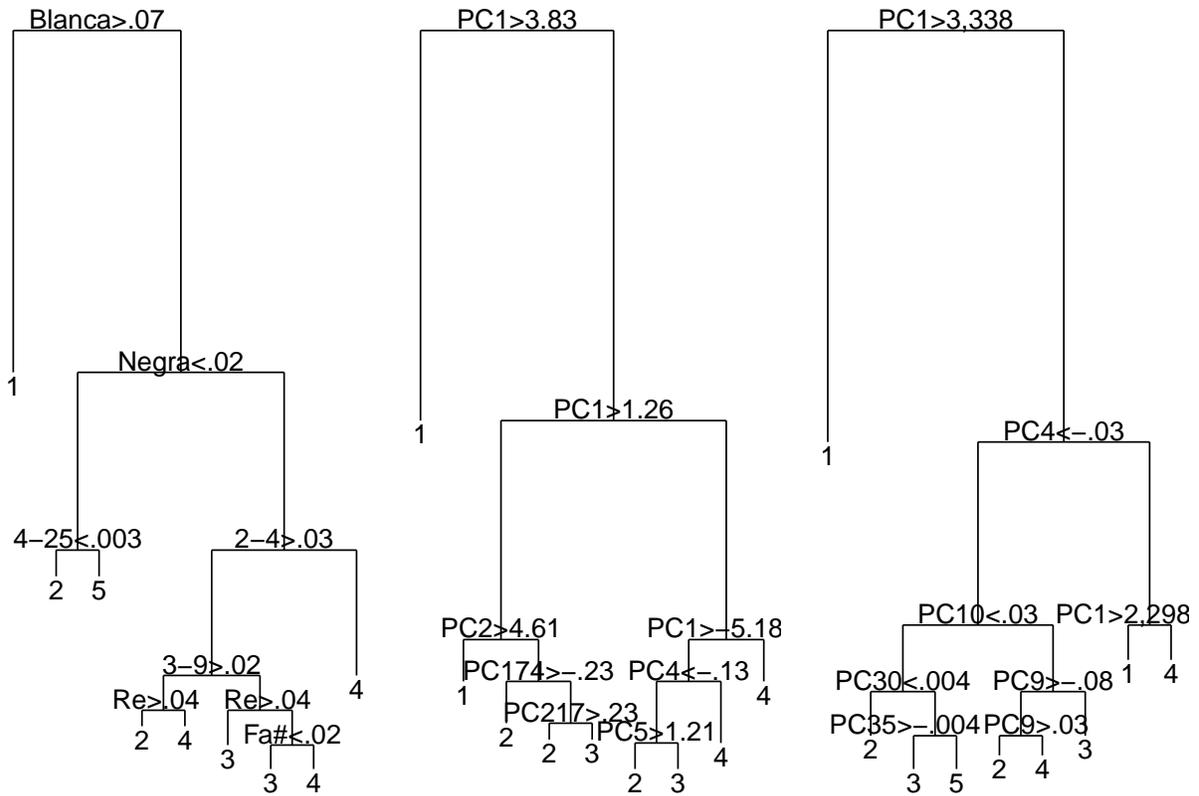
para cada  $i \in \{\text{Renacentista}, \text{Barroco}, \text{Clásico}\}$ . Nótese que de cumplirse el supuesto, los promedios de dicha diferencia sobre todas las observaciones rondarían el valor unitario. Sin embargo, las estimaciones superan la unidad en más de dos y doce unidades, como se muestran en el Cuadro 6.6, por lo que se tiene evidencia para sostener que tal supuesto no se cumple.

$i$	Media	Mínimo	Máximo	Desviación Estándar
Renacentista	12.13	0	31.23	7.57
Barroco	2.37	$1.89 \cdot 10^{-10}$	17.48	2.59
Clásico	2.53	$6.91 \cdot 10^{-5}$	24.01	2.47

Cuadro 6.6: Diferencia de logaritmos de las probabilidades acumuladas estimadas obtenidas mediante Regresión Multinomial para la verificación del supuesto de momios proporcionales del modelo ordinal ajustado con los primeros 61 PCs obtenidos sin reescalar las variables originales para la clasificación por corrientes musicales.

#### 6.1.4. Árboles de Clasificación

En esta sección se mostrarán los resultados correspondientes a haber construido Árboles de Decisión para la clasificación por corriente musical histórica. Primero, se realizó esta tarea mediante el método clásico utilizando como función de costo el Coeficiente de Gini y sobre las tres colecciones de datos contempladas hasta ahora. Los árboles obtenidos se muestran en la Figura 6.8.



(a) Variables originales.

(b)  $PC_{c/r}$ .

(c)  $PC_{s/r}$ .

Figura 6.8: Árboles de clasificación por corriente musical obtenidos con las variables originales y Componentes Principales obtenidos con reescalamiento en las variables ( $PC_{c/r}$ ) y sin reescalarlas ( $PC_{s/r}$ ). Las categorías de la variable respuesta están codificadas del 1 al 5 por presentación.

En primera instancia, los árboles asociados a ambas colecciones de Componentes Principales hacen uso principalmente de los primeros PCs, lo cual es lo natural; pero también resultan importantes los PCs 174 y 217, para los resultantes de reescalar las variables, y los PCs 30 y 35 en el caso restante. En el caso del árbol ajustado mediante las variables originales, cabe destacar que el algoritmo indica que la diferencia primordial entre las obras del Renacimiento y el resto es la probabilidad a largo plazo del uso de notas blancas, asignando a dicha clase los casos en que la probabilidad es mayor a 0.68. En la siguiente bifurcación del árbol se encuentra la única asignación al Modernismo, que constituye tener una probabilidad a largo plazo de realizar notas negras menor a 0.018, pero con una probabilidad mayor a 0.0027 de la utilización de tritonos, i.e., intervalos disonantes absolutamente, de acuerdo con la concepción contrapuntística de consonancia.

Por otro lado, en la Figura 6.9 se muestran los criterios de costo-complejidad aparentes y vali-

datos por *Cross-Validation* al aumentar el tamaño del árbol en términos de sus nodos finales. Como puede apreciarse, tales criterios se minimizan para el tamaño máximo contemplado, nueve nodos, por lo que carece de sentido podar los árboles en los tres casos. Los valores óptimos alcanzados por el criterio de costo-complejidad para los tres conjuntos de datos de manera respectiva son: 0.42, 0.502 y 0.52 (aparentes), y 0.48, 0.576 y 0.577 (validados). Por lo que puede decirse que el uso de las variables originales ante esta metodología es más eficiente, pues tales funciones de costo son menores a casi 0.1 para las estadísticas aparentes y validadas.

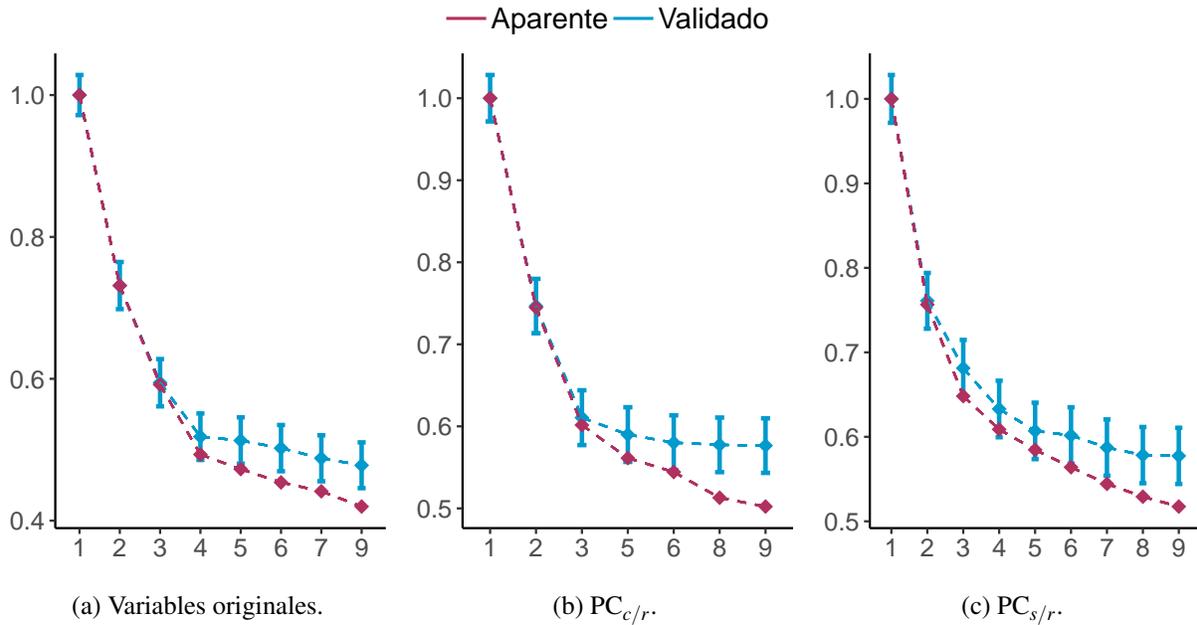


Figura 6.9: Criterio de costo-complejidad aparente y validado respecto el tamaño de los Árboles de Clasificación por corriente musical obtenidos con las variables originales y Componentes Principales obtenidos con reescalamiento en las variables ( $PC_{c/r}$ ) y sin reescalarlas ( $PC_{s/r}$ ). Se muestran también los intervalos de confianza del criterio validado.

Se prosiguió a la construcción de Árboles de Clasificación mediante *Bagging*, *Random Forests* (con  $l = \sqrt{p} \approx 17$ ) y *Boosting*, variando el número de Componentes Principales obtenidos al reescalar y sin reescalar las variables originales, así como con estas últimas. Obsérvese en la Figura 6.10 que para los tres métodos y las dos colecciones de Componentes Principales se obtuvieron resultados bastante fluctuantes al variar el número de PCs en la construcción, en particular, al aplicar *Random Forests*. También cabe hacer notar que con tales métodos resultaron tasas de clasificación errónea globales validadas sobre los Componentes Principales mayores a la obtenida con las variables originales, en concordancia con los resultados anteriores.

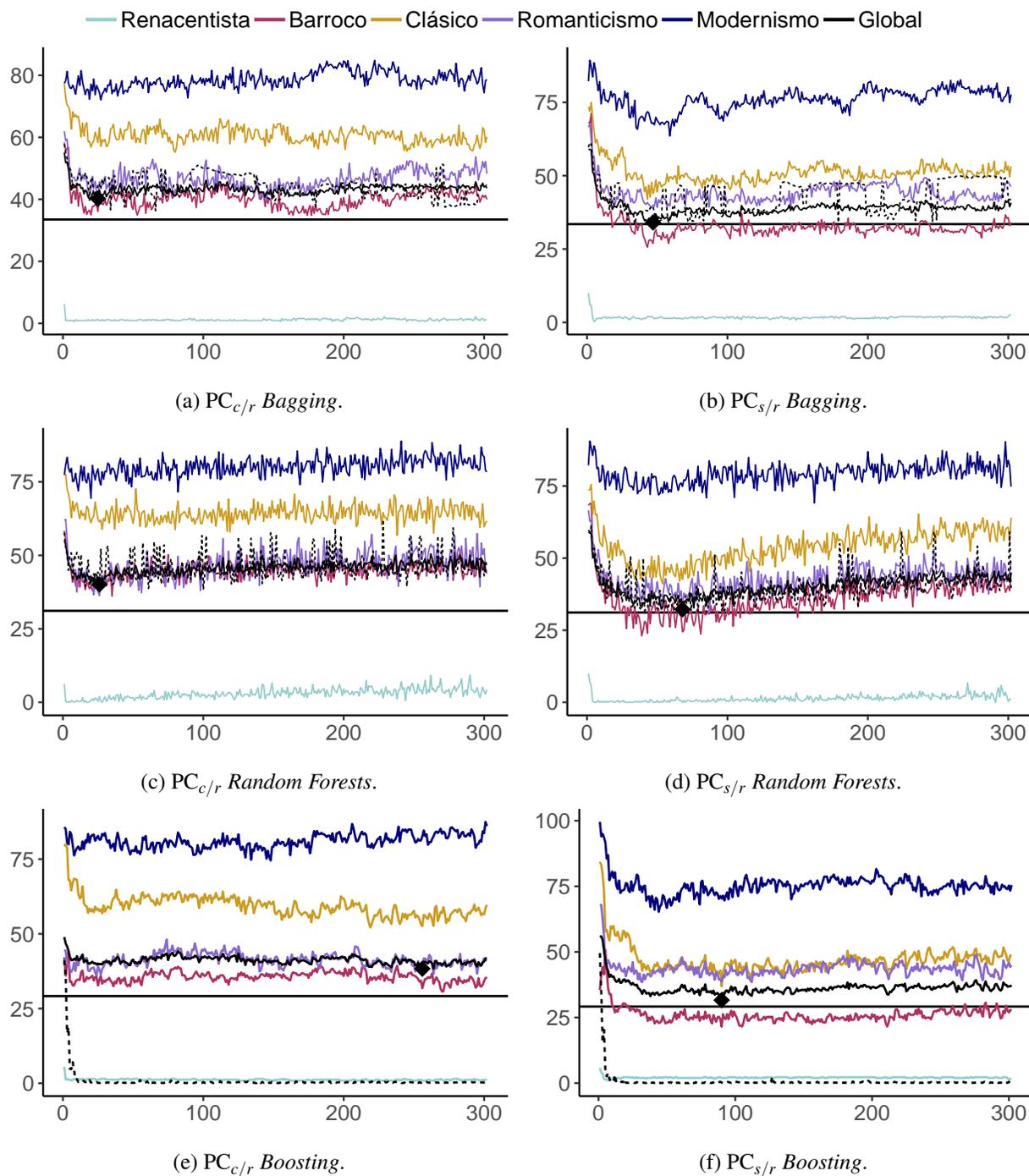


Figura 6.10: Tasas de error de clasificación (%) por corriente musical globales y por grupo validadas y aparente global estimadas mediante *Bagging*, *Random Forests* y *Boosting*, respecto al número de PCs considerados para el ajuste del modelo, obtenidos con ( $PC_{c/r}$ ) y sin reescalar las variables ( $PC_{s/r}$ ). El punto indica la tasa global validada mínima; la recta constante, la tasa global del modelo con las observaciones originales; y la línea punteada, el error global aparente.

Las mejores tasas de error globales validadas son alcanzadas para cada método y colección de datos con Componentes Principales, respectivamente, con: 25 y 26 PCs para *Bagging*, 26 y 68 para *Random Forests* y, 256 y 90 para *Boosting*. No obstante, como muestra el Cuadro 6.7, el mejor modelo de la sección corresponde al árbol ajustado por *Boosting* sobre las variables originales, cuya tasa de error de clasificación global validada resulta menor que el resto por al menos 1.96 %, pero la diferencia máxima es de 15.88 %.

Modelo		<i>Single Tree</i>	<i>Bagging</i>	<i>Random Forests</i>	<i>Boosting</i>
Aparentes	VARIABLES ORIGINALES	<b>32.21</b>	<b>34.58</b>	33.86	7.76
	PC <sub>c/r</sub>	38.53	44.67	36.4	0.55
	PC <sub>s/r</sub>	39.7	39.63	<b>33.72</b>	<b>0.21</b>
Validadas	VARIABLES ORIGINALES	<b>36.68</b>	<b>33.49</b>	<b>31.12</b>	<b>29.17</b>
	PC <sub>c/r</sub>	44.23	25, 40.2	26, 40.12	256, 38.33
	PC <sub>s/r</sub>	45.05	47, 34.05	68, 32.23	90, 31.61

Cuadro 6.7: Tasas de clasificación errónea globales, aparentes y validadas, obtenidas mediante Árboles de Clasificación para corrientes musicales con base en las variables originales y Componentes Principales obtenidos con reescalamiento en las variables (PC<sub>c/r</sub>) y sin reescalarlas (PC<sub>s/r</sub>). El número entero que precede las tasas validadas en los casos de modelos ajustados con PCs indica el número de componentes cuya inclusión al modelo minimiza el error del método.

### 6.1.5. *k*-Vecinos más Cercanos

Posteriormente, se llevó a cabo clasificación mediante *k*-Vecinos más Cercanos para las tres colecciones de datos completas, validando mediante *Cross-Validation* y variando el parámetro  $k \in \{1, \dots, 200\}$ . La variación de sus errores respecto  $k$  se muestra en la Figura 6.11.

Se obtuvo una tasa de error de clasificación global validada mínima fijando  $k = 78$  con el método aplicado a las variables originales, con  $k = 6$  para los Componentes Principales obtenidos reescalando las variables y con  $k = 172$  para el caso restante. Nótese que el mejor clasificador de esta sección corresponde al construido con base en los PCs obtenidos al reescalar las variables originales, resultando en una tasa de error global validada de 44.51 %; la cual, de hecho, no es competitiva respecto a las exhibidas hasta ahora, como se verá en las conclusiones.

Modelo	$k$	Renacentista	Barroco	Clásico	Romanticismo	Modernismo	Global
Originales	78	3.67	33.31	84.56	<b>58.15</b>	100	52.88
PC <sub>c/r</sub>	6	<b>0</b>	<b>9.51</b>	<b>65.56</b>	67.81	<b>94.17</b>	<b>44.51</b>
PC <sub>s/r</sub>	172	2.62	26.5	88.79	61.97	100	52.99

Cuadro 6.8: Tasas de clasificación errónea globales y por grupo validadas (%) obtenidas mediante  $k$ -Vecinos más Cercanos para corrientes musicales con base en las variables originales y Componentes Principales obtenidos con reescalamiento en las variables (PC<sub>c/r</sub>) y sin reescalarlas (PC<sub>s/r</sub>).

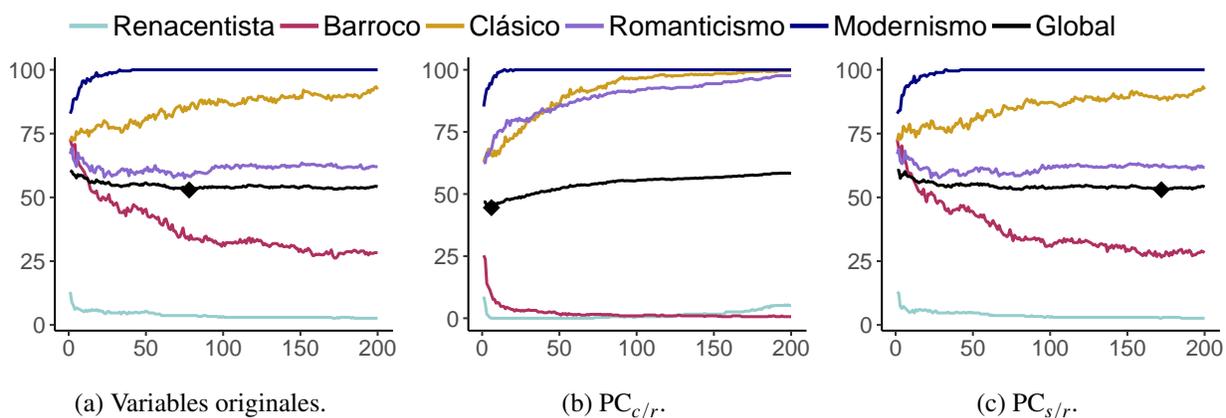


Figura 6.11: Tasas de error de clasificación (%) por corriente musical, globales y por grupo, validadas obtenidas mediante  $k$ -Vecinos más Cercanos, respecto al número de vecinos  $k$  y aplicado a las variables originales, Componentes Principales obtenidos sin reescalar las variables (PC<sub>c/r</sub>) y al reescalarlas (PC<sub>s/r</sub>).

### 6.1.6. Análisis de Discriminante

Siguiendo el mismo proceder expuesto hasta ahora, se ajustaron modelos de clasificación mediante Análisis de Discriminante Lineal y Cuadrático variando el número de Componentes Principales incluidos en la construcción; siendo la aplicación del segundo método posible hasta 157 PCs –debido al número de observaciones con que se cuenta–, e imposible de ajustar con las variables originales como predictores debido a deficiencia de rangos entre ellas derivada de su multicolinealidad.

Bajo esta metodología se observó que el LDA tendió menos al sobreajuste, como es evidenciado en las figuras 6.12a y 6.12b. Sin embargo, la aplicación del método para ambas colecciones de variables resultó en una mejora de apenas 0.5% en el caso de los 273 PCs obtenidos al reescalar las variables y del 6% con 92 PCs del restante conjunto, respecto al modelo ajustado con las variables originales, cuya tasa de error de clasificación global validada resultó en 34.7%.

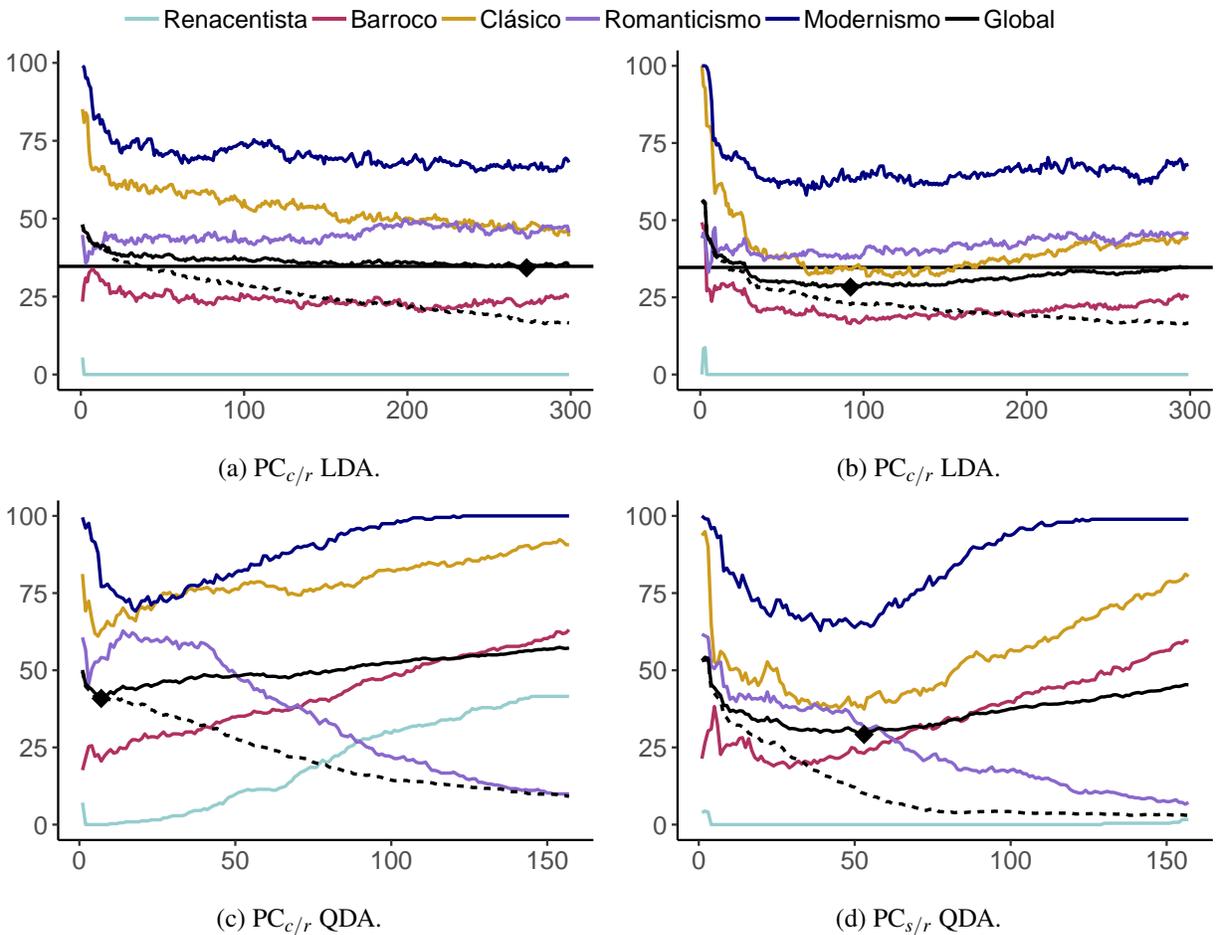


Figura 6.12: Tasas de error de clasificación (%) por corriente musical, globales y por grupo, validadas de los modelos de LDA y QDA respecto al número de PCs considerados, obtenidos con reescalamiento en las variables ( $PC_{c/r}$ ) y sin reescalarlas ( $PC_{s/r}$ ). La línea punteada corresponde al error global aparente y la recta constante a la tasa de clasificación errónea validada del modelo con las observaciones originales (para LDA).

Por otro lado, como puede apreciarse en los incisos c y d de la Figura 6.12, el sobreajuste de los modelos obtenidos mediante QDA es muy marcado y comienza antes de la inclusión de diez PCs para la primera colección de datos y alrededor de los 75 para el caso restante. No obstante, la mejor tasa de clasificación errónea global aparente del apartado fue lograda mediante este método y los primeros 53 Componentes Principales derivados de no realizar reescalamiento en las variables originales al realizar PCA, valuándose en 10.1%. La tasa de error global validada relativa al mismo modelo es de 29.21%, la cual es 0.9% mayor que la correspondiente al mejor modelo de esta sección: LDA aplicado a los primeros 92 PCs obtenidos sin reescalamiento en las variables. Es decir, con el mejor modelo derivado de estos métodos, se alcanzaron tasas de error de clasificación globales de 22.66% (aparente) y 28.31% (validada), como se presenta a continuación.

Método	Modelo	Aparentes	Validadas
LDA	Variables Originales	16.62	34.7
	PC <sub>c/r</sub> <sup>273</sup>	17.11	34.23
	PC <sub>s/r</sub> <sup>92</sup>	22.66	<b>28.31</b>
QDA	PC <sub>c/r</sub> <sup>7</sup>	40.87	40.88
	PC <sub>s/r</sub> <sup>53</sup>	<b>10.1</b>	29.21

Cuadro 6.9: Tasas de clasificación errónea globales (%), aparentes y validadas, obtenidas mediante LDA y QDA para corrientes musicales con base en las variables originales y Componentes Principales obtenidos con reescalamiento en las variables (PC<sub>c/r</sub>) y sin reescalarlas (PC<sub>s/r</sub>).

### 6.1.7. Clasificador Bayesiano sobre los factores de consonancia y disonancia

La modelación llevada a cabo mediante Campos Aleatorios sugiere una manera intuitiva de llevar a cabo clasificación: los factores de consonancia y disonancia  $\hat{\theta}_y = (\hat{\theta}_{\mathcal{C}}^y, \hat{\theta}_{\mathcal{D}}^y)$  para cada periodo histórico y autor ( $y$ ) son estimados a partir de la expresión

$$\pi(\underline{X} | \underline{\theta}_y) = \pi(\underline{X} | Y = y);$$

donde  $Y$  denota la clase de la obra musical  $\underline{X}$ .

De esta manera, puede obtenerse la función de probabilidad a posteriori con el Teorema de Bayes

$$\mathbb{P}(Y = y | \underline{X}) = \frac{\pi(\underline{X} | Y = y) \cdot \pi_Y(y)}{\sum_{z \in C} \pi(\underline{X} | Y = z) \cdot \pi_Y(z)};$$

donde  $\pi_Y$  es la función de probabilidad a priori de  $Y$  y  $C$  el conjunto de clases, valores que  $Y$  toma. La cual, para fines del presente se fijó como las proporciones de cada clase en la muestra para cada caso. Luego, la predicción de nuevas observaciones consistirá en la asignación de la etiqueta siguiente

$$\hat{y} = \operatorname{argm\acute{a}x}_{y \in C} \mathbb{P}(Y = y | \underline{X}).$$

Ahora bien, nótese que maximizar la función de probabilidad a posteriori es equivalente a maximizar respecto a  $y$  la expresión  $\pi(\underline{X} | Y = y) \cdot \pi_Y(y)$  y, a su vez, a maximizar su logaritmo, i.e.,  $\ell(\underline{X}; \hat{\theta}) + \log \pi_Y(y)$ , donde la función de log-Verosimilitud puede expresarse en términos de la medida de Gibbs del campo como se ha hecho en los Capítulos 3 y 5. De esta manera, para cada pieza fue calculada la expresión anterior donde la distribución a priori de  $Y$  está dada por las proporciones de obras correspondientes cada periodo en la muestra, i.e.,

$$\pi_Y(y) = \begin{cases} \frac{300}{1456} & \text{si } y = \text{Renacentista } \text{ó } y = \text{Clásico,} \\ \frac{315}{1456} & \text{si } y = \text{Barroco,} \\ \frac{339}{1456} & \text{si } y = \text{Romanticismo,} \\ \frac{202}{1456} & \text{si } y = \text{Modernismo.} \end{cases}$$

A continuación se presenta la matriz de confusión asociada a la clasificación descrita, así como las tasas de error de clasificación por grupo. La clasificación por esta vía representa una tasa de clasificación errónea global de 65.45%. Aquella clase que clasifica mejor es la correspondiente al Renacimiento; aunque el periodo Clásico y Moderno resultan en tasas de error del 100% debido a que el clasificador no asigna a ninguna obra musical tales corrientes históricas. Además, cabe hacer notar que las tasas mencionadas tienen carácter aparente, es decir, subestiman el error de predicción. Razón por la cual, desde este punto se reconoce que no es un buen clasificador para el problema en cuestión, por lo que, aunado al alto costo computacional de la estimación de los parámetros  $\hat{\theta}$  para cada clase, se decidió no validar las tasas.

$Y \setminus \hat{Y}$	Renacentista	Barroco	Clásico	Romanticismo	Modernismo
Renacentista	223	0	0	77	0
Barroco	14	180	0	121	0
Clásico	10	178	0	112	0
Romanticismo	7	231	0	100	1
Modernismo	9	134	0	59	0
TCEA (%)	25.67	42.86	100	70.5	100

Cuadro 6.10: Matriz de confusión asociada a la clasificación por corriente histórica y tasas de clasificación errónea aparentes por grupo mediante el clasificador Bayesiano sobre los factores de consonancia y disonancia.

### 6.1.8. Conclusiones

A manera de conclusión, obsérvese el siguiente cuadro, en que se presentan las tasas de error de clasificación globales aparentes y validadas de los mejores modelos de la sección, junto con sus intervalos de confianza al 5 % de significancia, desviación estándar y sesgo estimados de las últimas. En primera instancia, la mejor tasa global aparente es lograda mediante *Boosting* y las variables originales, la cual resultó en 7.76 %. No obstante, la tasa de error validada asociada al mismo modelo resultó en 29.17 %, con la que, de hecho, se constituye el modelo con mayor diferencia entre tales tasas, siendo de 21.42 %.

Método	Modelo	TECGA	TECGV	Intervalos de Confianza (95 %)
<b>R. Multinomial</b>	$PC_{s/r}^{41}$	<b>23.42</b>	<b>26.04</b>	
R. M. Lasso	$PC_{s/r}^{44}$ $\lambda=2.7e-4$	23.35	26.31	
R. M. Ridge	$PC_{s/r}^{49}$ $\lambda=2.7e-4$	21.91	26.41	
R. M. Elastic Net	$PC_{s/r}^{44}$ $\lambda=5.4e-4$	23.56	26.51	
R. Ordinal	$PC_{s/r}^{61}$	33.93	37.71	
Single Tree	V. O.	32.21	36.68	
Bagging	V. O.	34.58	33.49	
Random Forests	V. O., $l = 17$	33.86	31.12	
Boosting	V. O.	<b>7.76</b>	29.17	
k-NN	$PC_{c/r}^{302}, k = 6$	-	44.51	
LDA	$PC_{s/r}^{92}$	22.66	28.31	
QDA	$PC_{s/r}^{53}$	10.1	29.21	
C. Bayesiano	$\hat{\theta}_{\mathcal{L}}, \hat{\theta}_{\mathcal{D}}$	65.45	-	

Cuadro 6.11: Tasas de error de clasificación globales aparentes (TECGA) y validadas (TECGV) correspondientes a clasificación por corriente musical histórica por los métodos presentados. En la columna modelo se especifican las variables contempladas para el ajuste del mismo, así como los parámetros correspondientes al método utilizado. Las tasas de error se muestran en porcentajes. V. O. hace alusión a las variables originales.

Adicionalmente, cabe destacar que la mayor tasa aparente registrada corresponde al Clasificador Bayesiano sobre los factores de consonancia y disonancia planteados, la cual resultó en 65.45 %

y es casi 0.31 unidades mayor que la siguiente tasa de mayor a menor dada mediante *Bagging* y las variables originales. Por lo que, en realidad, tal clasificador no resulta ser nada eficiente, más aún, considerando que es una tasa aparente. Por otro lado, aquella tasa con sesgo estimado menor resulta ser la asociada al modelo con *Boosting*, tan pequeña que de hecho se pierden sus decimales por la precisión de la máquina. La cual viene seguida del modelo de Regresión Multinomial ajustado mediante los primeros 41 Componentes Principales obtenidos sin reescalar las variables originales, con un sesgo estimado de  $5.5 \cdot 10^{-18}$ .

Asimismo, aquel modelo cuya desviación estándar en la tasa de clasificación errónea global validada es menor que el resto corresponde al Análisis de Discriminante Lineal construido sobre los primeros 92 Componentes Principales de la misma colección mencionada anteriormente, con una estimación de 0.009. El cual, es seguido por la desviación estándar de la tasa validada del modelo de Regresión Multinomial *Elastic Net* construido con los primeros 44 Componentes Principales obtenidos sin reescalar las variables originales, parámetro de regularización  $\lambda = 0.00054$  y  $\alpha = 0.5$ , sólo mayor por 0.002 unidades. Finalmente, cabe destacar que le sigue el modelo de Regresión Multinomial sin regularización, con una desviación estimada de 0.013.

Fijando atención a las tasas validadas, nótese que la más baja está valuada en 26.04% y corresponde al mismo modelo de Regresión Multinomial. Tal tasa es seguida ascendentemente por el modelo de Regresión Multinomial Lasso ajustado sobre los primeros 44 PCs sin reescalamiento de las variables y parámetro de regularización  $\lambda = 0.00027$ , resultando en una tasa de error de clasificación global validada mayor por 0.27%, con desviación estándar prácticamente igual, aunque mayor sesgo estimado. Esta es la razón por la cual se concluye que el mejor modelo ajustado para la clasificación por corriente musical es el construido con base en los primeros 41 Componentes Principales obtenidos sin reescalamiento y mediante Regresión Multinomial –por lo que se analizó con tal detenimiento en el apartado respectivo–.

En el cuadro siguiente se presentan las tasas de clasificación errónea por grupo validadas asociadas a los modelos presentados en el Cuadro 6.11. Primero, la categoría que es más eficientemente clasificada es la Renacentista, como se esperaba desde el inicio con el análisis exploratorio de resultados del Capítulo 5, puesto que en la mayoría de los modelos ajustados se logra una tasa de error de clasificación validada del 0%, i.e., acierto del 100%. En cuanto a tasa de acierto, le sigue la clase Barroco, la cual es clasificada con mayor éxito mediante *k*-Vecinos más Cercanos con un 90.49% de acierto; aunque el modelo de Regresión Multinomial registra una tasa de acierto co-

rrespondiente a tal grupo del 79.98 %. Además, en general, las etiquetas Clásico y Romanticismo son mucho más difíciles de clasificar por los métodos aplicados, siendo en todos los casos las tasas validadas mayores a 30 %.

Método	Renacentista	Barroco	Clásico	Romanticismo	Modernismo
<b>R. Multinomial</b>	<b>0</b>	<b>20.02</b>	<b>31.98</b>	<b>32.68</b>	<b>54.07</b>
R. M. Lasso	0.3	19.38	32.3	34.2	53.58
R. M. Ridge	1.5	19.75	32.54	33.89	<b>51.89</b>
R. M. <i>Elastic Net</i>	0.29	19.38	32.3	34.17	54.99
R. Ordinal	0	31.42	50.93	45.26	70.23
<i>Single Tree</i>	0	25.4	48	43.66	80.2
<i>Bagging</i>	0	26.21	40.96	41.62	69.17
<i>Random Forests</i>	0	21.12	39.24	33.07	76.52
<i>Boosting</i>	0	23.29	34.05	35.05	62.99
<i>k</i> -NN	0	<b>9.51</b>	65.56	67.81	94.17
LDA	0	16.45	34.33	38.05	62.64
QDA	0	23.06	37.4	<b>31.62</b>	65.13
C. Bayesiano	25.67	42.86	100	70.5	100

Cuadro 6.12: Tasas de error de clasificación por grupo validadas correspondientes a clasificación por corriente musical histórica por todos los métodos y modelos presentados en la tabla anterior.

No obstante, la categoría de la variable respuesta en cuestión que resulta altamente problemática para clasificar por corrientes musicales es el Modernismo: la mejor tasa de acierto lograda para este grupo es de 48.11 %, la cual es apenas mayor por 28.11 % a la tasa correspondiente a su asignación al azar. Incluso, cabe destacar que tres de los modelos presentados<sup>12</sup> reportan tasas de error mayores a 80 % en cuanto a la clasificación de tal categoría, i.e., con rendimiento esperado peor que asignando el Modernismo al azar equiprobablemente sobre todas las clases de  $Y$ . Esta problemática puede ser resultado directo de la naturaleza de la misma corriente; puesto que durante ella se llevaron a cabo experimentaciones que sobrepasaron las líneas convencionales de la Teoría Musical, como en el caso de la Música Atonal o la Música Pandiatónica, por mencionar algunas. Como consecuencia, a pesar de pertenecer a tal categoría, las obras musicales contempladas contienen diferencias fundamentales entre sí.

<sup>12</sup>*Single Tree*, *k*-Vecinos más Cercanos y el Clasificador Bayesiano sobre los factores de consonancia y disonancia, con tasas de clasificación erróneas validadas para el grupo *Modernismo* valuadas en 80.2%, 94.17% y 100%, respectivamente.

Así como que, por su naturaleza ajena a la Música de armonía tonal en muchos de los casos, la dificultad de clasificar exitosamente a la Música Modernista podría recaer en el hecho de que en la modelación llevada a cabo falte algún elemento que la represente y diferencie del resto específicamente<sup>13</sup>. De manera que nace la motivación de llevar a cabo la clasificación restringida a la Música con armonía tonal, por así llamarle, es decir, obras musicales que no pertenezcan a la Música Postonal del Modernismo, como se lleva a cabo en la siguiente sección.

## 6.2. Clasificación por corrientes históricas con armonía tonal

Bajo la motivación introducida al final de la sección anterior, en la presente se muestran los resultados relativos a clasificación por corrientes musicales restringida a las obras musicales compuestas con base en armonía tonal, por lo que se contemplarán únicamente las anteriores al Modernismo. En este contexto, la variable respuesta  $Y$  toma valores en el conjunto  $\{\text{Renacentista}, \text{Barroco}, \text{Clásico}, \text{Romanticismo}\}$  y se cuenta con 1,254 observaciones de los mismos 302 predictores. Cabe destacar que la construcción de modelos se realizó de manera completamente análoga a la presentada en la sección anterior:

- Con base en las variables originales<sup>14</sup> y los Componentes Principales, obtenidos al reescalarlas y obtenidos sin reescalamiento. Cabe destacar que estos corresponden a los resultantes del PCA sobre la muestra completa, es decir, incluyendo el Modernismo, con la intención de trabajar sobre la representación de la información más *rica* posible.
- Variando el número de Componentes Principales de ambos conjuntos de datos incluidos en los modelos con el fin de minimizar las tasas de clasificación errónea globales.
- Validando las tasas de clasificación erróneas grupales y globales mediante *5-Fold Cross-Validation*, manteniendo los conjuntos de prueba y entrenamiento exactamente iguales para cada modelo con el fin de obtener estadísticas comparables entre sí.

---

<sup>13</sup>Recuérdese que la modelación *ideal* propuesta es prácticamente imposible de llevar a cabo por su costo computacional, por lo que se sustituyó con la alternativa presentada.

<sup>14</sup>Excepto en los casos de Regresión y Análisis de Discriminante, en que se excluyen las últimas entradas de las probabilidades límite estimadas de las Cadenas de Markov asociadas a cada pieza musical, con el fin de evitar la violación del supuesto de independencia de los predictores.

- Aplicando los métodos de Aprendizaje Estadístico: Regresión Multinomial, Regresión Multinomial Regularizada (Lasso, Ridge y *Elastic Net* –con  $\alpha = 0.5$  para las variables originales y los Componentes Principales obtenidos al reescalar las variables, y con  $\alpha = 0.517$  para los PCs obtenidos sin reescalamiento–), Regresión Ordinal, Árboles de Clasificación (*Single Tree*, *Pruning*, *Bagging*, *Random Forests* y *Boosting*), *k*-Vecinos más Cercanos, Análisis de Discriminante (Lineal y Cuadrático) y el Clasificador Bayesiano sobre los factores de disonancia y consonancia estimados asociados a la modelación por Campos Aleatorios.

Adicionalmente, con la intención de evitar repetitividad y, en medida de que en la sección anterior fue descrita con detenimiento la metodología aplicada, se presentan a continuación los resultados obtenidos de manera sintetizada, es decir, las tasas de error de clasificación validadas y aparentes correspondientes al mejor modelo resultante de la aplicación de cada método. Asimismo, se procederá a analizar detalladamente únicamente el modelo representativo de la sección, en cuanto a su poder predictivo.

Método	V. O.	Modelo		TECGV		
		PC <sub>c/r</sub>	PC <sub>s/r</sub>	V. O.	PC <sub>c/r</sub>	PC <sub>s/r</sub>
R. Multinomial	-	59	43	38.81	27.79	<b>18.62</b>
R. M. Lasso	$\lambda = 0.042$	70, $\lambda=0.00035$	42, $\lambda=0.0003$	29.17	27.64	<b>18.14</b>
R. M. Ridge	$\lambda = 0.042$	132, $\lambda=0.035$	84 $\lambda=0.03$	25.07	26.12	<b>18.21</b>
R. M. <i>Elastic Net</i>	$\alpha=0.5,$ $\lambda=0.084$	70, $\alpha=0.5,$ $\lambda=0.00069$	55, $\alpha=0.517,$ $\lambda=0.00068$	28.53	27.8	<b>17.9</b>
R. Ordinal	-	167	57	34.9	31.29	<b>26.82</b>
<i>Single Tree</i>	-	302	302	<b>27.99</b>	35.65	36.68
<i>Bagging</i>	-	135	34	<b>21.93</b>	30.4	25.37
<i>Random Forests</i>	$l = 17$	19, $l = 17$	67, $l = 17$	<b>20.38</b>	31.26	23.35
<i>Boosting</i>	-	287	70	<b>22.64</b>	28.69	24.98
<i>k</i> -NN	$k = 189$	$k = 5$	$k = 189$	45.04	<b>37.28</b>	45.2
LDA	-	131	80	27.37	25.45	<b>18.53</b>
QDA	-	7	35	-	32.94	<b>20.52</b>

Cuadro 6.13: Tasas de error de clasificación globales validadas (%) para clasificación por corriente musical histórica restringida a Música con armonía tonal con base en las variables originales (V.O.) y los Componentes Principales obtenidos al reescalarlas (PC<sub>c/r</sub>) y sin reescalamiento (PC<sub>s/r</sub>). Las columnas “Modelo” describen los parámetros de ajuste de cada modelo y la cantidad de PCs usados en los casos correspondientes.

Los modelos de Regresión ajustados resultaron en menores TECG validadas con los Componentes Principales obtenidos sin reescalar las variables originales como predictores. Al construir Árboles de Clasificación, las variables originales involucraron mayor poder predictivo que el uso de Componentes Principales, como en la sección anterior. Asimismo, con la clasificación llevada cabo mediante  $k$ -Vecinos más Cercanos sobre los Componentes Principales obtenidos con reescalamiento se consiguió una tasa de error menor por casi 0.08 unidades en comparación a los ajustados por el mismo método con los conjuntos de datos restantes. Por otro lado, el Análisis de Discriminante resultó ser más eficiente para predecir al ajustarse sobre los PCs obtenidos sin reescalar las variables originales, en comparación con su contraparte derivada del PCA. Cabe destacar que, debido a la cantidad de observaciones y predictores, así como a la multicolinealidad asociada a las variables originales, no fue numéricamente posible realizar Análisis de Discriminante con las mismas.

Método	Modelo	TECGA	TECGV	Intervalos de Confianza (95 %)
R. Multinomial	PC $_{s/r}^{43}$	15.31	18.62	
R. M. Lasso	PC $_{s/r}^{42}$ $\lambda=0.0003$	15.39	18.14	
R. M. Ridge	PC $_{s/r}^{49}$ $\lambda=0.03$	13.4	18.21	
<b>R. M. Elastic Net</b>	PC $_{s/r}^{55}$ , $\alpha=0.517$ $\lambda=0.00068$	<b>14.27</b>	<b>17.9</b>	
R. Ordinal	PC $_{s/r}^{57}$	23.29	26.82	
Single Tree	V. O.	21.29	27.99	
Bagging	V. O.	18.42	21.93	
Random Forests	V. O., $l = 17$	21.69	20.38	
Boosting	V. O.	<b>8.29</b>	22.64	
$k$ -NN	PC $_{c/r}^{302}$ , $k = 5$	-	37.28	
LDA	PC $_{s/r}^{80}$	15.47	18.53	
QDA	PC $_{s/r}^{35}$	12.92	20.52	
C. Bayesiano	$\hat{\theta}_{\mathcal{C}}, \hat{\theta}_{\mathcal{D}}$	59.89	-	

Cuadro 6.14: Tasas de error de clasificación globales aparentes (TECGA) y validadas (TECGV) correspondientes a clasificación por corriente musical histórica restringida a Música con armonía tonal por todos los métodos presentados (%). En la columna “Modelo” se especifican las variables contempladas para su ajuste, así como los parámetros correspondientes al método utilizado. V. O. hace alusión a las variables originales.

Nótese que, como se muestra en el Cuadro 6.14, en general, el mejor modelo para la clasificación de obras musicales en la corriente histórica a la que pertenecen (restringido a la Música con armonía tonal), en cuanto a su poder predictivo, corresponde al ajustado sobre los PCs obtenidos sin reescalar las variables originales mediante Regresión Multinomial regularizada con *Elastic Net* y los parámetros  $\alpha = 0.517$  y  $\lambda = 0.00068$ . Tal modelo implica una tasa de clasificación errónea global validada de 17.9%, la cual es entre 0.24% y 19.38% menor que las resultantes del resto de los métodos aplicados representados por los mejores modelos obtenidos respectivamente.

Además, la tasa global aparente asociada al modelo regularizado por *Elastic Net* es la tercera menor de todas las obtenidas, siendo, de hecho, menor que su contraparte validada por sólo 3.63%, lo que habla de que el modelo en cuestión no está saturado. Asimismo, el intervalo al 0.95 de confianza relativo a la estimación del error por *Cross-Validation* es [14.62, 21.19], contando con una desviación estándar estimada de 0.017 y un sesgo tan cercano a cero que se pierden sus decimales por la precisión de la máquina.

Método	Renacentista	Barroco	Clásico	Romanticismo
R. Multinomial	0	17.95	31.34	24.49
R. M. Lasso	0.34	18.26	<b>30.32</b>	23.01
R. M. Ridge	1.3	18.34	31.47	21.12
<b>R. M. <i>Elastic Net</i></b>	<b>1.01</b>	<b>18.4</b>	<b>30.6</b>	<b>20.75</b>
R. Ordinal	0	26.65	49.08	26.57
<i>Single Tree</i>	0	27.62	46	37.17
<i>Bagging</i>	2.38	27.73	40.97	29.47
<i>Random Forests</i>	0.29	24.25	42.28	26.33
<i>Boosting</i>	2.06	25.66	40.5	30.9
<i>k</i> -NN	0.29	<b>12.04</b>	64.614	68.39
LDA	0	17.31	31.97	24.46
QDA	0	19.46	35.11	25.82
C. Bayesiano	25.67	42.86	100	70.5

Cuadro 6.15: Tasas de error de clasificación por grupo validadas (%) correspondientes a clasificación por corriente histórica restringida a Música con armonía tonal por los métodos presentados en la tabla anterior.

Ante esta nueva tarea de clasificación, las categorías que los modelos suelen asignar con menos acierto son el *Clásico* y el *Romanticismo*. El modelo correspondiente a *Elastic Net* constituye aquél que más exitosamente clasifica la última categoría y el segundo mejor para las obras pertenecientes al periodo *Clásico*, con tasas de error validadas respectivas a dichos grupos de 18.4%

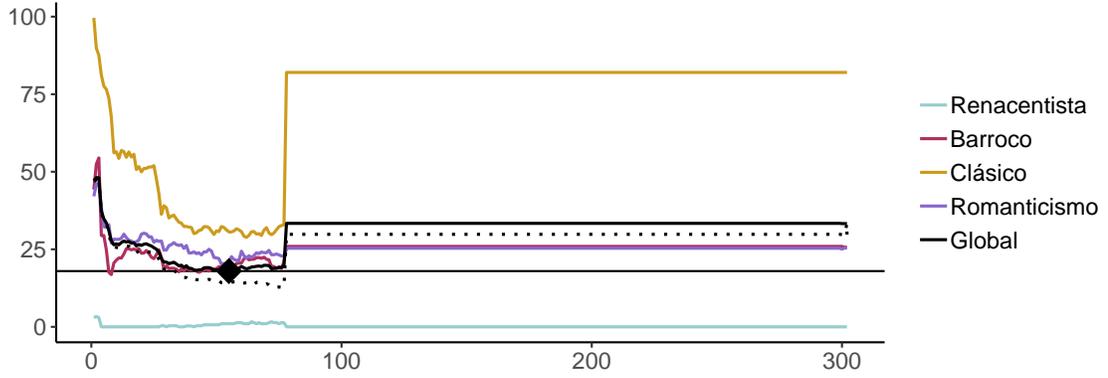
y 30.6%. Adicionalmente, se observa la misma tendencia a las tasas de error de clasificación validadas asociadas a la clase *Renacentista* que en la sección anterior, clasificándola con 100% de acierto mediante algunos modelos, como lo son: Regresión Multinomial y Ordinal, o Análisis de Discriminante.

Es de hacer notar que en esta sección se logró una tasa de error de clasificación global validada menor por 8.14% que la tasa de 26.04% correspondiente al modelo de Regresión Multinomial de la sección anterior. Asimismo, las tasas validadas relativas a la clasificación de los grupos *Barroco*, *Clásico* y *Romanticismo* se reducen respectivamente 1.62%, 0.64% y 2.08% al aplicar este ejercicio. Aunque la tasa asociada a la categoría *Renacentista* pasa de ser nula a valuarse en 1.01%.

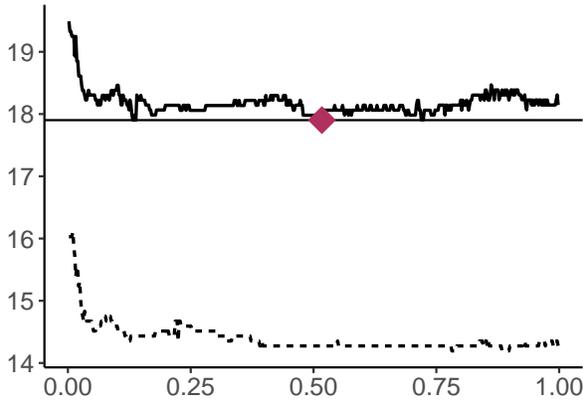
### **6.2.1. Regresión Mutlinomial regularizada mediante *Elastic Net***

A continuación se presenta el análisis del modelo obtenido aplicando Regresión Multinomial *Elastic Net* sobre los Componentes Principales derivados del análisis al omitir el reescalamiento de las variables originales. En primera instancia, se evaluaron las tasas de error de clasificación al variar el número de PCs incluidos en el ajuste del modelo fijando como parámetro  $\alpha = 0.5$  y obteniendo los parámetros de regularización  $\lambda$  óptimos para cada modelo. De aquí se obtuvo que con los primeros 55 PCs se minimiza la tasa de error de clasificación global validada, valuada en 17.98%; esto se ilustra en la Figura 6.13a, donde también se evidencia que a partir de la inclusión en el modelo de los primeros 85 PCs aproximadamente se llega a su saturación.

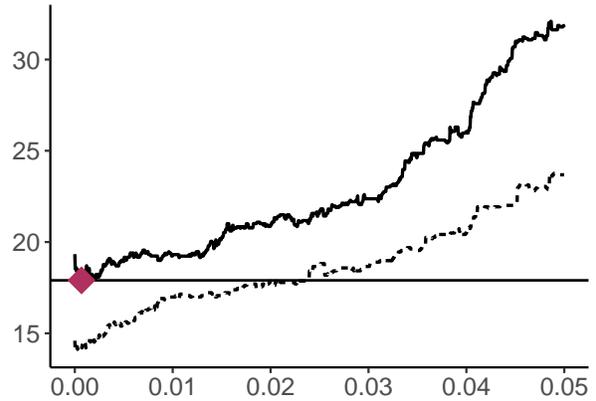
De manera posterior, se procedió a variar el parámetro  $\alpha$  fijando el ajuste de los modelos sobre los primeros 55 PCs y evaluar las tasas de clasificación errónea globales, aparentes y validadas, con el fin de encontrar su valor óptimo; de manera que con  $\alpha = 0.517$  y estableciendo  $\lambda = 0.00058$ , que es el valor automatizado del algoritmo computacional, se resulta en una TCEG validada de 17.902%. Finalmente, se decidió explorar sobre el parámetro de regularización manteniendo constantes el número de PCs incluidos (55) y el parámetro  $\alpha$ ; su comportamiento observado, junto con el del parámetro anterior, pueden consultarse en la Figura 6.13a y 6.13b. Finalmente, se halló que con  $\lambda = 0.00068$  el modelo ajustado resulta en su versión óptima en cuanto a poder predictivo y cuenta con las características enunciadas a continuación.



(a) Variación respecto  $x$ , donde  $PC_{s/r}^x$  y  $\alpha = 0.5$ .



(b) Variación respecto  $\alpha$ , con  $PC_{s/r}^{55}$ .



(c) Variación respecto  $\lambda$ , con  $PC_{s/r}^{55}$  y  $\alpha = 0.517$ .

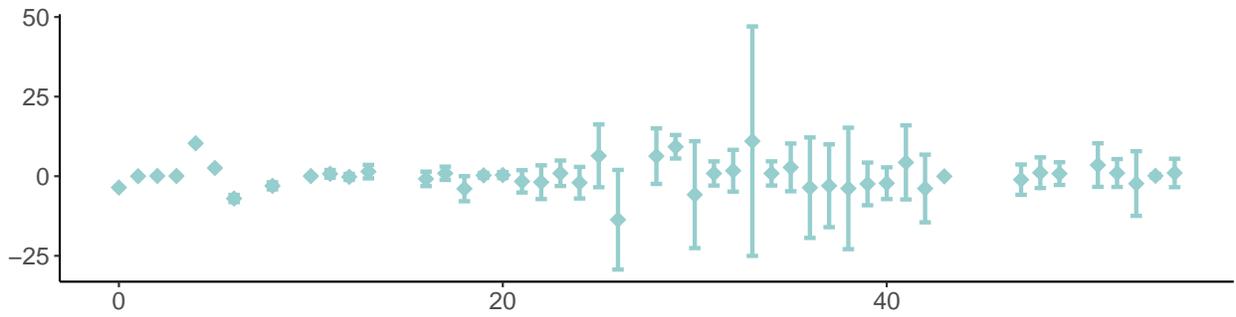
Figura 6.13: Tasas de clasificación errónea (%), validadas y aparentes, asociadas a los modelos de Regresión Multinomial *Elastic Net* ajustados sobre los Componentes Principales obtenidos sin reescalar las variables originales para clasificación por corriente histórica restringida a Música con armonía tonal. Las tasas se varían respecto al número de PCs incluidos en el modelo (para el caso en que se presentan también tasas validadas por grupo) y los parámetros  $\alpha$  y  $\lambda$ . Las líneas punteadas indican tasas globales aparentes.

Es momento de mencionar que la implementación de este método fue hecha mediante el planteamiento para el ajuste del modelo dado por las ecuaciones

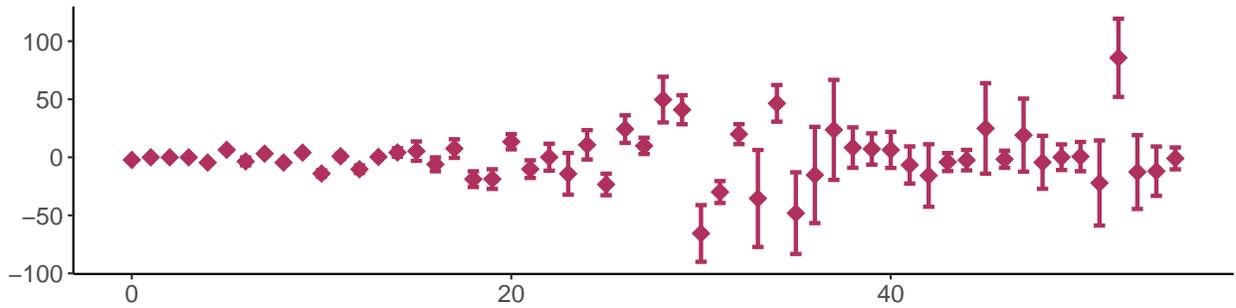
$$\ln \left( \frac{\mathbb{P}(Y = i | X)}{1 - \mathbb{P}(Y = i | X)} \right) = \beta_{0i} + \beta_i^t X;$$

para  $i \in \{1, \dots, k\}$ , donde la variable respuesta toma  $k$  clases. O, equivalentemente,

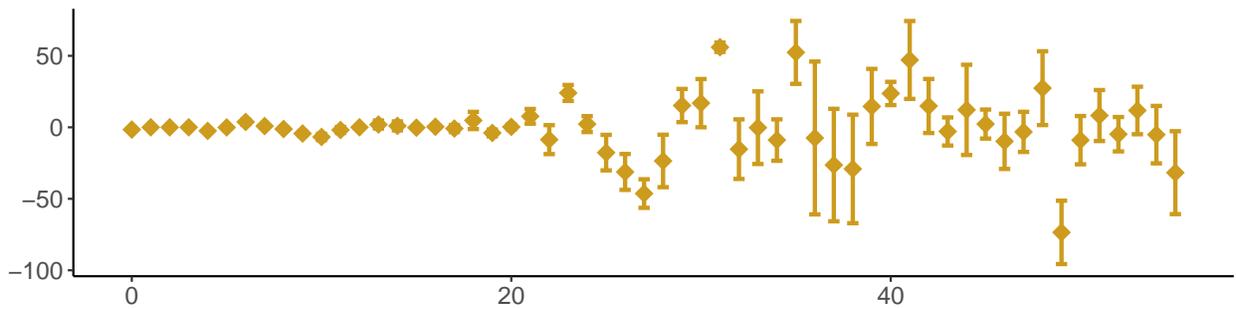
$$\mathbb{P}(Y = i | X) = \frac{\exp\{\beta_{0i} + \beta_i^t X\}}{\sum_{m=1}^k \exp\{\beta_{0m} + \beta_m^t X\}}.$$



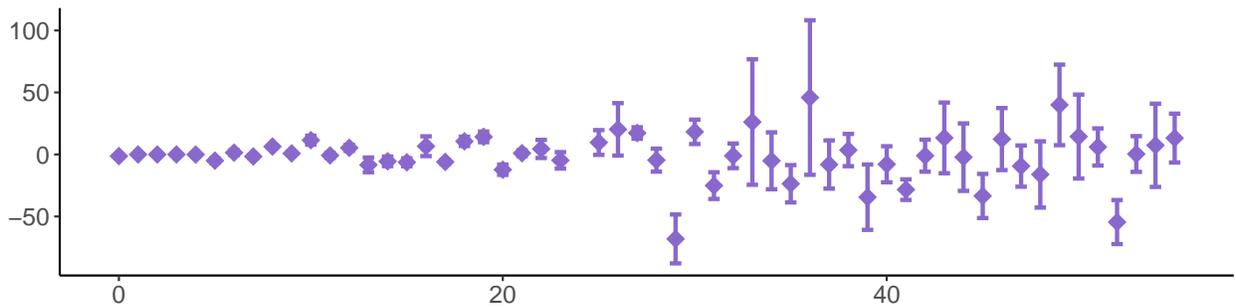
(a) Renacentista.



(b) Barroco.



(c) Clásico.



(d) Romanticismo.

Figura 6.14: Coeficientes de regresión estimados y sus intervalos al 0.95 de confianza por *Cross-Validation* del modelo de Regresión Multinomial *Elastic Net* para la clasificación por corriente histórica restringida a Música con armonía tonal ajustado con los primeros 55 PCs obtenidos sin reescalar las variables originales. El eje x indica el número del Componente Principal, el cero corresponde al intercepto.

Dicho modelo es equivalente al introducido en el Capítulo 4, pues nótese que la diferencia de los coeficientes de regresión derivados del primero con los de la clase  $k$  –para  $i \in \{1, \dots, k-1\}$ – corresponden a los del segundo planteamiento. Sin embargo, claramente con la aplicación de los mismos parámetros de regularización a las funciones de costo derivadas de cada modelo no se resulta en los mismos coeficientes estimados, razón por la cual es razonable aclararlo. Asimismo, por ello se cuenta con las estimaciones de  $\beta_i$  y  $\beta_{0i}$  para  $i \in \{1, \dots, p\}$  y  $j \in \{Renacentista, Barroco, Clásico, Modernismo\}$ . Tales estimaciones, junto con sus estadísticas asociadas son consultables en los cuadros A.15 al A.18 del *Anexo VI* –el cual puede consultarse en [github.com/danielapdv/music-mcs](https://github.com/danielapdv/music-mcs)– y se presentan gráficamente en la Figura 6.14.

### 6.3. Clasificación por autor

La presente sección está dedicada a la clasificación por autor de obras musicales. A diferencia de los apartados anteriores, la variable respuesta  $Y$  es categórica nominal y toma los valores  $\{Bach, Beethoven, Chopin, Handel, Haydn, Mozart, Palestrina, Vivaldi\}$ . Además, la clasificación se realizó con las tres bases de datos contempladas hasta ahora, i.e., 302 predictores, pero con únicamente las 1,016 observaciones, las obras musicales relativas a dichos autores.

Autor	Obras musicales	Porcentaje
Bach	105	10.33
Beethoven	100	9.84
Chopin	114	11.22
Handel	93	9.15
Haydn	100	9.84
Mozart	100	9.84
Palestrina	300	29.53
Vivaldi	104	10.24
Total	1,016	100

Cuadro 6.16: Cantidad de piezas musicales por clase en las bases de datos para clasificación por autor.

Para realizar la clasificación planteada en esta sección fueron aplicados los métodos de Aprendizaje Estadístico: Regresión Multinomial, Regresión Multinomial con regularización (Lasso, Rid-

ge y *Elastic Net*), Árboles de Clasificación (*Single Tree*, *Bagging*, *Boosting* y *Random Forests*), *k*-Vecinos más Cercanos, Análisis de Discriminante (Lineal y Cuadrático) y el clasificador Bayesiano análogo al presentado para la clasificación por corriente musical, pero sobre los factores de consonancia y disonancia estimados por autor que fueron anteriormente presentados en el Cuadro 5.5. Asimismo, la metodología fue llevada a cabo de la misma manera que en las secciones anteriores, esto es, sobre los variables originales y los Componentes Principales derivados del Análisis reescalando las variables y sin reescalarlas<sup>15</sup>. Los modelos representativos de cada bloque metodológico en cuanto a poder predictivo, junto con las tasas de error de clasificación respectivas, se enuncian a continuación.

Método	V. O.	Modelo		TECGV		
		PC <sub>c/r</sub>	PC <sub>s/r</sub>	V. O.	PC <sub>c/r</sub>	PC <sub>s/r</sub>
R. Multinomial	-	68	43	58.13	44.27	<b>32.75</b>
R. M. Lasso	$\lambda = 0.045$	25, $\lambda=0.0004$	42, $\lambda=0.00035$	51.72	43.82	<b>32.95</b>
R. M. Ridge	$\lambda = 0.045$	302, $\lambda=0.04$	45 $\lambda=0.00035$	40.74	42.03	<b>33.65</b>
R. M. <i>Elastic Net</i>	$\alpha=0.5,$ $\lambda=0.09$	26, $\alpha=0.5,$ $\lambda=0.0008$	42, $\alpha=0.5,$ $\lambda=0.0007$	52.11	43.79	<b>32.96</b>
<i>Single Tree</i>	-	302	302	<b>42.32</b>	50.49	51.08
<i>Bagging</i>	-	104	42	<b>35.72</b>	46.71	45.36
<i>Random Forests</i>	$l = 92$	29, $l = 17$	38, $l = 17$	<b>31.78</b>	47.79	41.01
<i>Boosting</i>	-	79	51	<b>39.95</b>	46.4	42.21
<i>k</i> -NN	$k = 160$	$k = 6$	$k = 172$	55.88	<b>50.41</b>	55.78
LDA	-	218	70	43.99	41.91	<b>35.04</b>
QDA	-	6	27	-	46.92	<b>35.99</b>

Cuadro 6.17: Tasas de error de clasificación globales validadas (%) para clasificación por autor con base en las variables originales (V.O.) y los Componentes Principales obtenidos al reescalarlas (PC<sub>c/r</sub>) y sin reescalamiento (PC<sub>s/r</sub>). Las columnas “Modelo” describen los parámetros de ajuste de cada modelo y la cantidad de PCs usados en los casos correspondientes.

Cabe destacar que en las tres tareas de clasificación de este capítulo se ha observado el mismo comportamiento en cuanto a que cada método resulta en mejores tasas de error de clasificación al aplicarse sobre conjuntos de datos específicos: los modelos de Regresión y los Análisis de Discriminante muestran mayor poder predictivo sobre los Componentes Principales obtenidos sin reescalar las variables originales, al reescalarlas mejoran las tasas validadas de acierto de *k*-Vecinos más

<sup>15</sup>Véase Anexo IV. Análisis de Componentes Principales Adicional.

Cercanos y, finalmente, los Árboles de Clasificación son mucho más eficientes sobre las variables originales.

Método	Modelo	TECGA	TECGV	Intervalos de Confianza (95 %)
<b>R. Multinomial</b>	$PC_{s/r}^{43}$	<b>20.96</b>	<b>32.75</b>	
R. M. Lasso	$PC_{s/r}^{42}$ $\lambda=0.00035$	22.05	32.95	
R. M. Ridge	$PC_{s/r}^{45}$ $\lambda=0.00035$	21.36	33.65	
R. M. Elastic Net	$PC_{s/r}^{42}$ , $\alpha=0.5$ $\lambda=0.0007$	22.74	32.96	
Single Tree	V. O.	34.74	42.32	
Bagging	V. O.	36.02	35.72	
<b>Random Forests</b>	V. O., $l = 92$	<b>32.78</b>	<b>31.78</b>	
Boosting	V. O.	<b>7.68</b>	39.95	
k-NN	$PC_{c/r}^{302}$ , $k = 6$	-	50.41	
LDA	$PC_{s/r}^{70}$	27.17	35.04	
QDA	$PC_{s/r}^{27}$	<b>18.11</b>	35.99	
C. Bayesiano	$\hat{\theta}_{\mathcal{L}}, \hat{\theta}_{\mathcal{D}}$	73.13	-	

Cuadro 6.18: Tasas de error de clasificación globales aparentes (TECGA) y validadas (TECGV) correspondientes a clasificación por autor por todos los métodos presentados (%). En la columna “Modelo” se especifican las variables contempladas para el ajuste del mismo, así como los parámetros correspondientes al método utilizado. V. O. hace alusión a las variables originales.

En primer lugar, los dos modelos con TCEG aparentes menores resultaron ser *Boosting* sobre las variables originales y Análisis de Discriminante Cuadrático sobre los primeros 27 PCs obtenidos sin reescalar las variables, con tasas de 7.68% y 18.11%, respectivamente. Sin embargo, al validar las tasas globales asociadas los mismos métodos, aumentan por 32.27% y 17.88%, lo que indica que los modelos ajustados sobre todas las observaciones están saturados y carecen de valor predictivo *real*, pues las tasas respectivas validadas se valoraron en 39.95% y 35.99%.

Por otro lado, los modelos con menores tasas de error de clasificación globales validadas son *Random Forests* sobre las variables originales (31.78%) y Regresión Multinomial con base en

los primeros 43 PCs obtenidos sin reescalamiento en el análisis (32.75%). El modelo asociado a *Random Forests* cuenta con un sesgo por *Cross-Validation* de  $-1.1e-17$  y una desviación estándar de 0.018, por lo que el intervalo al 0.95 de confianza correspondiente a la tasa validada es de [28.3, 35.27]. Asimismo, la tasa global relativa al ajuste de Regresión Multinomial se estima dentro del intervalo [26.03, 39.47], con 0.034 de desviación y  $-1.1e-17$  de sesgo. Nótese que mediante estos dos métodos se logran las menores TCEG validadas, por lo que representan los mejores modelos para clasificación de obras musicales por autor y serán analizados detalladamente en los siguientes apartados.

Método	Bach	Beethoven	Chopin	Handel	Haydn	Mozart	Palestrina	Vivaldi
<b>R. M.</b>	<b>43.32</b>	<b>56.75</b>	<b>22.64</b>	<b>47.59</b>	<b>61.31</b>	<b>59.55</b>	<b>0</b>	<b>28.19</b>
R. M. Lasso	33.87	63.96	20.22	<b>45.13</b>	68.63	63.86	0.63	25.48
R. M. Ridge	35.72	64.36	23.59	45.64	65.01	60.97	2.29	26.02
R. M. El. Net	32.46	63.96	<b>19.38</b>	46.18	66.23	65.6	0.93	26.22
<i>Single Tree</i>	54.29	82	38.6	66.67	78	65	0	40.38
<i>Bagging</i>	41.04	68.99	34.98	61.4	62.89	68.35	0	21.74
<b>Random F.</b>	<b>29.78</b>	<b>64.17</b>	<b>27.08</b>	<b>49.54</b>	<b>60.9</b>	<b>64.65</b>	<b>0</b>	<b>18.66</b>
<i>Boosting</i>	48.49	76.48	28.58	69.69	70.75	75.11	0	30.98
<i>k</i> -NN	48.69	87.14	89.84	86.4	75.89	89.18	0.32	25.95
LDA	35.68	71.68	26.17	61.42	62.77	68.41	0	25.84
QDA	43.57	78.48	25.14	73.73	63.58	<b>56.22</b>	0	<b>18.36</b>
C. Bayesiano	50	85.65	100	91.96	100	100	6.81	100

Cuadro 6.19: Tasas de error de clasificación por grupo validadas (%) correspondientes a clasificación por autor por los métodos y modelos presentados en el cuadro anterior.

En general, los autores que con más acierto logran clasificarse son *Palestrina* y *Vivaldi*, el promedio de la tasa de acierto de clasificación validada sobre todos los métodos es de 99.09% para *Palestrina* y 67.68% para *Vivaldi*; 99.62% y 73.83%, respectivamente, si se excluye del promedio el clasificador Bayesiano sobre los factores de consonancia y disonancia. De hecho, únicamente los modelos de Regresión Multinomial regularizada, *k*-Vecinos más Cercanos y el clasificador Bayesiano resultan en tasas de error de clasificación validadas no nulas para *Palestrina*. Adicionalmente, las mejores tasas de clasificación errónea validadas en el caso de *Vivaldi* son alcanzadas con QDA (18.36%) y *Random Forests* (18.66%).

Por otro lado, los autores que con más dificultad son clasificados exitosamente son Beethoven, Mozart y Haydn, con tasas de acierto de apenas de 28.03%, 30.26% y 30.34% promediadas sobre

todos los modelos mostrados; 29.28 %, 33.01 % y 33.09 %, de manera respectiva si se excluye el clasificador Bayesiano sobre factores de consonancia y disonancia del promedio. Recuérdese de las secciones anteriores que la corriente musical que mayor reto representa en cuanto a su clasificación, seguida del *Modernismo*, es la música del periodo Clásico, corriente musical histórica al cual pertenecen estos tres autores. Las mejores tasas de error de clasificación a las que se llega en su asignación son: 56.75 % para Beethoven mediante Regresión Multinomial, 56.22 % con QDA en el caso de Mozart, y 60.9 % mediante *Random Forests* para Haydn.

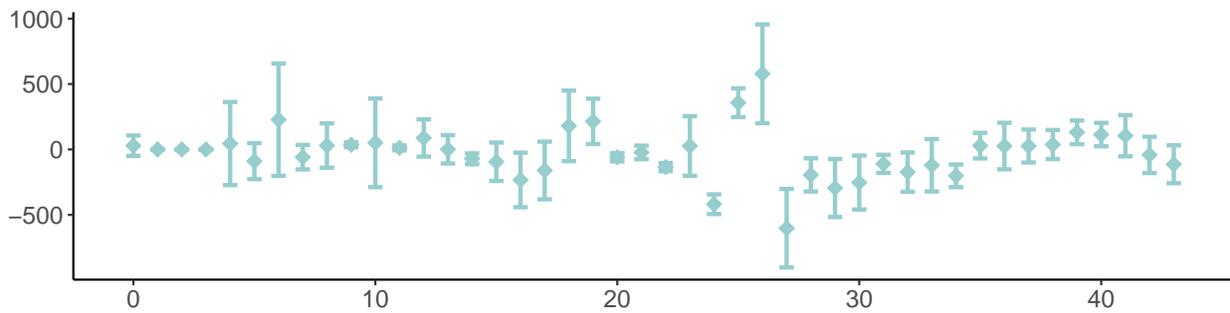
Cabe destacar que el clasificador Bayesiano sobre los factores de consonancia y disonancia resulta en una TEGC aparente sumamente alta, específicamente, de 73.13, la cual es cercana al error esperado de asignar las clases al azar equiprobablemente, 87.5 %. Esto es derivado directamente del hecho de que, en realidad, asigna exitosamente sólo la clase con mayor número de observaciones, *Palestrina*, a la que corresponde casi el 30 % de los datos. En otras palabras, no es un clasificador útil en este caso.

### 6.3.1. Regresión Multinomial

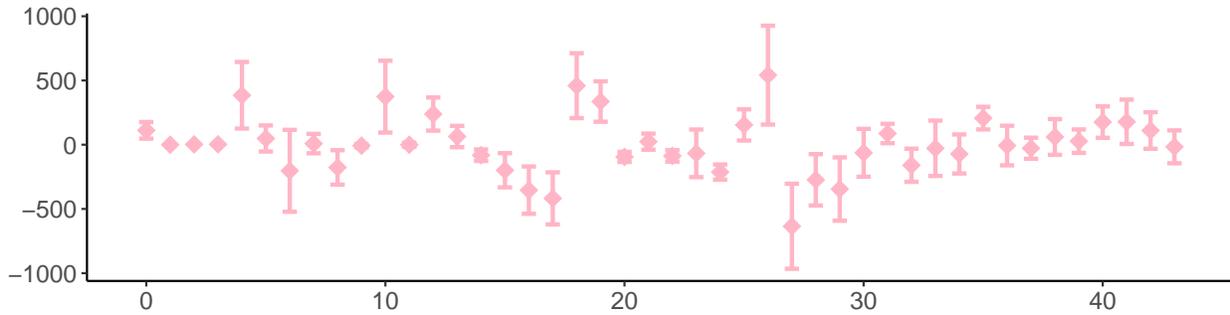
Uno de los dos mejores modelos, en cuanto a su tasa de acierto global, es aquel ajustado mediante Regresión Multinomial sobre los primeros 43 Componentes Principales correspondientes al análisis sin reescalamiento de las variables originales. Dicho modelo, como se muestra a continuación, resulta significativo al hacer Análisis de Devianza; esto es, para el contraste cuya hipótesis nula es que todos los coeficientes asociados a los predictores sean cero contra que al menos uno no lo sea.

Estadística	AIC	BIC	Devianza	Devianza Residual	<i>p-value</i>	TEGC (%)
Aparente	2,935.83	3,152.48	2,847.83	1,169.32	< 2e-16	20.96
Validada	2,441.37	2,658.01	2,353.37	858.84		32.75

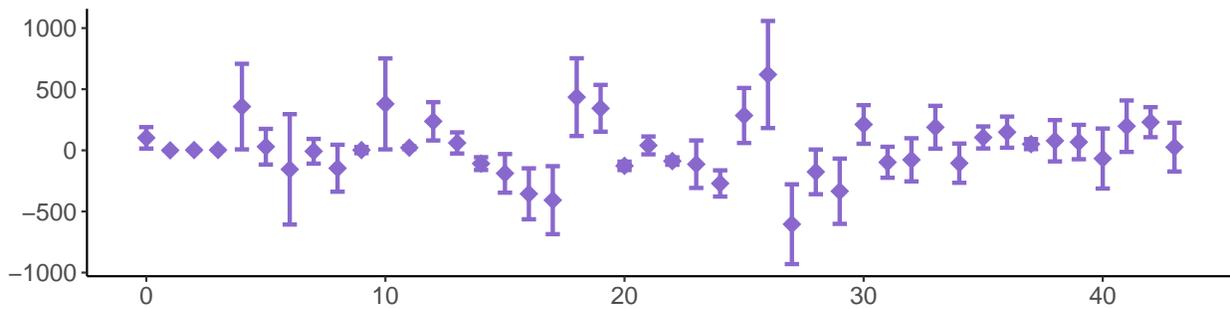
Cuadro 6.20: Criterios de Información de Akaike y Bayesiano, Análisis de Devianza y tasa de error de clasificación global, aparentes y validados, del modelo de Regresión Multinomial para la clasificación por autor ajustado con los primeros 43 PCs obtenidos sin reescalar las variables originales.



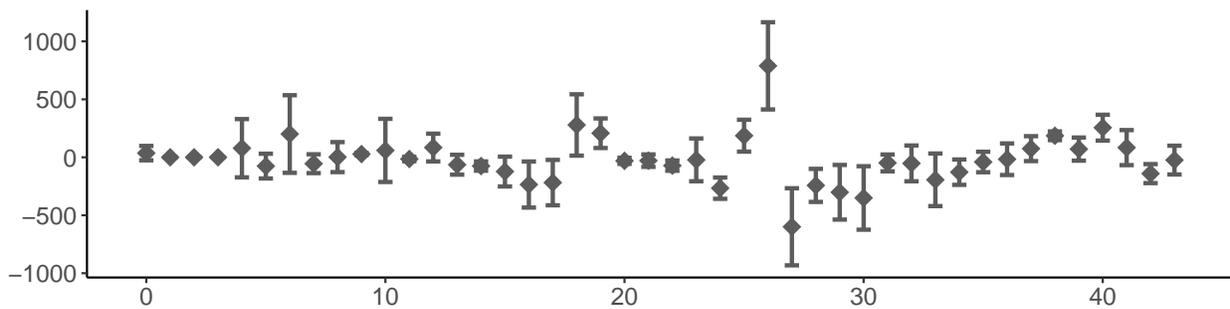
(a) Bach.



(b) Beethoven.

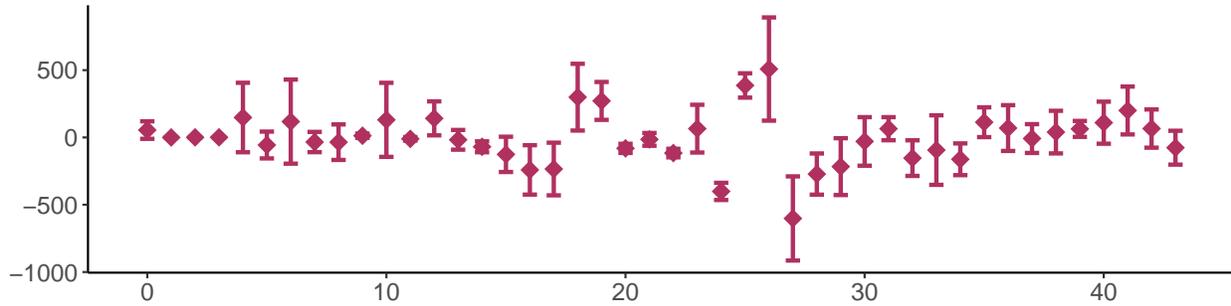


(c) Chopin.

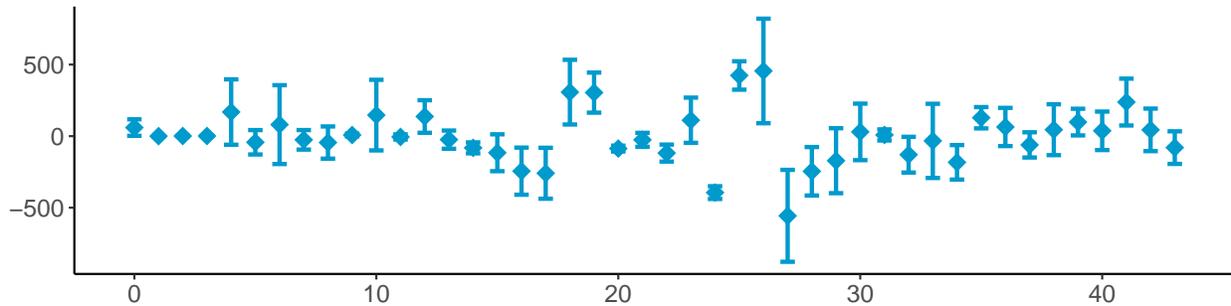


(d) Handel.

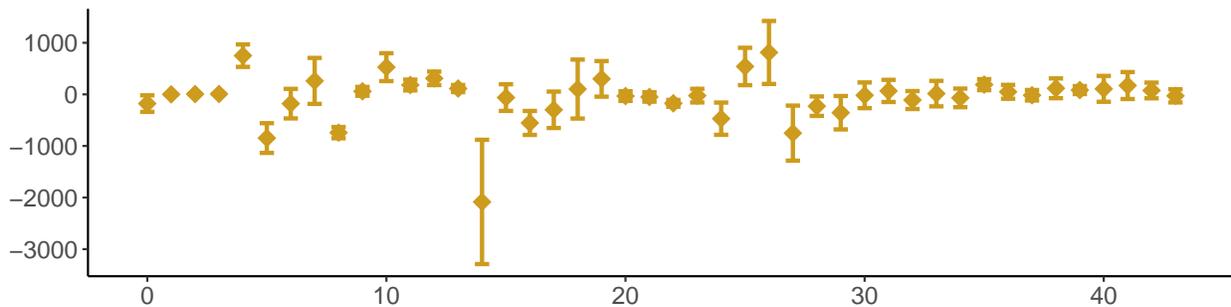
Figura 6.15: Coeficientes de regresión estimados y sus intervalos al 0.95 de confianza por *Cross-Validation* del modelo de Regresión Multinomial para la clasificación por autor ajustado con los primeros 43 PCs obtenidos sin reescalar las variables originales. El eje x indica el número del Componente Principal, el cero corresponde al intercepto.



(e) Haydn.



(f) Mozart.



(g) Palestrina.

Figura 6.15: Coeficientes de regresión estimados y sus intervalos al 0.95 de confianza por *Cross-Validation* del modelo de Regresión Multinomial para la clasificación por autor ajustado con los primeros 43 PCs obtenidos sin reescalar las variables originales. El eje x indica el número del Componente Principal, el cero corresponde al intercepto.

Como se evidencia en los cuadros A.19 al A.25 del *Anexo VI* –disponible en [github.com/danielapdv/music-mcs](https://github.com/danielapdv/music-mcs)–, no todos los coeficientes de regresión estimados son estadísticamente significativos: trece asociados a la clase *Bach*; ocho a *Beethoven*, *Chopin* y *Palestrina*; cinco a *Handel*; diez a *Haydn*; y siete a *Mozart*. Sin embargo, se decidió concluir con esta modelación en función de que se logra una tasa de clasificación global acertada validada de 64.5%.

$Y \setminus \hat{Y}$	Bach	Beethoven	Chopin	Handel	Haydn	Mozart	Palestrina	Vivaldi
Bach	76	5	0	3	6	2	0	13
Beethoven	3	61	8	3	14	9	0	2
Chopin	1	7	101	3	1	1	0	0
Handel	13	0	3	67	0	5	0	5
Haydn	10	9	3	6	51	20	0	1
Mozart	6	7	5	5	21	53	0	3
Palestrina	0	0	0	0	0	0	300	0
Vivaldi	5	2	0	3	0	0	0	94
Aparente	27.62	39	11.4	27.96	49	47	0	9.62
Validada	43.32	56.75	22.64	47.59	61.31	59.55	0	28.19

Cuadro 6.21: Matriz de confusión y tasas de clasificación errónea, aparentes y validadas, por grupo del modelo de Regresión Multinomial para la clasificación por autor ajustado con los primeros 43 PCs obtenidos sin reescalar las variables originales.

En el cuadro anterior puede apreciarse que el clasificador suele confundir a Beethoven con Haydn, a Haydn con Mozart y viceversa. De aquí que resultan en las tasas de error por grupos, tanto aparentes, como validadas, más altas. Adicionalmente, esto concuerda con el hecho de que el Clasicismo, corriente histórica a la que pertenecen dichos autores, es aquella clase más difícil de clasificar en la sección anterior. También cabe destacar que trece obras de Bach las clasifica como de Vivaldi, cinco de Vivaldi, las asigna a Bach y trece de Handel son etiquetadas como obras de Bach (los tres del Barroco). Aunado a lo anterior, tal hecho brinda evidencia de que la clasificación por corriente histórica es más adecuada y, con razón, puesto que en las secciones correspondientes a ello se logran tasas de error de clasificación globales validadas de 26.04% y 17.9% respectivamente, mucho menores que el 32.75% y 31.78% resultantes de esta sección.

### 6.3.2. *Random Forests*

Finalmente, se presentarán características más detalladas del mejor modelo para clasificación por autor del capítulo, que es el modelo de *Random Forests* sobre las variables originales. Primero, cabe hacer notar que para los árboles construidos intrínsecamente se fijó como parámetro una profundidad máxima de 20 bifurcaciones. Asimismo, se llevó a cabo con la construcción de cien árboles, i.e., se construyeron cien árboles simples bajo *Bootstrap*. Adicionalmente, se impuso al algoritmo el criterio de detenimiento: si después de dos rondas de ajuste la tasa de clasificación errónea global no se reduce en 0.001 unidades, la construcción de árboles por *Bootstrap* se detiene.

Recuérdese que *Random Forests* consiste en seleccionar aleatoria y equiprobablemente  $l < p$  predictores que considerar para realizar la bifurcación en cada iteración de la construcción de los árboles simples. Luego, a pesar de que el parámetro  $l$  recomendado en la bibliografía para clasificación es  $\sqrt{p} \approx 17$ , se aplicó el método variando  $l \in \{1, \dots, 302\}$  con el fin de minimizar la TEGC. A continuación se presentan los resultados de aplicar tal ejercicio, en que se concluyó que fijando  $l = 92$  se optimiza el error, con una TEGCV de 31.78%. En la Figura 6.16 no se incluyen las tasas de clasificación errónea globales aparentes puesto que son muy cercanas a las validadas, lo cual implica que justamente la naturaleza del método está evitando el sobreajuste de los modelos; además, se tomó tal medida por fines de presentación.

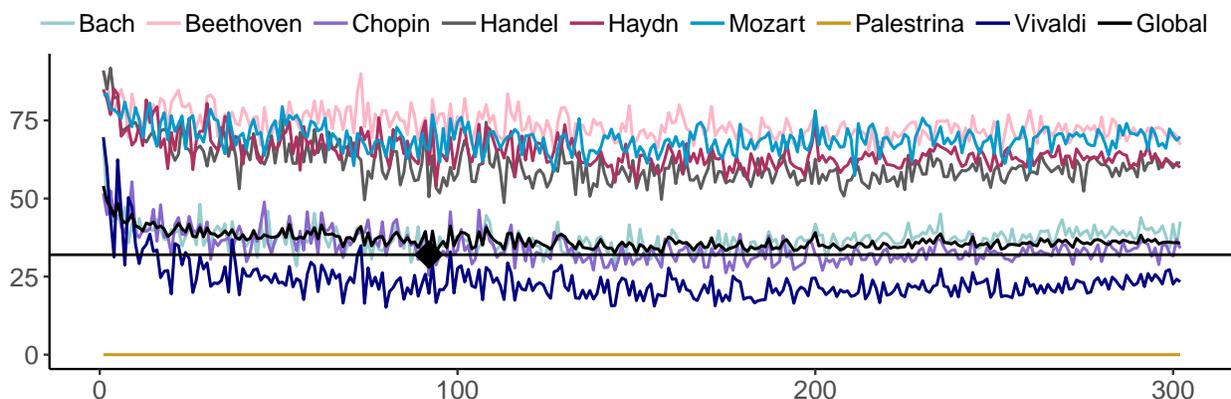


Figura 6.16: Tasas de error de clasificación validadas, global y por grupo, respecto a  $l$  (al número de variables a tomar en cuenta) para el ajuste de Árboles de Clasificación en cada iteración del modelo de *Random Forests* sobre las variables originales para clasificación por autor. El punto y la línea constante indican la tasa global mínima.

Por otro lado, una ventaja que otorgan los Árboles de Clasificación sobre otros métodos es que puede calcularse la importancia relativa de las variables explicativas para predecir la variable respuesta con base en cuánto *mejora* en promedio la función de costo utilizada para la construcción de los árboles al bifurcarlos respecto cada variable. A continuación, se muestra la importancia relativa resultante del modelo en cuestión correspondiente a los diez predictores más influyentes; los cuales, en conjunto, representan el 49.49% de la importancia de las 302 variables predictoras utilizadas para ajustar el modelo. Nótese que la probabilidad a largo plazo de realizar notas blancas, negras, corchea y semicorchea, silencios, y tritonos se encuentran dentro de las diez variables con mayor importancia; mismas que resultaron influyentes para la construcción de árboles simples para clasificación por corriente histórica<sup>16</sup> y que pertenecen también al conjunto de variables que

<sup>16</sup>Véase Figura 6.8.

mayores pesos tienen en los primeros tres Componentes Principales –relativos tanto a las variables reescaladas, como a su contraparte–.

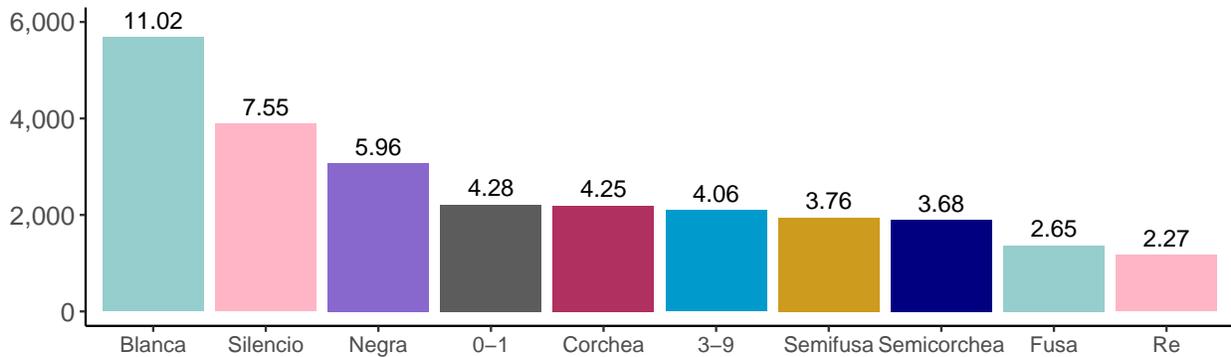


Figura 6.17: Importancia relativa de las variables originales en el modelo de *Random Forests* para clasificación por autor. Se presentan aquellas diez con mayor influencia y se indica en la parte superior de cada barra el porcentaje que implica.

Finalmente, se presenta la matriz de confusión asociada al mejor modelo de la sección, en que se evidencia la dificultad del clasificador de asignar las clases correspondientes a los autores que pertenecen a la misma corriente musical histórica, como se discutió al final del apartado anterior. Con *Random Forests* se logra mejorar la tasa de clasificación errónea validada derivada del modelo de Regresión Multinomial por 0.97%; además, se reducen las tasas asociadas a los grupos: Bach, Haydn, Palestrina y Vivaldi.

$Y \setminus \hat{Y}$	Bach	Beethoven	Chopin	Handel	Haydn	Mozart	Palestrina	Vivaldi
Bach	63	2	3	8	1	5	0	23
Beethoven	4	26	18	5	23	19	0	5
Chopin	2	5	84	2	9	11	0	1
Handel	11	2	4	44	5	18	0	9
Haydn	3	9	15	9	40	18	0	6
Mozart	3	15	4	9	28	37	0	4
Palestrina	0	0	0	0	0	0	300	0
Vivaldi	8	1	0	5	0	1	0	89
Aparente	40	74	26.32	52.69	60	63	0	14.42
Validada	29.78	64.17	27.08	49.54	60.9	64.65	0	18.66

Cuadro 6.22: Matriz de confusión y tasas de clasificación errónea, aparentes y validadas, por grupo del modelo de *Random Forests* para la clasificación por autor ajustado con las variables originales.



## CAPÍTULO 7

# SIMULACIÓN DE OBRAS MUSICALES

El presente capítulo trata sobre simulación de obras musicales conceptualizándolas como Modelos de Markov de orden  $k - 1$  y con base en la paquetería del MIT anteriormente mencionada, `music21`. Es debido hacer notar que la modelación introducida en el Capítulo 5 no es adecuada para esta finalidad debido a que con ella se pretende caracterizar de manera representativa el tempo armónico de una pieza musical de manera generalizada, por lo que se excluyen aspectos fundamentales para la simulación, como por ejemplo, la melodía de manera aislada. Desde luego que es posible llevar a cabo simulación sobre las Cadenas de Markov definidas en tal capítulo, sin embargo, recuérdese que en un principio se transforman las intervenciones de la obra a acordes en su posición básica, por lo tanto, el resultado estaría dado por intervenciones de la misma forma y sería sumamente disonante por lo mismo.

Considérese los procesos que denotan, primero, las intervenciones y, segundo, las duraciones respectivas de las realizaciones de una obra musical en tiempo discreto. Esto, para la melodía y acompañamiento por separado, i.e., se estará considerando cuatro procesos. Aquéllos asociados a las duraciones toman valores en el mismo conjunto que la Cadena de Markov de duraciones especificada en el Capítulo 5 –véase Cuadro 5.2–, excepto por el valor “otras”, en cuyo caso se asignó “negra”. Mientras que las intervenciones toman valores en el conjunto de notas y acordes caracterizados mediante la notación inglesa de las notas y su octava. Por ejemplo, el estado  $A_4$  hace alusión a la nota  $La_4$  y  $(C_4, E_4, G_4)$  al acorde *Do Mayor* en la cuarta octava. Al hecho de que no se contemplen silencios en los procesos asociados a las intervenciones se le dará sentido posteriormente.

Tales procesos se considerarán Markovianos de orden  $k - 1$ , i.e., para  $n \geq k$

$$\mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1}, \dots, X_1 = x_1) = \mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1}, \dots, X_{n-k+1} = x_{n-k+1}).$$

Adicionalmente, se asumirá que son homogéneos, es decir, para todo  $n \geq k$

$$\mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1}, \dots, X_{n-k+1} = x_{n-k+1}) = \mathbb{P}(X_k = x_k \mid X_{k-1} = x_{k-1}, \dots, X_1 = x_1);$$

donde  $X_1$  es la primera realización el proceso. Nótese que tal planteamiento es análogo a una Cadena de Markov homogénea para  $k = 2$ .

El primer ejercicio de simulación consiste en lo siguiente, dada una obra musical *base* sobre la cual se simulará, para cada uno de los cuatro procesos definidos anteriormente:

1. Obtener equiprobablemente las primeras  $k - 1$  realizaciones del proceso sobre el espacio de estados  $E$ , cuya cardinalidad se denotará por  $m$ .
2. Para  $i \in \{k, \dots, n\}$ , donde  $n$  corresponde al número de saltos a simular,

a) Estimar por Máxima Verosimilitud el vector de probabilidades de transición

$$\left( P(x_{i-k+1}, \dots, x_{i-1}, s_1), \dots, P(x_{i-k+1}, \dots, x_{i-1}, s_m) \right);$$

donde

$$P(x_{i-k+1}, \dots, x_{i-1}, x_i) = \mathbb{P}(X_k = x_i \mid X_{k-1} = x_{i-1}, \dots, X_1 = x_{i-k+1}),$$

y  $s_j \in E$ ,  $j \in \{1, \dots, m\}$ . El estimador es análogo al presentado en la Sección 2.3, en que se obtiene el Estimador Máximo Verosímil de la matriz de transición para una Cadena de Markov Homogénea, extendiendo el procedimiento al orden del modelo en  $k - 1$ . Es decir, la estimación de dicho vector de probabilidades está dado por

$$\left( \frac{n(x_{i-k+1}, \dots, x_{i-1}, s_1)}{\sum_{j=1}^m n(x_{i-k+1}, \dots, x_{i-1}, s_j)}, \dots, \frac{n(x_{i-k+1}, \dots, x_{i-1}, s_m)}{\sum_{j=1}^m n(x_{i-k+1}, \dots, x_{i-1}, s_j)} \right);$$

donde  $n(x_{i-k+1}, \dots, x_{i-1}, x_i)$  corresponde al número de veces que en la obra musical base se dio la realización  $x_{i-k+1}, \dots, x_{i-1}, x_i$  del proceso correspondiente.

- b) Generar el siguiente paso del proceso,  $X_i$ , con base en el vector de probabilidades, i.e., simular la variable aleatoria que toma el valor  $s_j$  con probabilidad  $\hat{P}(x_{i-k+1}, \dots, x_{i-1}, s_j)$ . Esto es, generar un número aleatorio  $U$  de distribución uniforme continua  $(0, 1)$  y asignar al salto del proceso el valor  $s_j$  tal que

$$\sum_{k=1}^{j-1} \hat{P}(x_{i-k+1}, \dots, x_{i-1}, s_k) \leq U \leq \sum_{k=1}^j \hat{P}(x_{i-k+1}, \dots, x_{i-1}, s_k);$$

usando la convención  $\sum_{k=1}^0 \hat{P}(x_{i-k+1}, \dots, x_{i-1}, s_k) = 0$ .

3. Finalmente, al contar con las simulaciones de los cuatro procesos, construir la obra musical con base en las intervenciones y duraciones asociadas a la melodía y acompañamiento, respectivamente.

Nótese que, a diferencia de la modelación propuesta en el Capítulo 5, en este caso no se incluyen suavizamientos en la estimación de las probabilidades de transición debido a que se busca que la simulación de la pieza sea *lo más fiel posible* a la original, en cuanto a su mera sensación auditiva. Adicionalmente, por la manera en que fueron definidos los procesos asociados a las intervenciones, dada una secuencia de tamaño  $k - 1$ , la realización del siguiente salto está restringida a un número *pequeño* de estados. De manera que, al aumentar un sesgo a las entradas de los vectores de probabilidad, aunque fuese cercano a cero, implicó resultados poco deseables en el sentido de que, justo en los casos mencionados, las realizaciones de los procesos fueron generadas casi equiprobable e independientemente de las secuencias anteriores, lo cual claramente no es la intención de la tarea de simulación musical<sup>1</sup>.

El archivo *Joplin (ind).mp3* consultable en [github.com/danielapdv/music-mcs](https://github.com/danielapdv/music-mcs) corresponde a la simulación musical bajo este planteamiento tomando como base la obra *Maple Leaf Rag* de Scott Joplin, con  $k = 2$ . En él se hace evidente la necesidad de considerar no independientes la melodía y el acompañamiento, lo cual es un supuesto intrínseco de la modelación propuesta en este apartado, y lo cual tiene sentido, naturalmente.

Ahora bien, la intuición dicta que al aumentar el orden del modelo se obtendrían *mejores* simulaciones. Sin embargo, desde la utilización de  $k = 3$  y derivado de la forma de los procesos

---

<sup>1</sup>Sin mencionar que suena realmente *mal*.

asociados a las intervenciones, resulta que las simulaciones llegan con rapidez a sus estados absorbentes. Tal hecho se puede apreciar en la siguiente figura, en que se muestran los primeros dos compases de la simulación sobre la misma obra musical que el ejemplo anterior<sup>2</sup>, con  $k = 3$ . De aquí se concluye que la utilización de órdenes mayores a 2 para la simulación bajo este planteamiento es impráctica. Y, de hecho, es razón por la cual se excluyen los silencios del espacio de estados de los procesos asociados a las intervenciones: al modelar las obras mediante un Modelo de Markov de orden  $k = 2$  se tiene la equivalencia con Cadenas de Markov y, entonces, si la realización anterior del proceso fuese un silencio, la siguiente estaría dada de manera completamente independiente de la nota o acorde anterior, lo cual no es adecuado. Adicionalmente, por la misma razón enunciada al inicio, no es viable la idea de considerar un proceso que incluya las realizaciones de las intervenciones conjuntamente con sus duraciones, pues la problemática aumenta: los procesos llegan a sus estados absorbentes casi desde el inicio, aún con  $k = 2$ ; o bien, se replica casi idénticamente la pieza musical en la que está basada la simulación.

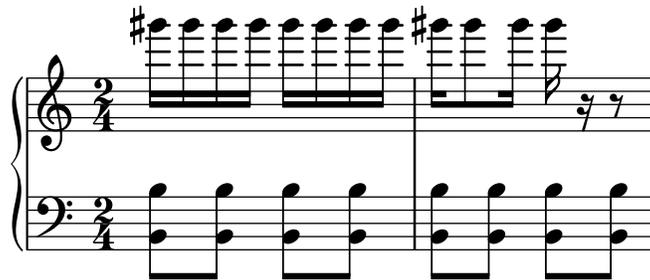


Figura 7.1: Primeros dos compases de la simulación con  $k = 3$  y considerando melodía y acompañamiento independientes sobre *Maple Leaf Rag* de Scott Joplin.

Dicho lo anterior, en lo que queda espacio para mejorar la simulación es en la independencia de la melodía y el acompañamiento<sup>3</sup>. Esto fue llevado a cabo considerando una modelación análoga a introducida anteriormente, con la diferencia de que en lugar de obtener y simular las intervenciones y sus duraciones para melodía y acompañamiento de manera separada, previamente estos dos fueron *concatenados*.

De tal manera, el ejercicio de simulación se simplifica a generar el proceso asociado a las intervenciones y el proceso correspondiente a sus duraciones, implicando aún independencia entre

<sup>2</sup>Se muestran únicamente los dos primeros en medida de que se repiten a lo largo de toda la simulación, sólo con duraciones distintas.

<sup>3</sup>Así debió ser desde un inicio; sin embargo, era de interés visualizar si su adherencia quedaría impresa intrínsecamente en las transiciones de la melodía y el acompañamiento de las obras.

las duraciones y las cualidades de las intervenciones a lo largo de la obra musical. Nótese que la problemática expuesta anteriormente se hereda al presente planteamiento, por lo que es inminente la simulación de las duraciones e intervenciones de forma independiente y con  $k = 2$ . No obstante, se obtuvieron simulaciones *aceptables* en el sentido de que parecería que se sigue cierta tendencia armónica, como puede escucharse en los ejemplos simulados con base en las siguientes obras y en los archivos indicados a continuación siguiendo la liga [github.com/danielapdv/music-mcs](https://github.com/danielapdv/music-mcs).

Autor	Obra musical base Título	Archivo con la simulación
Bach	<i>Chorale BWV 1.6</i>	<i>Bach.mp3</i>
Beethoven	<i>Cuarteto de Cuerdas Opus 18 no. 3</i>	<i>Beethoven.mp3</i>
Joplin	<i>Maple Leaf Rag</i>	<i>Joplin.mp3</i>
Mozart	<i>Cuarteto de Cuerdas No.2 en Re Mayor K.155 Mov. 1</i>	<i>Mozart.mp3</i>
Palestrina	<i>Agnus I 68</i>	<i>Palestrina.mp3</i>
Vivaldi	<i>Las Cuatro Estaciones 8 I I en Mi Mayor</i>	<i>Vivaldi.mp3</i>

Cuadro 7.1: Ejemplos de simulación musical llevados a cabo: su obra musical base y el nombre del archivo para su consulta en [github.com/danielapdv/music-mcs](https://github.com/danielapdv/music-mcs).

Finalmente, es momento de aclarar que la simulación llevada a cabo está restringida a una pieza musical como base. Desde luego que sería tentador aplicarla sobre un autor o, incluso, corriente musical, adecuando las estimaciones de las probabilidades de transición como fue llevado a cabo para encontrar las probabilidades límite asociadas a cada uno de ellos en el Capítulo 5. Sin embargo, lo generado mediante tal ejercicio no seguiría una estructura musical establecida, puesto que al simular sobre una obra solamente, el proceso generado asociado a las intervenciones toma de ella su tonalidad, signo de compás y uso de las progresiones armónicas, por mencionar algunos conceptos. De manera que extender la estimación de las transiciones a más piezas musicales violaría su fundamento teórico musical.



## CONCLUSIONES

La investigación sobre el contexto musical teórico llevó a conceptualizar a las obras musicales como sistemas dinámicos no deterministas que están completamente caracterizados por la variabilidad de su tensión y reposo, disonancia y consonancia, mediante los cuales el autor busca causar las sensaciones que les dan esencia. De tal manera, la tarea llevada a cabo consistió en modelar dichas fluctuaciones y su tendencia a largo plazo, que, de acuerdo con la Teoría Musical, son completamente intencionadas por parte del autor.

Un primer acercamiento fue hecho con la modelación mediante Cadenas de Markov, caracterizando las piezas musicales mediante las raíces de sus intervenciones, su Clase Forte y sus duraciones. Con ello, se pretendió cuantificar el tempo armónico de la obra, mismo objetivo que puede decirse logrado, puesto que los resultados arrojan que tal conceptualización refleja el uso de las progresiones armónicas a largo plazo, con la estimación de las probabilidades límite asociadas a cada cadena. No obstante, la propuesta con base en Cadenas de Markov incluye las notas de paso en los acordes sobre los que se mueven los procesos, lo cual no es adecuado estrictamente hablando, pues en teoría las notas de paso no alteran los acordes a los que se superponen; su fin es meramente ornamental, no estructural. Adicionalmente, no es verosímil que la intervención en una pieza musical dependa únicamente de su inmediata anterior, por lo que tomó inminencia la generalización de la modelación en cuestión.

Posteriormente, como vía para establecer una modelación generalizada del primer enfoque tomado, se hizo uso de la teoría de Campos Aleatorios. La propiedad Markoviana se consideró indispensable para la conceptualización de una obra, al considerarla secuencial; razón por la cual, en primer lugar, se consideró a una pieza musical como un Modelo de Markov de orden  $k - 1$ , con lo

que se relaja el supuesto de la propiedad Markoviana para Cadenas de Markov. Ahora bien, se procedió a modelarlas como Campos Aleatorios dinámicos, de Gibbs y de Markov, en medida de que es tanto intuitivo, como musicalmente adecuado, afirmar que las interacciones de las intervenciones de una pieza derivadas de su realización consecutiva generan cierto flujo, energía, en términos de la consonancia y disonancia causadas<sup>1</sup>.

De tal manera, se construyó un Campo Aleatorio dinámico cuya topología subyace en los vectores de clase de intervalo reducidos asociados a las mismas intervenciones de la obra musical ordenados de manera secuencial, dependientes de los  $k - 1$  anteriores y cuya energía está determinada por la consonancia y disonancia resultantes de sus sucesiones, las cuales fueron cuantificadas mediante dos herramientas musicológicas propuestas: medidas de consonancia y disonancia para una sucesión de acordes. Cabe destacar que, al analizarlas, se concluye que éstas son sensibles, tanto a la resolución de la progresión, ya sea en tensión o cadencia, como a la ganancia o pérdida de disonancia y consonancia en los acordes anteriores a la resolución. Al definir la energía del campo de manera paramétrica, las piezas musicales se caracterizan meramente por dichos parámetros, pues determinan completamente la medida de Gibbs del campo y, además, fue posible darles interpretación: factores de disonancia y consonancia, su noción para la concepción de la pieza musical, la fuerza de la tendencia al uso de cada una.

Es posible aplicar la modelación a piezas musicales de manera aislada, así como extenderla hasta autores y corrientes históricas musicales completos. Cabe destacar que al estimar los conceptos relativos a ambos planteamientos sobre una base de datos, se obtuvieron resultados completamente congruentes con los estudios musicológicos, por mencionar algunos:

- Se tiende a la consonancia sobre la disonancia pese a la concepción que se les de históricamente. Esto es reflejado en el hecho de que la diferencia mínima observada entre los factores de consonancia y disonancia estimados para las corrientes musicales a lo largo de la historia corresponde a 10.71 unidades. De hecho, tal diferencia mínima es la asociada al Modernismo, siendo la noción de la consonancia 1.86 veces mayor que la asociada a su contraparte.
- A lo largo de la historia, las raíces fundamentales predilectas de los acordes utilizados resultan ser *Re* y *Sol*, con probabilidades de ocurrencia a largo plazo de entre 0.04 y 0.2 a lo largo de las corrientes históricas; seguidas por *Do* y *La*, con probabilidades entre 0.05 y 0.1.

---

<sup>1</sup>Y, con ello, el peso de las notas ornamentales en la modelación disminuye.

- Al Renacimiento corresponden composiciones con estructura más simple en cuanto al uso mayormente de triadas mayores y menores, cuya probabilidad de uso a largo plazo es de 0.2. Asimismo, para la utilización de notas fundamentales meramente naturales se registran probabilidades valuadas entre 0.1 y 0.2, contando las notas no naturales con probabilidades menores a 0.01. También se evidencia el uso de sonidos de mayor duración, siendo las notas blancas y negras las más utilizadas –0.5 y 0.4 de probabilidad de ocurrencia–; uso casi nulo de silencios; y tendencia a la consonancia al triple que en el resto de los periodos analizados.
- A partir del Barroco, se hace común el uso de las notas individualmente con probabilidad entre 0.05 y 0.08; así como las intervenciones de duraciones menores, pasando las corcheas a ser las notas preferidas con probabilidades de ocurrencia a largo plazo entre 0.4 y 0.5, seguidas por semicorcheas y fusas, con probabilidad de 0.2. También se evidencia la diversificación del uso de escalas ajenas a las naturales, aunque aún con preferencia a las notas naturales como raíz de los acordes en medida de que se uniformizan las probabilidades asociadas a las raíces, pero aquéllas correspondientes a las notas naturales son mayores a su contraparte por magnitudes entre 0.01 y 0.05. Adicionalmente, Vivaldi y Bach, pertenecientes al Barroco, son aquellos autores con mayor noción de consonancia después de Palestrina<sup>2</sup>.
- En el periodo Clásico se observa la segunda mayor noción de consonancia, justo detrás del Renacimiento, lo cual hace sentido puesto que en dicho periodo el paradigma fue la exaltación de la belleza. También se evidencia la pluralidad que la Música comenzaba a tomar, en medida de que Beethoven resultó ser aquel con menor noción de disonancia de todos los autores analizados, valuado en 3.15. Mientras que Mozart es el único cuyo factor de consonancia es menor que el de disonancia –por 1.69 unidades–, siendo éste el mayor de los registrados con un valor de 30.49.
- A partir del Romanticismo aumenta en el uso de tetracordes y, en general, sonidos simultáneos de más de seis notas, junto con el uso prácticamente equiprobable a largo plazo de notas fundamentales naturales y no naturales –con probabilidad entre 0.04 y 0.05–. Esto concuerda con la negación de los valores de la estética del Clasicismo que caracteriza dicha corriente.
- Lo anterior es concretado en el Modernismo, con la negación y alteración de la armonía tonal de parte de algunos autores, periodo en el que se muestra aumento de la utilización de acordes de hasta doce notas –aunque aún con probabilidades de ocurrencia a largo plazo casi nulas, por su disonancia–, intervenciones de duraciones a la par con las *tradicionales* y la mayor

---

<sup>2</sup>Representante del Renacimiento en la base de datos.

incurrencia en la disonancia de los periodos posteriores al Renacimiento, correspondiendo la estimación del factor de disonancia modernista a 12.39.

Tras la estimación de la modelación propuesta sobre una base de datos, se prosiguió a la clasificación por corriente histórica y por autor con tales estimaciones como predictores. Con el mejor modelo para clasificación por corriente musical histórica se tiene una tasa de clasificación acertada global validada de 73.96 %. Mediante todos los métodos de Aprendizaje Estadístico aplicados para la clasificación por corriente, se observó clasificación exitosa de la clase Renacentista en 100 %, por lo general. Así como gran dificultad para asignar con veracidad el Modernismo, con tasas de acierto validadas menores que 50 % en todos los casos, valuándose en 45.93 % para el modelo de Regresión Multinomial. Mostrando así que la modelación propuesta parece ser sensible a la utilización de la Teoría de la Armonía, al menos en su uso antes de su consumación.

Tal hecho motivó a la clasificación por corriente musical histórica restringida a la Música con armonía tonal, en medida de que en la modelación pudiese faltar algún factor que permita distinguir si en una obra se desarrolla armonía tonal o no, como ocurre en algunas obras del Modernismo, por ejemplo, en la Música Atonal<sup>3</sup>. Mediante dicho ejercicio de clasificación, se alcanzó una tasa de acierto global validada de 82.1 % con el modelo de Regresión Multinomial regularizada mediante *Elastic Net*, ajustado sobre los primeros 55 Componentes Principales obtenidos sin reescalamiento en las variables originales y con los parámetros  $\alpha = 0.517$  y  $\lambda = 0.0007$ . La tasa de acierto resultante es mayor por 8.14 % que la obtenida considerando al Modernismo en los datos.

Finalmente, para la clasificación por autor, el mejor modelo obtenido está dado por *Random Forests* aplicado a las variables originales y con parámetro  $l = 92$ , mediante el cual se logra una tasa de clasificación acertada global validada de 68.22 %. Cabe destacar que se aplicaron trece métodos de Aprendizaje Estadístico sobre las variables originales y sus Componentes Principales obtenidos al reescalar tales variables y sin rescalarlas; lo que muestra que la clasificación sobre obras musicales no es tarea fácil. Sin embargo, se sostiene que las tasas de acierto resultantes son competitivas al tratarse de un problema de reconocimiento de patrones; sobre todo aquéllas asociadas a la clasificación por corriente, que de hecho es generalización de la clasificación por autor.

---

<sup>3</sup>Recuérdese que para poder aplicar la modelación por Campos Aleatorios, se tuvo que restringir a los vectores de clase de intervalo reducidos debido al alto costo computacional derivado de la modelación alternativa, que de hecho no es nada factible. Podría ser justamente esa especificación entre los tipos de consonancia y disonancia los detalles que faltan para permitir a los clasificadores dilucidar entre la Música compuesta con base en la armonía tonal y sus variaciones postonales.

Clasificación	Método	Modelo	TCAGA	TCAGV
Corriente histórica	Regresión Multinomial	Primeros 41 PCs obtenidos sin reescalamiento	76.58	73.96
Corriente histórica restringida a Música no Postonal	Regresión Multinomial <i>Elastic Net</i>	Primeros 55 PCs obtenidos sin reescalamiento, $\alpha = 0.517$ y $\lambda = 0.0007$	85.73	82.1
Autor	<i>Random Forests</i>	Variables originales y $l = 92$	67.22	68.22

Cuadro C.1: Tasas de clasificación acertada globales, aparentes (TCAGA) y validadas (TCAGV), de los mejores modelos para la clasificación por corriente musical histórica, por corriente restringida a Música con armonía tonal y por autor. Tasas en porcentajes.

Asimismo, en general, se observó que los modelos ajustados con los Componentes Principales obtenidos al reescalar las variables originales necesitaron centenas de ellos para minimizar las tasas de error de clasificación validadas; mientras que para los obtenidos sin reescalamiento, no fue necesario el uso de más de sesenta. Este comportamiento es natural, dado que con los primeros dos Componentes Principales derivados de no reescalar las variables originales se acumula prácticamente el 100 % de la varianza explicada, lo que es debido a que los factores de consonancia y disonancia y la evaluación de la función de log-Verosimilitud presentan magnitudes mucho mayores, como es discutido en el *Anexo IV*. No obstante, el hecho de que mediante ellos se logren tasas de error validadas menores puede implicar que darle tal peso *extra* a las variables asociadas a la modelación por Campos Aleatorios es adecuado.

Finalmente, se llevó a cabo simulación musical con base en Cadenas de Markov asociadas a las intervenciones y duraciones a lo largo de una pieza musical base, respectivamente para la melodía y acompañamiento. El procedimiento que proporcionó mejores resultados, como era de esperarse, fue considerar la melodía y acompañamiento no independientes. Cabe hacer notar que las simulaciones generadas parecerían respetar la armonía y esencia de la obra musical original; sin embargo, ésta es justamente su limitación, la necesidad de *aprender* de otra.

La presente investigación muestra que es posible construir acercamientos hacia la modelación de la Música mediante enfoques probabilísticos con resultados *acceptables*. No obstante, es claro que aún queda un largo camino que recorrer para entender el genio musical, si es que fuese verdaderamente posible descriptarlo objetivamente.



## ANEXO I. CLASIFICACIÓN DE ALLEN FORTE

Allen Forte, elaboró una clasificación de conjuntos tonales en su forma básica, herramienta que se utiliza en la presente investigación. La Clase Forte de un conjunto tonal está dada por dos números enteros separados por un guión. El primero corresponde la cantidad de notas que conforman el acorde asociado a tal conjunto tonal [21, p. 180]. Mientras que el segundo es un índice que denota la clase de conjunto de acuerdo a los intervalos entre los tonos que la conforman, y al cual se antepone una *Z* en los casos en que tal clase de conjunto es *Z*-relativa a alguna otra.

A continuación se presentan las Clases Forte observadas en la base de datos<sup>4</sup>, i.e., el espacio de estados de la Cadena de Markov propuesta<sup>5</sup> denotada por  $\{C_n\}_{n \in \mathbb{N}}$ . Cabe destacar que a la lista fueron añadidas las clases 0-1, 1-1, 2-1, 2-2, 2-3, 2-4, 2-5 y 2-6 para denotar respectivamente los silencios, notas y dos sonidos simultáneos con intervalos de longitud 1 (11), 2 (10), 3 (9), 4 (8), 5 (7) y 6. La lista completa de clases puede ser consultada directamente en [11, p. 179] o en [21, p. 315].

Clase Forte	Forma básica	Clase Forte	Forma básica	Clase Forte	Forma básica
0-1	[]	2-3	[0 3]	3-1	[0 1 2]
1-1	[0]	2-4	[0 4]	3-2	[0 1 3]
2-1	[0 1]	2-5	[0 5]	3-2B	[0 1 3]
2-2	[0 2]	2-6	[0 6]	3-3	[0 1 4]

Cuadro A.1: Clases Forte y sus conjuntos tonales en forma básica asociados observados en la base de datos (espacio de estados de la Cadena de Markov  $\{C_n\}_{n \in \mathbb{N}}$ ).

<sup>4</sup>Véase *Anexo V*, disponible en [github.com/danielapdv/music-mcs](https://github.com/danielapdv/music-mcs).

<sup>5</sup>Véase Capítulo 5, Sección 1.

Clase Forte	Forma básica	Clase Forte	Forma básica	Clase Forte	Forma básica
3-3B	[0 1 4]	4-18B	[0 1 4 7]	5-16B	[0 1 3 4 7]
3-4	[0 1 5]	4-19	[0 1 4 8]	5-19	[0 1 3 6 7]
3-4B	[0 1 5]	4-19B	[0 1 4 8]	5-19B	[0 1 3 6 7]
3-5	[0 1 6]	4-20	[0 1 5 8]	5-20	[0 1 3 7 8]
3-5B	[0 1 6]	4-21	[0 2 4 6]	5-20B	[0 1 3 7 8]
3-6	[0 2 4]	4-22	[0 2 4 7]	5-21	[0 1 4 5 8]
3-7	[0 2 5]	4-22B	[0 2 4 7]	5-21B	[0 1 4 5 8]
3-7B	[0 2 5]	4-23	[0 2 5 7]	5-22	[0 1 4 7 8]
3-8	[0 2 6]	4-24	[0 2 4 8]	5-23	[0 2 3 5 7]
3-8B	[0 2 6]	4-25	[0 2 6 8]	5-23B	[0 2 3 5 7]
3-9	[0 2 7]	4-26	[0 3 5 8]	5-24	[0 1 3 5 7]
3-10	[0 3 6]	4-27	[0 2 5 8]	5-24B	[0 1 3 5 7]
3-11	[0 3 7]	4-27B	[0 2 5 8]	5-25	[0 2 3 5 8]
3-11B	[0 3 7]	4-28	[0 3 6 9]	5-26	[0 2 4 5 8]
3-12	[0 4 8]	5-1	[0 1 2 3 4]	5-26B	[0 2 4 5 8]
4-1	[0 1 2 3]	5-2	[0 1 2 3 5]	5-27	[0 1 3 5 8]
4-2	[0 1 2 4]	5-2B	[0 1 2 3 5]	5-27B	[0 1 3 5 8]
4-2B	[0 1 2 4]	5-3	[0 1 2 4 5]	5-28	[0 2 3 6 8]
4-3	[0 1 3 4]	5-3B	[0 1 2 4 5]	5-28B	[0 2 3 6 8]
4-4	[0 1 2 5]	5-4	[0 1 2 3 6]	5-29	[0 1 3 6 8]
4-4B	[0 1 2 5]	5-4B	[0 1 2 3 6]	5-29B	[0 1 3 6 8]
4-5	[0 1 2 6]	5-5	[0 1 2 3 7]	5-30	[0 1 4 6 8]
4-5B	[0 1 2 6]	5-5B	[0 1 2 3 7]	5-30B	[0 1 4 6 8]
4-6	[0 1 2 7]	5-6	[0 1 2 5 6]	5-31	[0 1 3 6 9]
4-7	[0 1 4 5]	5-6B	[0 1 2 5 6]	5-31B	[0 1 3 6 9]
4-8	[0 1 5 6]	5-7	[0 1 2 6 7]	5-32	[0 1 4 6 9]
4-9	[0 1 6 7]	5-7B	[0 1 2 6 7]	5-32B	[0 1 4 6 9]
4-10	[0 2 3 5]	5-8	[0 2 3 4 6]	5-33	[0 2 4 6 8]
4-11	[0 1 3 5]	5-9	[0 1 2 4 6]	5-34	[0 2 4 6 9]
4-11B	[0 1 3 5]	5-9B	[0 1 2 4 6]	5-35	[0 2 4 7 9]
4-12	[0 2 3 6]	5-10	[0 1 3 4 6]	6-1	[0 1 2 3 4 5]
4-12B	[0 2 3 6]	5-10B	[0 1 3 4 6]	6-2	[0 1 2 3 4 6]
4-13	[0 1 3 6]	5-11	[0 2 3 4 7]	6-2B	[0 1 2 3 4 6]
4-13B	[0 1 3 6]	5-11B	[0 2 3 4 7]	6-5	[0 1 2 3 6 7]
4-14	[0 2 3 7]	5-13	[0 1 2 4 8]	6-5B	[0 1 2 3 6 7]
4-14B	[0 2 3 7]	5-13B	[0 1 2 4 8]	6-7	[0 1 2 6 7 8]
4-16	[0 1 5 7]	5-14	[0 1 2 5 7]	6-8	[0 2 3 4 5 7]
4-16B	[0 1 5 7]	5-14B	[0 1 2 5 7]	6-9	[0 1 2 3 5 7]
4-17	[0 3 4 7]	5-15	[0 1 2 6 8]	6-9B	[0 1 2 3 5 7]
4-18	[0 1 4 7]	5-16	[0 1 3 4 7]	6-15	[0 1 2 4 5 8]

Cuadro A.1: Clases Forte y sus conjuntos tonales en forma básica asociados observados en la base de datos (espacio de estados de la Cadena de Markov  $\{C_n\}_{n \in \mathbb{N}}$ ).

Clase Forte	Forma básica	Clase Forte	Forma básica
6-15B	[0 1 2 4 5 8]	7-11	[0 1 3 4 5 6 8]
6-16	[0 1 4 5 6 8]	7-11B	[0 1 3 4 5 6 8]
6-16B	[0 1 4 5 6 8]	7-13	[0 1 2 4 5 6 8]
6-18	[0 1 2 5 7 8]	7-13B	[0 1 2 4 5 6 8]
6-18B	[0 1 2 5 7 8]	7-14	[0 1 2 3 5 7 8]
6-20	[0 1 4 5 8 9]	7-14B	[0 1 2 3 5 7 8]
6-21	[0 2 3 4 6 8]	7-15	[0 1 2 4 6 7 8]
6-21B	[0 2 3 4 6 8]	7-16	[0 1 2 3 5 6 9]
6-22	[0 1 2 4 6 8]	7-16B	[0 1 2 3 5 6 9]
6-22B	[0 1 2 4 6 8]	7-19	[0 1 2 3 6 7 9]
6-27	[0 1 3 4 6 9]	7-19B	[0 1 2 3 6 7 9]
6-27B	[0 1 3 4 6 9]	7-20	[0 1 2 4 7 8 9]
6-30	[0 1 3 6 7 9]	7-20B	[0 1 2 4 7 8 9]
6-30B	[0 1 3 6 7 9]	7-21	[0 1 2 4 5 8 9]
6-31	[0 1 3 5 8 9]	7-21B	[0 1 2 4 5 8 9]
6-31B	[0 1 3 5 8 9]	7-22	[0 1 2 5 6 8 9]
6-32	[0 2 4 5 7 9]	7-23	[0 2 3 4 5 7 9]
6-33	[0 2 3 5 7 9]	7-23B	[0 2 3 4 5 7 9]
6-33B	[0 2 3 5 7 9]	7-24	[0 1 2 3 5 7 9]
6-34	[0 1 3 5 7 9]	7-24B	[0 1 2 3 5 7 9]
6-34B	[0 1 3 5 7 9]	7-25	[0 2 3 4 6 7 9]
6-35	[0 2 4 6 8 10]	7-25B	[0 2 3 4 6 7 9]
7-1	[0 1 2 3 4 5 6]	7-26	[0 1 3 4 5 7 9]
7-2	[0 1 2 3 4 5 7]	7-26B	[0 1 3 4 5 7 9]
7-2B	[0 1 2 3 4 5 7]	7-27	[0 1 2 4 5 7 9]
7-3	[0 1 2 3 4 5 8]	7-27B	[0 1 2 4 5 7 9]
7-3B	[0 1 2 3 4 5 8]	7-28	[0 1 3 5 6 7 9]
7-4	[0 1 2 3 4 6 7]	7-28B	[0 1 3 5 6 7 9]
7-4B	[0 1 2 3 4 6 7]	7-29	[0 1 2 4 6 7 9]
7-5	[0 1 2 3 5 6 7]	7-29B	[0 1 2 4 6 7 9]
7-5B	[0 1 2 3 5 6 7]	7-30	[0 1 2 4 6 8 9]
7-6	[0 1 2 3 4 7 8]	7-30B	[0 1 2 4 6 8 9]
7-6B	[0 1 2 3 4 7 8]	7-31	[0 1 3 4 6 7 9]
7-7	[0 1 2 3 6 7 8]	7-31B	[0 1 3 4 6 7 9]
7-7B	[0 1 2 3 6 7 8]	7-32	[0 1 3 4 6 8 9]
7-8	[0 2 3 4 5 6 8]	7-32B	[0 1 3 4 6 8 9]
7-9	[0 1 2 3 4 6 8]	7-33	[0 1 2 4 6 8 10]
7-9B	[0 1 2 3 4 6 8]	7-34	[0 1 3 4 6 8 10]
7-10	[0 1 2 3 4 6 9]	7-35	[0 1 3 5 6 8 10]
7-10B	[0 1 2 3 4 6 9]	8-1*	[0 1 2 3 4 5 6 7]

Cuadro A.1: Clases Forte y sus conjuntos tonales en forma básica asociados observados en la base de datos (espacio de estados de la Cadena de Markov  $\{C_n\}_{n \in \mathbb{N}}$ ).

Clase Forte	Forma básica	Clase Forte	Forma básica
8-2	[0 1 2 3 4 5 6 8]	8-25	[0 1 2 4 6 7 8 10]
8-2B	[0 1 2 3 4 5 6 8]	8-26	[0 1 2 4 5 7 9 10]
8-3	[0 1 2 3 4 5 6 9]	8-27	[0 1 2 4 5 7 8 10]
8-4	[0 1 2 3 4 5 7 8]	8-27B	[0 1 2 4 5 7 8 10]
8-4B	[0 1 2 3 4 5 7 8]	8-28	[0 1 3 4 6 7 9 10]
8-5	[0 1 2 3 4 6 7 8]	9-1	[0 1 2 3 4 5 6 7 8]
8-5B	[0 1 2 3 4 6 7 8]	9-2	[0 1 2 3 4 5 6 7 9]
8-6	[0 1 2 3 5 6 7 8]	9-2B	[0 1 2 3 4 5 6 7 9]
8-7	[0 1 2 3 4 5 8 9]	9-3	[0 1 2 3 4 5 6 8 9]
8-8	[0 1 2 3 4 7 8 9]	9-3B	[0 1 2 3 4 5 6 8 9]
8-9	[0 1 2 3 6 7 8 9]	9-4	[0 1 2 3 4 5 7 8 9]
8-10	[0 2 3 4 5 6 7 9]	9-4B	[0 1 2 3 4 5 7 8 9]
8-11	[0 1 2 3 4 5 7 9]	9-5	[0 1 2 3 4 6 7 8 9]
8-11B	[0 1 2 3 4 5 7 9]	9-5B	[0 1 2 3 4 6 7 8 9]
8-12	[0 1 3 4 5 6 7 9]	9-6	[0 1 2 3 4 5 6 8 10]
8-12B	[0 1 3 4 5 6 7 9]	9-7	[0 1 2 3 4 5 7 8 10]
8-13	[0 1 2 3 4 6 7 9]	9-7B	[0 1 2 3 4 5 7 8 10]
8-13B	[0 1 2 3 4 6 7 9]	9-8	[0 1 2 3 4 6 7 8 10]
8-14	[0 1 2 4 5 6 7 9]	9-8B	[0 1 2 3 4 6 7 8 10]
8-14B	[0 1 2 4 5 6 7 9]	9-9	[0 1 2 3 5 6 7 8 10]
8-16	[0 1 2 3 5 7 8 9]	9-10	[0 1 2 3 4 6 7 9 10]
8-16B	[0 1 2 3 5 7 8 9]	9-11	[0 1 2 3 5 6 7 9 10]
8-17	[0 1 3 4 5 6 8 9]	9-11B	[0 1 2 3 5 6 7 9 10]
8-18	[0 1 2 3 5 6 8 9]	9-12	[0 1 2 4 5 6 8 9 10]
8-18B	[0 1 2 3 5 6 8 9]	10-1	[0 1 2 3 4 5 6 7 8 9]
8-19	[0 1 2 4 5 6 8 9]	10-2	[0 1 2 3 4 5 6 7 8 10]
8-19B	[0 1 2 4 5 6 8 9]	10-3	[0 1 2 3 4 5 6 7 9 10]
8-20	[0 1 2 4 5 7 8 9]	10-4	[0 1 2 3 4 5 6 8 9 10]
8-21	[0 1 2 3 4 6 8 10]	10-5	[0 1 2 3 4 5 7 8 9 10]
8-22	[0 1 2 3 5 6 8 10]	10-6	[0 1 2 3 4 6 7 8 9 10]
8-22B	[0 1 2 3 5 6 8 10]	11-1	[0 1 2 3 4 5 6 7 8 9 10]
8-23	[0 1 2 3 5 7 8 10]	12-1	[0 1 2 3 4 5 6 7 8 9 10 11]
8-24	[0 1 2 4 5 6 8 10]		

Cuadro A.1: Clases Forte y sus conjuntos tonales en forma básica asociados observados en la base de datos (espacio de estados de la Cadena de Markov  $\{C_n\}_{n \in \mathbb{N}}$ ).

## ANEXO II. APLICACIÓN DE LA MODELACIÓN PROPUESTA SOBRE UNA OBRA MUSICAL

Con el objetivo de ilustrar la modelación llevada a cabo de manera específica, así como su implementación, en el presente apartado se describirá el proceso de manejo de datos, estimación paramétrica y resultados obtenidos para *Polonaise Op. 89 en Do Mayor* de Ludwig van Beethoven.

En primera instancia, a través de la paquetería `music21` de Python es posible trabajar computacionalmente con obras musicales vía objetos denominados `streams`, parecidos a listas, los cuales contienen la información de una obra musical: signo de compás, armadura, melodía y acompañamiento –notas, acordes y sus duraciones–, etc. Ahora bien, la obra musical con que se ilustrará está contenida en la base de datos *Yale Classical Music Corpus* [36], que contiene en formato `csv` las intervenciones de la pieza musical junto con sus duraciones respectivas –melodía y acompañamiento juntos–. Así, a partir del corpus de Yale, las intervenciones de *Polonaise Op. 89 en Do Mayor* son guardadas en un `stream` de `music21` almacenando de manera iterativa la información correspondiente a las notas que constituyen a cada intervención, así como sus duraciones respectivas.

Posteriormente, la obra musical es transformada como fue descrito a inicios del Capítulo 5 (p. 90). De hecho, la Figura 5.1 muestra los primeros cuatro compases de la obra en cuestión transformada. Así, para cada intervención en la obra musical, si no constituye un silencio, es llevada a su posición básica<sup>6</sup> y a la cuarta octava con el comando: `.closedPosition(inplace = True, forceOctave = 4)`.

---

<sup>6</sup>Véase Figura 1.9.

## Modelación mediante Cadenas de Markov

Tras haber llevado a cabo la transformación de la obra musical, son recabadas las realizaciones de las Cadenas de Markov asociadas a los siguientes atributos de sus intervenciones:

- Duraciones ( $\{D_n\}_{n \in \mathbb{N}}$ ), con el comando `.duration.type`.
- Notas fundamentales ( $\{R_n\}_{n \in \mathbb{N}}$ ). Si la intervención es un acorde o nota, con la aplicación de `.pitch.midi - 59` sobre la raíz del acorde –o, simplemente, la nota– mediante el cual se obtiene su numeración midi, a la que se resta 59 para obtener la codificación de 1 a 12 que representa los semitonos de la octava de *Do* a *Si*. En caso de que la intervención sea un silencio, se asigna a la realización de la cadena el número cero.
- Clases Forte ( $\{C_n\}_{n \in \mathbb{N}}$ ). En caso de ser un silencio, se asigna la clase “0 – 1”, de lo contrario, se obtiene su clasificación de Allen Forte mediante el comando `.forteClassTnI`.

Por ejemplo, para los primeros dos compases de *Polonaise Op. 89 en Do Mayor*, las realizaciones de las tres cadenas respectivamente están dadas como muestra el siguiente cuadro, las cuales obedecen la descripción del espacio de estados dada en el Cuadro 5.2.



$n$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
$D_n$	negra	s	s	s	s	s	s	s	s	s	s	s	s	s	s	s	s	s	s	s	s
$R_n$	1	0	1	3	5	6	5	6	8	1	1	6	5	6	5	6	8	1	1	3	5
$C_n$	3-11	0-1	4	3	3	4	3	4	4	4	4	6	3	4	3	3	4	4	4	3	3

Cuadro A.2: Transformación de las intervenciones en los primeros dos compases de *Polonaise Op. 89 en Do Mayor* de Beethoven y las Cadenas de Markov asociadas a su duración, raíz y Clase Forte. Para las realizaciones 3 a 21 de la cadena  $C_n$  se omite el “2-” que les precede, por presentación. Por “s” se denota “semicorchea”.

Habiendo obtenido las realizaciones de las tres cadenas asociadas a las intervenciones de la obra musical –que se asumen homogéneas–, se prosigue a estimar por Máxima Verosimilitud la matriz de transición de cada una, aumentando al estimador un sesgo de suavizamiento. Esto es,

la entrada correspondiente a la  $i$ -ésima fila de la  $j$ -ésima columna de cada matriz de transición se estima como sigue:

$$\widehat{P}_{ij} = \frac{n_{ij}}{n_i} + \frac{1}{n_i};$$

donde  $n_{ij}$  corresponde al número de veces que a lo largo de la obra musical el estado  $j$  de la cadena fue precedido por el estado  $i$ ,  $n_i = \sum_{k \in S} n_{ik}$  (denotando por  $S$  el espacio de estados de la cadena) y el segundo término es el suavizamiento aumentado.

En el Cuadro A.3 se muestran las matrices de transición estimadas para las Cadenas de Markov asociadas a la duración (A.3a) y nota fundamental (A.3b) de las intervenciones de *Polonaise Op. 89 en Do Mayor* de Beethoven. Por fines de presentación, se omite la matriz de transición correspondiente a la cadena asociada a las Clases Forte, ya que la cardinalidad de su espacio de estados es 277.

Por construcción, las cadenas definidas tienen espacios de estados finitos, por lo que cuentan con la característica de ser recurrentes positivas, de acuerdo con el Teorema 2.2. Adicionalmente, como fue descrito en el Capítulo 5, son irreducibles. Por lo tanto, por el Teorema 2.1, existen sus distribuciones estacionarias (probabilidades límite), son únicas y sus entradas son estrictamente positivas.

Estado	Redonda	Negra	Corchea	Semicorchea	Fusa	Semifusa	Otra
Redonda	0.14	0.14	0.14	0.14	0.14	0.14	0.14
Negra	0.003	0.02	0.01	0.7	0.23	0.03	0.003
Corchea	0.001	0.02	0.16	0.43	0.39	0.01	0.0004
Semicorchea	0.001	0.15	0.85	0.002	0.002	0.001	0.001
Fusa	0.001	0.13	0.86	0.001	0.004	0.001	0.002
Semifusa	0.03	0.81	0.03	0.03	0.03	0.03	0.03
Otra	0.12	0.12	0.12	0.25	0.12	0.12	0.12

(a) Duraciones.

Cuadro A.3: Matriz de transición de las Cadenas de Markov asociadas a las duraciones y notas fundamentales de las intervenciones de *Polonaise Op. 89 en Do Mayor* de Beethoven. Para la cadena asociada a las notas fundamentales se presentan las cifras en porcentajes por presentación.

Estado	Sil	Do	Do♯	Re	Re♯	Mi	Fa	Fa♯	Sol	Sol♯	La	La♯	Si
Sil	0.04	18	2	6	3	16	9	3	25	3	7	1	5
Do	96	0.4	0.2	0.2	0.2	0.2	0.4	0.2	1	0.2	0.2	0.2	0.2
Do♯	83	1	1	1	1	1	1	1	1	1	1	1	1
Re	90	1	1	3	1	1	1	1	1	1	1	1	1
Re♯	87	1	1	1	1	1	1	1	1	1	1	1	1
Mi	92	0.2	0.2	0.2	0.2	5	1	0.2	0.5	0.2	0.2	0.2	0.2
Fa	92	1	0.4	0.4	0.4	1	2	0.4	1	0.4	0.4	0.4	0.4
Fa♯	87	1	1	1	1	1	1	1	1	1	1	1	1
Sol	94	1	0.2	0.3	0.2	0.3	0.3	0.2	3	0.2	0.2	0.2	0.2
Sol♯	87	1	1	1	1	1	1	1	1	1	1	1	1
La	94	1	1	1	1	1	1	1	1	1	1	1	1
La♯	72	2	2	2	2	2	2	2	2	2	2	2	2
Si	91	1	1	1	1	1	1	1	1	1	1	1	1

(b) Notas fundamentales.

Cuadro A.3: Matriz de transición de las Cadenas de Markov asociadas a las duraciones y notas fundamentales de las intervenciones de *Polonaise Op. 89 en Do Mayor* de Beethoven. Para la cadena asociada a las notas fundamentales se presentan las cifras en porcentajes por presentación.

Luego, recuérdese que la Proposición 2.2 indica que, para Cadenas de Markov Homogéneas,  $\mathbb{P}(X_n = j | X_0 = i) = (P^n)_{ij}$  donde  $i$  y  $j$  son estados de la cadena. Tal hecho, aunado al argumento anterior, proporciona una vía para la estimación numérica de las probabilidades límite correspondientes a cada Cadena de Markov, puesto que

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n = j | X_0 = i) = \lim_{n \rightarrow \infty} (P^n)_{ij}.$$

Entonces, las estimaciones de las probabilidades límite de cada cadena fueron calculadas elevando sus matrices de transición a una potencia suficientemente grande, a saber, 10,000 –punto a partir del cual dejó de observarse cambio en las filas de las matrices de transición–.

$j$	Breve	Redonda	Blanca	Negra	Corchea	Semicorchea	Fusa	Semifusa	Otra
$\pi_j$	-	0.002	-	0.08	0.46	0.25	0.2	0.006	0.002

(a) Duraciones.

Cuadro A.4: Probabilidades límite de las Cadenas de Markov asociadas a las duraciones y notas fundamentales de las intervenciones de *Polonaise Op. 89 en Do Mayor* de Beethoven. Por  $j$  se denota el estado de la cadena y por  $\pi_j$  la probabilidad límite correspondiente a dicho estado.

$j$	Sil	Do	Do $\sharp$	Re	Re $\sharp$	Mi	Fa	Fa $\sharp$	Sol	Sol $\sharp$	La	La $\sharp$	Si
$\pi_j$	0.48	0.09	0.01	0.03	0.02	0.08	0.04	0.02	0.13	0.02	0.04	0.01	0.03

(b) Notas fundamentales.

Cuadro A.4: Probabilidades límite de las Cadenas de Markov asociadas a las duraciones y notas fundamentales de las intervenciones de *Polonaise Op. 89 en Do Mayor* de Beethoven. Por  $j$  se denota el estado de la cadena y por  $\pi_j$  la probabilidad límite correspondiente a dicho estado.

En el cuadro anterior se muestran los valores estimados de las probabilidades límite de Cadenas de Markov Homogéneas asociadas a las duraciones y notas fundamentales de las intervenciones de *Polonaise Op. 89 en Do Mayor*. Cabe destacar que las estimaciones de las matrices de transición –y, por lo tanto, de las probabilidades límite– fueron hechas con base en los estados observados en la obra. Razón por la cual, a las entradas de los vectores de las probabilidades límite correspondientes a los estados que las cadenas no visitaron en la obra se les asignó un valor nulo, para especificar que no fueron observados.

Finalmente, es debido hacer notar que las entradas de los vectores de probabilidades límite estimados para cada una de las tres cadenas planteadas constituyen 299 de los 302 predictores utilizados para el problema de clasificación expuesto en el Capítulo 6: nueve son las entradas del vector de probabilidades límite de la cadena asociada a las duraciones, trece de la asociada a las raíces de los acordes y los restantes 277 corresponden a las entradas del vector de las probabilidades límite de la cadena asociada a las Clases Forte.

## Modelación mediante Campos Aleatorios

En el presente apartado se ejemplificará la aplicación de la modelación por Campos Aleatorios propuesta para *Polonaise Op. 89 en Do Mayor*. En primer lugar, se crea un nuevo stream con las intervenciones de la obra musical transformadas de misma manera que para la modelación anterior, con la diferencia de que se omiten los silencios y las notas –i.e., se trabajará solamente con los acordes–. Luego, de cada intervención es extraído el vector de clase de intervalo con el comando `.intervalVector` y se obtiene su vector de clase de intervalo reducido  $[C D]$  sumando las entradas 2-3-4 y 1-2-6, de acuerdo con la Definición 1.16. En el siguiente cuadro se muestran los vectores de clase de intervalo reducidos asociados a las intervenciones de los primeros dos compases de *Polonaise Op. 89 en Do Mayor*, de Beethoven.



$n$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$D$	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
$C$	3	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1

Cuadro A.5: Transformación de las intervenciones en los primeros dos compases de *Polonaise Op. 89 en Do Mayor* de Beethoven y sus vectores de clase de intervalo reducidos.  $C$  corresponde a la entrada asociada a los intervalos consonantes de la intervención y  $D$  la asociada a los disonantes.

Ahora, se mostrará la aplicación llevada a cabo, la cual fue hecha asumiendo que las intervenciones de la obra musical constituyen un Modelo de Markov de orden  $k - 1 = 2 - 1 = 1$ . Es decir, considerando la secuencia de vectores de clase de intervalo reducidos de los conjuntos tonales  $A_1, \dots, A_n$  asociados a las intervenciones de *Polonaise Op. 89 en Do Mayor* denotados por  $\mathcal{I}_{m_1}^*(A_1), \mathcal{I}_{m_2}^*(A_2), \dots, \mathcal{I}_{m_n}^*(A_n)$ , para todo  $j \in \{2, \dots, n\}$  se asume que ocurre

$$\mathbb{P}(\mathcal{I}_{m_j}^*(A_j) = i_j \mid \mathcal{I}_{m_{j-1}}^*(A_{j-1}) = i_{j-1}, \dots, \mathcal{I}_{m_1}^*(A_1) = i_1) = \mathbb{P}(\mathcal{I}_{m_j}^*(A_j) = i_j \mid \mathcal{I}_{m_{j-1}}^*(A_{j-1}) = i_{j-1}).$$

Luego, siguiendo la modelación planteada, se considerará un Campo Aleatorio con la topología  $(S, N)$  dada por  $S = \{1, 2\} \times \{1, 2\}$  y  $N$  tal que para todo  $s \in S$ ,  $N_s = S \setminus s$ . Asimismo, el espacio de

configuraciones del campo está determinado por su espacio fase  $\Lambda$  y está dado por

$$\Lambda^S = \{(7, 3), (6, 9), (15, 13), (10, 11), (3, 0), (2, 1), (13, 8), (5, 1), (3, 7), (0, 3), (12, 9), \\ (36, 30), (1, 2), (11, 10), (3, 3), (5, 5), (17, 11), (6, 3), (1, 5), (14, 14), (12, 3), (6, 4), \\ (8, 2), (19, 17), (13, 15), (8, 7), (4, 2), (1, 0), (30, 25), (9, 6), (16, 12), (0, 1), (10, 5), \\ (21, 15), (4, 6), (14, 7), (24, 21), (25, 20), (18, 18), (7, 8), (2, 4), (9, 12), (20, 16)\}^2.$$

Ahora bien, recuérdese que el Campo Aleatorio se definió como un campo dinámico, por lo que las observaciones del campo son consideradas de manera secuencial, i.e.,

$$\left(\mathcal{I}_{m_1}^*(A_1), \mathcal{I}_{m_2}^*(A_2)\right), \left(\mathcal{I}_{m_2}^*(A_2), \mathcal{I}_{m_3}^*(A_3)\right), \dots, \left(\mathcal{I}_{m_{n-1}}^*(A_{n-1}), \mathcal{I}_{m_n}^*(A_n)\right).$$

Por otro lado, el Campo Aleatorio asociado a los conjuntos tonales de las intervenciones de la obra es caracterizado por el potencial de activación

$$V_C(x) = \begin{cases} \frac{66+x(2,1)-x(1,2)}{2 \cdot 66} & \text{si } C = \{(1, 2), (2, 1)\} =: \mathcal{C} \\ \frac{66+x(2,2)-x(1,1)}{2 \cdot 66} & \text{si } C = \{(1, 1), (2, 2)\} =: \mathcal{D} \\ 0 & \text{e.o.c.} \end{cases}$$

De manera que la energía del campo de forma paramétrica está dada por

$$\varepsilon(x) = \theta_{\mathcal{C}} \cdot V_{\mathcal{C}}(x) - \theta_{\mathcal{D}} \cdot V_{\mathcal{D}}(x) = \theta_{\mathcal{C}} \cdot \frac{66 + x(2, 1) - x(1, 2)}{2 \cdot 66} - \theta_{\mathcal{D}} \cdot \frac{66 + x(2, 2) - x(1, 1)}{2 \cdot 66};$$

donde  $\theta_{\mathcal{C}}, \theta_{\mathcal{D}} \geq 0$  son los parámetros asociados a los potenciales correspondientes a los *cliques* de la forma  $\mathcal{C}$  y  $\mathcal{D}$  –los factores de consonancia y disonancia asociados a *Polonaise Op. 89 en Do Mayor* de Beethoven–, respectivamente.

Posteriormente, se procede a estimar los parámetros  $\theta_{\mathcal{C}}$  y  $\theta_{\mathcal{D}}$  maximizando la función de log-Verosimilitud del campo, cuya expresión analítica general fue obtenida en la Sección 3.3.2, Capítulo 3 (p. 63). Para el modelo específico que se ha construido en este apartado, la función de log-Verosimilitud puede expresarse como sigue. Para  $i \in \{2, \dots, n\}$ , considérese la  $i$ -ésima observación del campo asociado a los conjuntos tonales de la pieza musical denotada por  $X^i$ , esto es,

$X^i = (\mathcal{I}_{m_{i-1}}^*(A_{i-1}), \mathcal{I}_{m_i}^*(A_i))$  y las observaciones del campo  $\underline{X} = (X^1, \dots, X^n)$ . Entonces,

$$\begin{aligned} \ell(\underline{X}; \theta_{\mathcal{E}}, \theta_{\mathcal{D}}) = & - \sum_{i=2}^n \left[ \theta_{\mathcal{E}} \cdot V_{\mathcal{E}}(X^i) - \theta_{\mathcal{D}} \cdot V_{\mathcal{D}}(X^i) \right] - \ln Z(\theta_{\mathcal{E}}, \theta_{\mathcal{D}}) + \\ & - \sum_{i=3}^n \ln \left( \sum_{\lambda \in \Lambda} \exp \left\{ - \theta_{\mathcal{E}} \cdot V_{\mathcal{E}}(X^i(\{(1,1), (1,2)\}), \lambda) + \right. \right. \\ & \left. \left. + \theta_{\mathcal{D}} \cdot V_{\mathcal{D}}(X^i(\{(1,1), (1,2)\}), \lambda) \right\} \right). \end{aligned}$$

Tal expresión fue optimizada mediante métodos numéricos, a saber, el Algoritmo Evolutivo de optimización estocástica con 1,000 iteraciones. Para ello fue considerado un Campo Aleatorio con las siguientes características:

- Sitios  $S = (P[0, 300])^2$  donde  $P[0, 300]$  es la partición equiespaciada del intervalo  $[0, 300]$  en 3,000 subintervalos. Nótese que los sitios representan el espacio paramétrico de  $(\theta_{\mathcal{E}}, \theta_{\mathcal{D}})$ .
- Vecindades tipo  $\beta$ , i.e., para cada sitio  $\theta = (\theta_1, \theta_2) \in S$ ,

$$\begin{aligned} N_{\theta} = & \{(\theta_1 + \delta, \theta_2 + \delta), (\theta_1, \theta_2 + \delta), (\theta_1 - \delta, \theta_2 + \delta), (\theta_1 + \delta, \theta_2), (\theta_1 - \delta, \theta_2), \\ & (\theta_1 + \delta, \theta_2 - \delta), (\theta_1, \theta_2 - \delta), (\theta_1 - \delta, \theta_2 - \delta)\}; \end{aligned}$$

donde  $\delta = 1/3,000$ .

- Espacio fase dado por  $\ell(\underline{X}; S)$ , donde el campo es tal que para cada sitio  $\theta = (\theta_1, \theta_2)$ , con probabilidad unitaria ocurre  $x(\{\theta\}) = \ell(\underline{X}; \theta_1, \theta_2)$ .

A continuación se describe brevemente el algoritmo de optimización. Para ello, deberá considerarse una constante  $\alpha \in (0, 1)$  a elección como parámetro:

1. Iniciar el contador de iteraciones  $i = 1$ .
2. Escoger un par ordenado  $\theta_i$  equiprobablemente sobre los sitios.
3. Generar un número aleatorio  $u$  con distribución Uniforme(0,1):  
Si  $u \leq \alpha$ , elegir  $\theta_j$  equiprobablemente sobre los sitios en la vecindad de  $\theta_i$ .  
De lo contrario, si  $u > \alpha$ , elegir  $\theta_j$  equiprobablemente sobre los sitios del campo.
4. Evaluar  $\ell(\underline{X}; \theta_i)$  y  $\ell(\underline{X}; \theta_j)$ . Si  $\ell(\underline{X}; \theta_i) < \ell(\underline{X}; \theta_j)$ , actualizar  $\theta_i = \theta_j$ .

5. Actualizar  $i = i + 1$ . Si  $i \leq 1,000$ , regresar al paso 2.

Es debido hacer notar que se llevó a cabo la optimización con parámetro  $\alpha = 0.7$ . Nótese que el complemento de tal cifra constituye la probabilidad de escoger aleatoriamente un sitio sobre  $S$ , en lugar de sobre la vecindad del sitio en cuestión; de manera que el algoritmo es eficiente para evitar la obtención de un máximo local.

La implementación de la maximización de la función de log-Verosimilitud para las observaciones del campo asociado a *Polonaise Op. 89 en Do Mayor* de Beethoven, resulta en que su punto máximo es  $(44.79, 21.9)$  y su valor máximo está dado por  $-11,230.77$ . Entonces, el factor de consonancia estimado asociado a Beethoven en su obra musical *Polonaise Op. 89 en Do Mayor* es  $\hat{\theta}_{\mathcal{C}} = 44.79$ , el factor de disonancia es  $\hat{\theta}_{\mathcal{D}} = 21.9$  y la evaluación de su función de log-Verosimilitud en tales puntos resulta en  $\hat{\ell}(\underline{X}; 44.79, 21.9) = -11,230.77$ . Por lo tanto, la medida de Gibbs del campo asociado a tal pieza musical está dada por

$$\pi(x) = \frac{1}{Z(44.79, 21.9)} \cdot \exp\{0.17 \cdot [x(2, 2) - x(1, 1)] - 0.34 \cdot [x(2, 1) - x(1, 2)] - 11.22\}.$$

Por último, cabe destacar que para la tarea de clasificación de obras musicales, además de los 299 obtenidos en el apartado anterior (probabilidades límite de las Cadenas de Markov asociadas a la pieza), se contemplan como predictores los valores estimados  $\hat{\theta}_{\mathcal{C}}$ ,  $\hat{\theta}_{\mathcal{D}}$  y  $\hat{\ell}(\underline{X}; \hat{\theta}_{\mathcal{C}}, \hat{\theta}_{\mathcal{D}})$ .



### ANEXO III. OBSERVACIONES ATÍPICAS

Autor	Obra
Beethoven	Sonata 28 i Op 101 A major String Quartet #10 Op 74 ii Ab major Violin Sonata #7 Op 30 No 2 ii Ab major
Dvorak	Slavonic Dances in B Op72No1 B major
Handel	Allegro HWV 325 Bb Major Gigue HWV 327 F Major Gigue HWV 350 G Major Largo affettuoso HWV 324 G Minor Largo HWV 323 B Minor Lentement HWV 349 D Major Minuet HWV 323 D Major
Haydn	Tempore Belli Mass Hob22-9 vi Agnus Dei C major The Creation Hob21-2 1 Introduction C minor
Hough	Cheetham's Close Eb Major
Palestrina	Gloria 15 H, Gloria 68 I, Gloria 69 I Gloria 74 F, Kyrie 62 B, Sanctus 40
Rameau	Nouvellas Suites de Pieces de Clavecin 5 Fanfarinette A major
Ravel	Valses Nobles et Sentimentales 1 Modere Tres Franc G major
Schubert	Im Dorfe D911 XVII D Major Landler 1 D790 D Major
Shostakovich	Aphorisms 7 Dance of Death Op13No7 F minor
Vivaldi	Concerto 129 III G Major Concerto 3 7 III F Major Concerto 537 I C Major

Cuadro A.6: Observaciones atípicas en la base de datos con base en el último Componente Principal correspondiente al análisis por corriente musical histórica.

Autor	Obra
Bach	Ich will den Kreuzstab gerne tragen BWV 56 Ich will den Kreuzstab gerne tragen G minor Jauchzet Gott in allen Landen BWV 51 Jauchzet Gott in allen Landen C major WTC 1.8 Fugue D# minor
Handel	Gigue HWV 350 G Major Largo affettuoso HWV 324 G Minor Largo HWV 323 B Minor
Haydn	Tempore Belli Mass Hob22-9 vi Agnus Dei C major The Creation Hob21-2 1 Introduction C minor
Khachaturian	Album for Children 8 Fuga C major
Liszt	Cantique S173 X E Major
Palestrina	Gloria 15 H Gloria 68 I Gloria 69 I Gloria 74 F Kyrie 62 B Sanctus 40
Saint-Saens	6 Etudes for the Left Hand Op 135 Alla fuga G major 6 Etudes for the Left Hand Op 135 Elegie Db major
Schubert	Im Dorfe D911 XVII D Major Landler 1 D790 D Major
Vivaldi	Concerto 129 III G Major Concerto 537 I C Major

Cuadro A.7: Observaciones atípicas en la base de datos con base en el último Componente Principal correspondiente al análisis por autor.

## ANEXO IV. ANÁLISIS DE COMPONENTES PRINCIPALES ADICIONAL

En el *Capítulo 5: Modelación matemática de obras musicales*, se expuso un Análisis de Componentes Principales, el cual fue llevado a cabo aplicando tal método sobre las estimaciones asociadas a las obras musicales de la base de datos correspondientes a las probabilidades límite de las Cadenas de Markov, factores de consonancia y disonancia y log-Verosimilitud en términos de la medida de Gibbs. Cabe destacar que, para tal finalidad, las variables fueron reescaladas para tener media cero y varianza unitaria; y con justa razón, ya que a pesar de que todas las estimaciones de los vectores de probabilidad están en la escala  $[0, 1]$ , claramente, para los factores de consonancia y disonancia y la evaluación de la función de log-Verosimilitud no ocurre lo mismo.

Adicionalmente, aunado al análisis exploratorio que brinda realizar un PCA, nótese que incluir los Componentes Principales a la metodología de clasificación es indispensable, en especial para los GLM, LDA y QDA. Esto, puesto que teóricamente mediante ellos se asume independencia en los predictores, lo cual no es verdadero para las variables introducidas inicialmente, como es evidenciado por el Análisis de Conglomerados Jerárquico llevado a cabo en el Capítulo 5 (véase Figura 5.8).

Además, en la siguiente figura se exponen las 34 variables que muestran coeficientes de correlación en valor absoluto mayores a 0.5 con al menos otra: las entradas de los vectores asociados a las probabilidades límite de duraciones *Redonda*, *Blanca*, *Negra*, *Corchea*, *Semicorchea* y *Fusa*; las probabilidades del uso de silencios a largo plazo; las correspondientes a las notas fundamentales *Do*, *Re*, *Fa*, *Sol*, *Sol#*, *La* y *Si*; las Clases Forte *0-1*, *1-1*, *2-3*, *2-4*, *2-5*, *3-4*, *3-7*, *3-9*, *3-10*, *3-11*,

4-14, 4-20, 4-22, 4-23, 4-26, 4-27, 4-28 y 5-31; los parámetros de consonancia  $\hat{\theta}_c$  y la función log-Verosimilitud maximizada en términos de las medidas de Gibbs de cada obra musical.

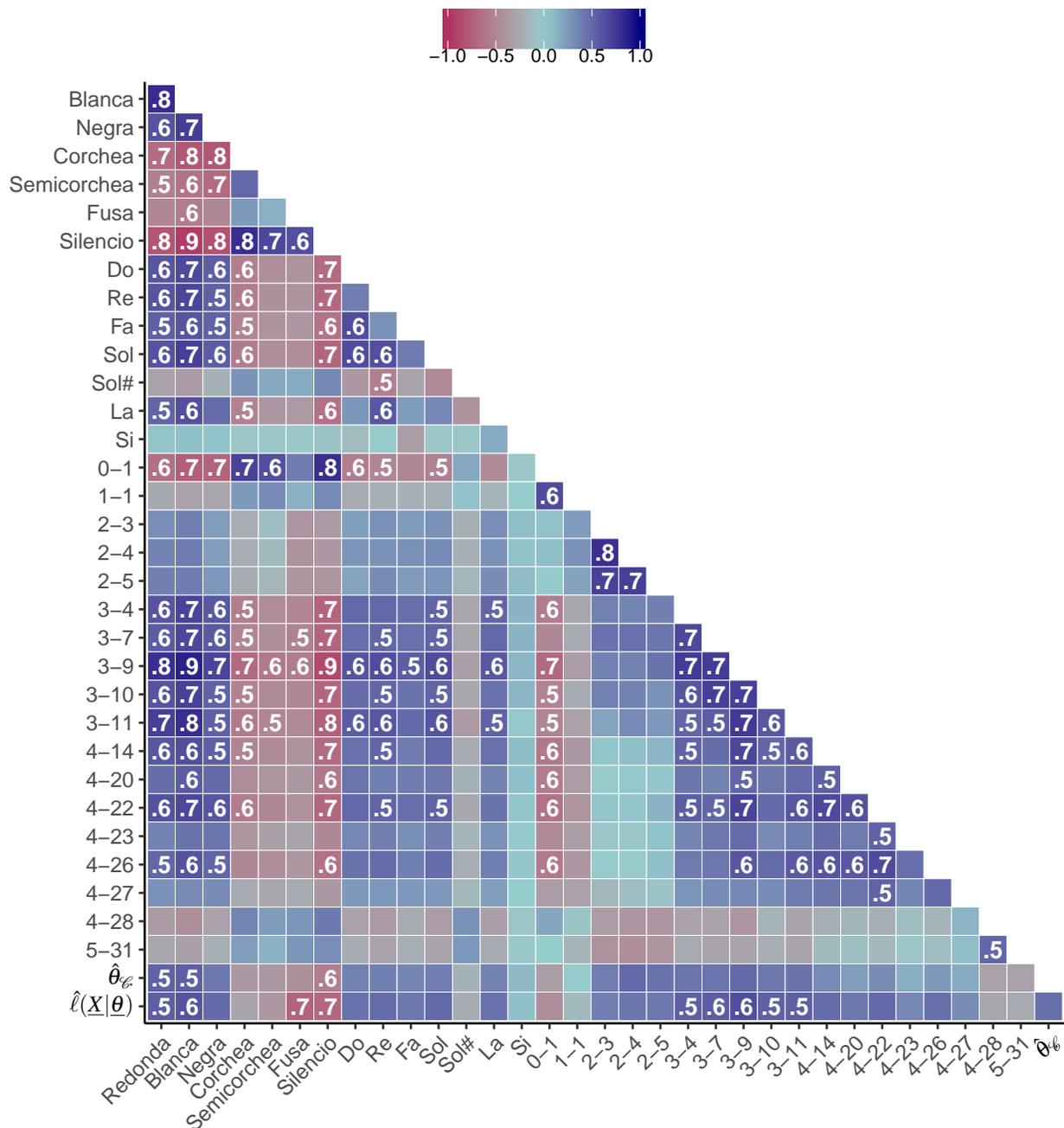


Figura A.1: Matriz de correlación de Pearson de las estimaciones de las probabilidades límite, factores de consonancia y disonancia y evaluación de la función de log-Verosimilitud en términos de la medida de Gibbs. Se presentan sólo aquellas variables cuya correlación con al menos una de las restantes sea mayor o igual a 0.5 en valor absoluto, en esos casos se reporta tal valor.

Ahora bien, en medida de que el objetivo del Capítulo 6 es la clasificación de obras musicales, se llevó a cabo también un Análisis de Componentes Principales sin reescalar las variables para aplicarlo a la clasificación e, inicialmente, con el fin único de averiguar su desempeño en la metodología en términos de error de clasificación. Como se discute en el mismo capítulo, en general, se alcanzaron tasas de error menores a través de este conjunto de datos al aplicarlos a Modelos Lineales Generalizados y Análisis de Discriminante Lineal, razón por la cual se presenta a continuación el análisis correspondiente.

En primer lugar, la varianza mínima observada en las variables asociadas a las probabilidades límite es de  $1.87e-6$  y la máxima de 0.048; mientras que las asociadas a los factores de consonancia, disonancia y log-Verosimilitud, respectivamente, son 444.97, 50.83 y 11,145,669. Dado que los Componentes Principales son construidos de manera que su varianza sea decreciente respecto su índice, el no reescalar las variables tiene como resultado que la varianza de los primeros componentes (los eigenvalores de la matriz de varianzas y covariancias) sea *altísima*, en proporción al resto. Tal hecho puede apreciarse en el siguiente cuadro, en que se muestran los datos hasta el sexto Componente Principal puesto que a partir de él, las desviaciones estándar de los componentes son menores a 0.1 y las proporciones de varianza explicada tan cercanas a cero que se pierden sus decimales por la precisión de la máquina.

Componente Principal	1	2	3	4	5	6
Desviación estándar	3,338.53	18,76	6.79	0.32	0.11	0.1
Proporción de varianza explicada	99.996	0.0003	<2e-16	<2e-16	<2e-16	<2e-16
Varianza explicada acumulada	99.996	100	100	100	100	100

Cuadro A.8: Desviación estándar y varianza explicada por los Componentes Principales obtenidos sin reescalar las variables.

Asimismo, el hecho de que las varianzas de las últimas tres variables mencionadas sean tantas veces mayores que el resto, tiene impacto directo en los *loadings* asociados a ellas. En el cuadro siguiente se presentan los diez *loadings* con mayor magnitud (en milésimas) correspondientes a los primeros dos Componentes Principales, con los que se acumula prácticamente el 100% de la proporción de varianza explicada. Nótese que, en efecto, los pesos asociados a los factores de consonancia y disonancia y a la evaluación de la función de log-Verosimilitud son mucho mayores en proporción al resto. Normalmente, esto se interpretaría como que son variables *más importantes*, pues se les da más peso para explicar la varianza de los datos. Sin embargo, en este caso tal impor-

tancia podría catalogarse como *artificial*, en el sentido de que es derivado de sus altísimas varianzas (constantes por que se multiplican las variables centradas), no de qué tanto logren explicar la información dada por los datos. Justo esta es la razón por la que, en estos casos, lo adecuado es aplicar el análisis sobre las variables reescaladas.

PC <sub>1</sub>	$\hat{\ell}(\underline{X} \underline{\theta})$ 999.99	$\hat{\theta}_{\mathcal{L}}$ 2.9	$\hat{\theta}_{\mathcal{D}}$ -0.58	Sil. -0.39	Blanca 0.032	Negra 0.026	Fusa -0.025	Corchea -0.02	0-1 -0.019	S. -0.014
PC <sub>2</sub>	$\hat{\theta}_{\mathcal{L}}$ 998.46	$\hat{\theta}_{\mathcal{D}}$ 55.17	Corchea -3.14	$\hat{\ell}(\underline{X} \underline{\theta})$ -2.87	Blanca 2.79	Sil. -2.59	Negra 1.93	S. -1.43	0-1 -1.08	3-11 0.81

Cuadro A.9: *Loadings* asociados a los primeros dos Componentes Principales obtenidos sin reescalar las variables. Se presentan aquellos diez de magnitud mayor. Por presentación “Sil.” representa Silencio y “S.”, Semicorchea. Cifras en milésimas.

Dicho lo anterior, resulta sorprendente que mediante este conjunto de datos se logren mejores tasas de clasificación errónea en el Capítulo 6. No obstante, obsérvese que mediante los primeros dos Componentes Principales se logra ver la separación al menos del grupo *Renacentista* ó *Palestrina* del resto. Para las otras clases la separación no es tan evidente, pero ello tampoco se logra mediante los primeros dos PCs obtenidos al reescalar las variables<sup>7</sup>.

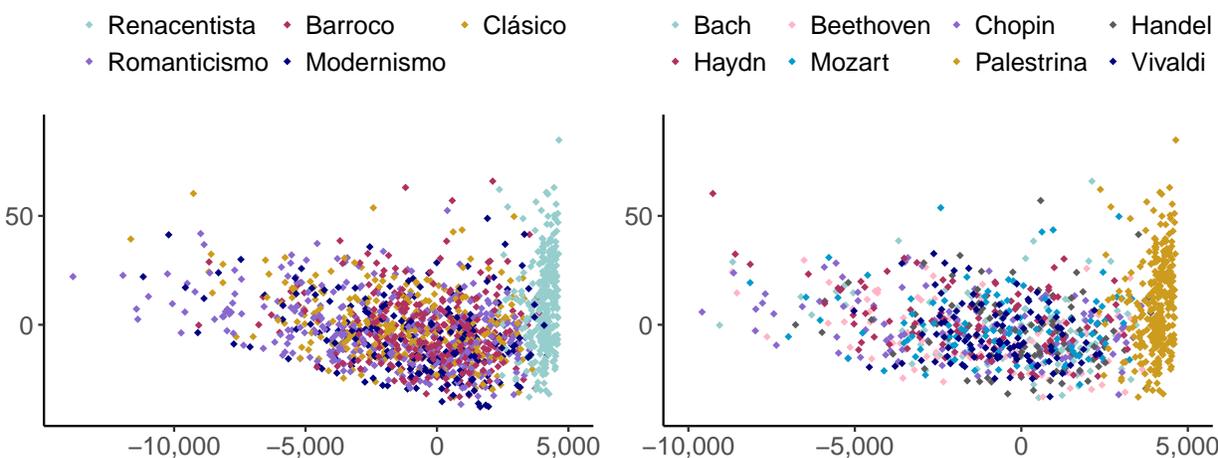


Figura A.2: Proyección de las observaciones a los primeros dos Componentes Principales obtenidos sin reescalar las variables.

<sup>7</sup>Véase Figura 5.14 y Figura 5.16.

## BIBLIOGRAFÍA

- [1] Agresti, A. (2007). *An Introduction to Categorical Data Analysis*. Wiley Series in Probability and Statistics, Wiley, 2a Edición, New Jersey.
- [2] Allen, R., Cohen, D., Hager, N., Taylor, J. (2006). *Music: Its language, history, and culture*. Conservatory of Music at Brooklyn College of the City University of New York.
- [3] Alvarez García, J. L. (2009). *Galileo y Vincenzo. La Música y el nacimiento del método experimental*. Revista Ciencias No. 100, Facultad de Ciencias, UNAM. URL: <http://www.revistaciencias.unam.mx/pt/102-revistas/revista-ciencias-100/715-galileo-y-vincenzo-la-musica-y-el-nacimiento-del-metodo-experimental.html>.
- [4] Ayala Velázquez, D., Lonngi V., P.A. (2008). *Contribuciones de Leonhard Euler a la acústica*. Miscelánea Matemática 46 1-21, UAM. URL: <http://albertofest.matcuer.unam.mx/Misc46/Lonngi.pdf>.
- [5] Bertos, M. del C. (S.F.). *Música y Matemáticas*. IES Villanueva del Mar, España.
- [6] Brémaud, P. (1999). *Markov Chains, Gibbs Fields, Monte Carlo Simulation, and Queues*. Springer, Nueva York.
- [7] Daubs, K. (2011). *The house that math built*. The Star. URL: [https://www.thestar.com/news/gta/2011/02/04/the\\_house\\_that\\_math\\_built.html](https://www.thestar.com/news/gta/2011/02/04/the_house_that_math_built.html).
- [8] Descartes, R. (1618). *Compendio de Música*. Ed. Tecnos, España.
- [9] de la Ossa Martínez, M. A. (2013). *Un acercamiento a la Música 'Clásica': Rompiendo tópicos*. Didáctica de la Expresión Musical, Facultad de Educación de Cuenca.

- [10] de Régules, S. (2015). *Cien años de la teoría general de la relatividad*. Revista de Divulgación Científica de la UNAM: ¿Cómo ves?, No. 204. URL: <http://www.comoves.unam.mx/numeros/articulo/204/cien-anos-de-la-teoria-general-de-la-relatividad>.
- [11] Forte, A. (1973). *The Structure of Atonal Music*. Yale University Press, New Haven and London.
- [12] Friedman, J., Hastie, T., Tibshirani, R. (2000). *Additive Logistic Regression: a Statistical View of Boosting*. Stanford University; The Annals of Statistics, Vol. 28, No. 2, pp. 337-407. URL: <https://web.stanford.edu/~hastie/Papers/AdditiveLogisticRegression/alr.pdf>.
- [13] Grout, D., Palisca, C. (2001). *Historia de la Música Occidental*. Vol. I y II, Ed. Alianza.
- [14] Guillotel-Nothmann, C. (2009). *Dissonance and harmonic progression. The impact of the seconda pratica on the advent of tonality*. 2nd International Conference for PhD Music Students, Thessalonique, Greece.
- [15] Gupte, M. (2018). *Tabla plays a crucial role in Fields Medal winner Manjul Bhargava's mathematical research*. The Economic Times. URL: [http://www.economictimes.indiatimes.com/articleshow/63496944.cms?utm\\_source=contentofinterest&utm\\_medium=text&utm\\_campaign=cppst](http://www.economictimes.indiatimes.com/articleshow/63496944.cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst).
- [16] Hastie, T., Tibshirani, R., Friedman, J. (2009). *Elements of Statistical Learning*. Springer, Nueva York.
- [17] Hastie, T., Tibshirani, R., Wainwright, M. (2005). *Statistical Learning with Sparsity: The Lasso and Generalizations*. Chapman & Hall.
- [18] Howat, R. (1994). *Debussy, Bartók and Nature Forms*. Symmetry: Culture and Science, Vol. 5, No. 3, 269-285.
- [19] James, G., Witten, D., Hastie, T., Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer, Nueva York.
- [20] Kellner, H. A. (1978). *Was Bach a Mathematician?*. The English Harpsichord Magazine, Vol. 2 No. 2.
- [21] Kostka, S. (2012). *Materials and Techniques of Post-tonal Music*. 4<sup>o</sup> Edición, Ed. Routledge, Nueva York.
- [22] Kuusi, T. (2001). *Set-Class and Chord: Examining Connection between Theoretical Resemblance and Perceived Closeness*. Studia Musica 12, Sibelius Academy, Helsinki.
- [23] Laborecky, J. (1984). *Music Terminological Dictionary*. SPN, Bratislava, 2000.

- [24] Mandereau, J., Ghisi, D., Amiot, E., Andreatta, M., Agon, C. (2011). *Z-Relation and Homometry in Musical Distributions*. Dipartimento di Matematica, Università di Pisa, Italy; Coursus de composition, IRCAM, Paris, France; CPGE, Perpignan, France.
- [25] Marsík, L. (2013). *Music Harmony Analysis: Towards a Harmonic Complexity of Musical Pieces*. Facultad de Matemáticas, Física e Informática, Comenius University, Bratislava.
- [26] Maxwell, H. J. (1992). *An expert system for harmonic analysis of tonal music*. Understanding music with AI; Cambridge, MIT Press.
- [27] May, B. H. (2007). *A survey of radial velocities in the zodiacal dust cloud*. Department of Physics, Imperial College London.
- [28] Real Academia Española. (S.F.) *Diccionario de la Lengua Española*. Edición del Tricentenario.
- [29] Schmidt-Jones, C. (2007). *Understanding Basic Music Theory*. Rice University, Houston, Texas.
- [30] Sethares, W. A. (1999). *Tuning, Timbre, Spectrum, Scale*. Springer, E.E.U.U.
- [31] Temperley, D. (1997). *An Algorithm for Harmonic Analysis*. Music Perception: An Interdisciplinary Journal, Vol. 15, No. 1 (Fall, 1997), University of California Press, pp. 31-68.
- [32] Tibshirani, R. (2011). *The lasso: some novel algorithms and applications*. Stanford University. Disponible en: <http://statweb.stanford.edu/~tibbs/ftp/lassotalk.pdf>
- [33] Tomasini, M. C. (S.F.). *El fundamento matemático de la escala musical y sus raíces pitagóricas*. Universidad de Palermo.
- [34] Venables, W. N., Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer, 4a Edición.
- [35] Xenakis, I. (1991). *Formalized Music. Thought and Mathematics in Composition*. Pendragon Press.
- [36] Yale Classical Archives Corpus. (2019). Disponible en: <https://ycac.yale.edu/downloads>.