



**INSTITUTO POTOSINO DE INVESTIGACIÓN  
CIENTÍFICA Y TECNOLÓGICA, A.C.**

**Prueba de irreversibilidad en el análisis de  
Electrocardiogramas**

Tesis que presenta

**Nazul Bonfilio Merino Negrete**

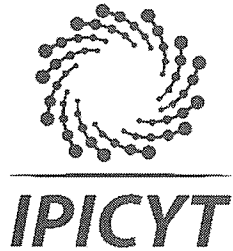
Para obtener el grado de

**Maestro en Ciencias en Control y Sistemas Dinámicos**

**Director de la Tesis:**

Dr. César Octavio Maldonado Ahumada

San Luis Potosí, S.L.P., 16 de Agosto de 2018



## Constancia de aprobación de la tesis

La tesis "***Prueba de irreversibilidad en el análisis de Electrocardiogramas***" presentada para obtener el Grado de Maestro en Control y Sistemas Dinámicos, fue elaborada por **Nazul Bonfilio Merino Negrete** y aprobada el **dieciséis de agosto del dos mil dieciocho** por los suscritos, designados por el Colegio de Profesores de la División de Matemáticas Aplicadas del Instituto Potosino de Investigación Científica y Tecnológica, A.C.

**Dr. Cesar Octavio Maldonado Ahumada**  
Director de la tesis

**Dr. Haret-Codratian Rosu Barbus**  
Jurado en el Examen

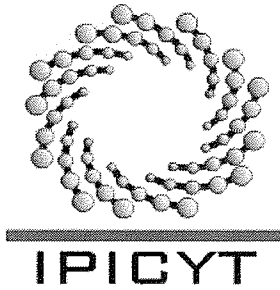
**Dr. Raúl Salgado García**  
Jurado en el Examen



## **Créditos Institucionales**

Esta tesis fue elaborada en la División de Matemáticas Aplicadas del Instituto Potosino de Investigación Científica y Tecnológica, A.C., bajo la dirección del Dr. César Octavio Maldonado Ahumada.

Durante la realización del trabajo el autor recibió una beca académica del Consejo Nacional de Ciencia y Tecnología con número de registro 612806 y del Instituto Potosino de Investigación Científica y Tecnológica, A. C.



# Instituto Potosino de Investigación Científica y Tecnológica, A.C.

## Acta de Examen de Grado

El Secretario Académico del Instituto Potosino de Investigación Científica y Tecnológica, A.C., certifica que en el Acta 033 del Libro Primero de Actas de Exámenes de Grado del Programa de Maestría en Control y Sistemas Dinámicos está asentado lo siguiente:

En la ciudad de San Luis Potosí a los 16 días del mes de agosto del año 2018, se reunió a las 16:00 horas en las instalaciones del Instituto Potosino de Investigación Científica y Tecnológica, A.C., el Jurado integrado por:

<b>Dr. Raúl Salgado García</b>	<b>Presidente</b>	<b>UAEM</b>
<b>Dr. Cesar Octavio Maldonado Ahumada</b>	<b>Secretario</b>	<b>IPICYT</b>
<b>Dr. Haret-Codratian Rosu Barbus</b>	<b>Sinodal</b>	<b>IPICYT</b>

a fin de efectuar el examen, que para obtener el Grado de:

**MAESTRO EN CONTROL Y SISTEMAS DINÁMICOS**

sustentó el C.

**Nazul Bonfilio Merino Negrete**

sobre la Tesis intitulada:

*Prueba de irreversibilidad en el análisis de Electrocardiogramas*

que se desarrolló bajo la dirección de

**Dr. Cesar Octavio Maldonado Ahumada**

El Jurado, después de deliberar, determinó

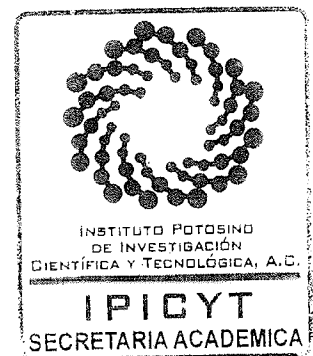
**APROBARLO**

Dándose por terminado el acto a las 17:35 horas, procediendo a la firma del Acta los integrantes del Jurado. Dando fe el Secretario Académico del Instituto.

A petición del interesado y para los fines que al mismo convengan, se extiende el presente documento en la ciudad de San Luis Potosí, S.L.P., México, a los 16 días del mes de agosto de 2018.

  
**Dr. Horacio Flores Zúñiga**  
Secretario Académico

  
**Mtra. Ivonne Lizette Cuevas Vélez**  
Jefa del Departamento del Posgrado



*Para mis adoradas sobrinas Hypatia Neith y  
Andoeni Yael...son mi continua fuente de  
inspiración.*

# Agradecimientos

De manera especial, agradezco al Dr. César Maldonado, por la confianza depositada en mi para realización de esta tesis, por sus enseñanzas, por su infinita paciencia.

Extiendo este agradecimiento a mis sinodales Dr. Raúl Salgado y Dr. Haret Rosu, quienes revisaron la presente tesis y aportaron sugerencias para mejorarla. Al CONACYT, por la beca 612806 otorgada y al departamento de Matemáticas Aplicadas del IPICYT, por abrirme las puertas de sus instalaciones para la realización de mis estudios de maestría.

A mis padres, María y Bonfilio, por todo el apoyo recibido. Al resto de mi familia: Osiris, Cuauhtemoc, Andoeni, Hypatia y Nefthalí, por el apoyo mutuo...que siempre sea así.

A mis apreciados compañeros: Elizabeth, Ivan, Carlos, Antonio y al Richie, por los momentos compartidos.

A todos los mencionados y en especial a algunos de los omitidos, mi testimonio de gratitud, la rosa blanca de Martí, la mano franca.

# Resumen

El desarrollo de métodos para el análisis de señales biomédicas ha sido objeto de estudio en las últimas décadas; en el área de las investigaciones biomédicas hay una necesidad constante por mejores herramientas de diagnóstico. En el presente trabajo se propone analizar electrocardiogramas mediante la producción de entropía en modelos markovianos. El objetivo es exhibir evidencia numérica de la diferencia en producción de entropía promedio entre pacientes sanos y pacientes con arritmia cardíaca.

Los electrocardiogramas analizados se obtuvieron de la base de datos PhysioBank. Las secuencias simbólicas se generaron a partir de tres particiones del espacio de estados y dos tiempos de muestreo. Las secuencias simbólicas permitieron calcular la probabilidad estacionaria de los estados, estimar las probabilidades de transición entre los estados y así obtener el modelo markoviano del sistema. La producción de entropía promedio en pacientes sanos y pacientes con arritmia se estimó a partir del modelo markoviano y a partir de un estimador de irreversibilidad empírico.

En general, la probabilidad empírica de los estados converge en un tiempo promedio de 15 minutos para los dos grupos de pacientes, a lo que suponemos es el vector de probabilidad estacionario. La probabilidad empírica de los estados, como función del número de datos, presenta mayor fluctuación en los pacientes con arritmia cardíaca que en los pacientes sanos. Respecto a la producción de entropía, los resultados numéricos exhiben una diferencia en la producción de entropía promedio entre los dos grupos de pacientes, teniendo una mayor producción de entropía los pacientes sanos. Los resultados respecto al comportamiento de la producción de entropía, fueron consistentes para los dos enfoques que adoptamos, a saber: el enfoque del modelo markoviano y el enfoque de un estimador de irreversibilidad empírico.

En este trabajo se presenta evidencia numérica de que la producción de entropía puede ser una cantidad relevante para distinguir entre electrocardiogramas generados por pacientes con arritmia y aquellos generados por pacientes sanos. Sin embargo, para dar resultados más contundentes, es necesario un estudio más exhaustivo, aumentando la estadística y restringiendo más las condiciones, además de proponer una forma de estandarización de los datos, entre otros.

# Abstract

Recently there has been an important development of the methods employed for the analysis of biomedical data, since there is an increasingly need for better diagnostic tools. In this manuscript, we propose analyze electrocardiograms by using Markov models. The purpose is to show numerical evidence in the mean entropy production between healthy patients and patients with cardiac arrhythmia.

The analyzed electrocardiograms were obtained from the PhysioBank database. From the time-series we have generated symbolic sequences using three different partitions of the state space and two sampling times, trying to encode the physiological meaning of the data. The symbolic sequences allowed calculating the stationary probability of the states, and estimating the transition probabilities, in this way the markovian model was obtained. The mean entropy production in healthy patients and in patients with arrhythmia was estimated from the inferred markovian model and using an empirical irreversibility estimator.

We observed, that, the empirical probability of the states generally converges in an average time of 15 minutes for the two groups of patients. We assume the empirical probability converges to the stationary probability vector. The empirical probability states exhibit greater fluctuations in patients with cardiac arrhythmia than in the healthy ones. Regarding the entropy production, the numerical results show a difference in the mean entropy production between both groups of patients, having a higher entropy production the healthy patients, as expected. The results obtained with the two approaches we adopted, namely: the approach of the Markov model and the empirical irreversibility estimator approach, are consistent.

In this thesis we show numerical evidence that entropy production might be a relevant quantity to distinguish between electrocardiograms generated by patients with cardiac arrhythmia and those generated by healthy patients. However, to give more conclusive results, a more comprehensive study is necessary, let say, by increasing the statistics and by further restricting the conditions, proposing a form of standardization for data, among others.



# Índice general

<b>1. Introducción</b>	<b>1</b>
<b>2. Electrocardiogramas</b>	<b>5</b>
2.1. ECG y Derivaciones . . . . .	5
2.2. ECG normales . . . . .	6
2.3. Caracterización de ECG . . . . .	8
2.4. Ritmo sinusal y arritmia cardíaca . . . . .	10
<b>3. Cadenas de Márkov</b>	<b>11</b>
3.1. Propiedad de Márkov . . . . .	11
3.2. Matriz de transición . . . . .	11
3.3. Probabilidades de transición de orden superior . . . . .	12
3.4. Algunas propiedades (irreducibilidad, recurrencia y estacionariedad) . . . . .	12
3.5. Ley de los grandes números para cadenas de Márkov . . . . .	13
3.6. Distribuciones estacionarias y distribución límite . . . . .	14
3.7. Cadenas reversibles . . . . .	15
<b>4. Representación simbólica de series de tiempo, entropía y producción de entropía</b>	<b>17</b>
4.1. Representación simbólica de series de tiempo . . . . .	17
4.2. Entropía en sistemas dinámicos simbólicos . . . . .	18
4.3. Producción de entropía en sistemas dinámicos simbólicos . . . . .	18
4.3.1. Ejemplo de proceso reversible: Cadena Ehrenfest . . . . .	20
4.3.2. Ejemplo de sistema irreversible: Ciclo de 3 estados . . . . .	21
<b>5. Modelo y Metodología</b>	<b>25</b>
5.1. Modelo . . . . .	25
5.2. Esquema general . . . . .	26
5.3. Adquisición de datos . . . . .	26
5.4. Partición del espacio de estados . . . . .	28
5.4.1. Partición del espacio de estados de los ECG completos . . . . .	29
5.4.2. Partición del espacio de estados de la línea isoelectrica . . . . .	30
5.5. Obtención de la línea isoelectrica . . . . .	32
5.6. Codificación simbólica, matriz de transición y distribución de probabilidad invariante . . . . .	33
5.7. Cálculo de la producción de entropía . . . . .	34
5.8. Cálculo de un estimador de irreversibilidad empírico . . . . .	35

<b>6. Resultados y discusión</b>	<b>37</b>
6.1. Modelo . . . . .	37
6.2. Probabilidad invariante . . . . .	38
6.3. Producción de entropía: modelo markoviano . . . . .	42
6.4. Producción de entropía: estimador de irreversibilidad . . . . .	42
<b>7. Conclusiones y trabajo futuro</b>	<b>45</b>
7.1. Conclusiones . . . . .	45
7.2. Trabajo futuro . . . . .	45
<b>Bibliografía</b>	<b>47</b>
<b>Apéndice A. Espacio de probabilidad, variable aleatoria y entropía</b>	<b>51</b>
A.1. Espacio de probabilidad . . . . .	51
A.2. Variable aleatoria . . . . .	51
<b>Apéndice B. Obtención de la línea isoelectrica</b>	<b>53</b>
<b>Apéndice C. Probabilidad estacionaria.</b>	<b>57</b>
<b>Apéndice D. Probabilidad y matriz de transición</b>	<b>59</b>
<b>Apéndice E. Estimación de la producción de entropía</b>	<b>65</b>
<b>Apéndice F. Estimador de irreversibilidad</b>	<b>69</b>

# Índice de figuras

2.1. Obtención de electrocardiogramas. . . . .	6
2.2. ECG normal real. . . . .	7
2.3. Diagrama del corazón. . . . .	7
2.4. Despolarización auricular y ventricular. . . . .	8
4.1. Cadena Ehrenfest. . . . .	20
4.2. Una realización de la Cadena Ehrenfest. . . . .	21
4.3. Grafo del ciclo de tres estados. . . . .	22
4.4. Realizaciones del ciclo de tres estados para $p=0.02$ , $p=1/2$ y $p=0.98$ . . . . .	23
5.1. Esquema general de la producción de entropía en ECGs. . . . .	27
5.2. Partición del espacio de estados del ECG completo. . . . .	30
5.3. Distribución de la línea isoeleétrica. . . . .	31
5.4. Partición del espacio de estados del la línea isoeleétrica. . . . .	32
5.5. Obtención de la línea isoeleétrica. . . . .	33
6.1. Grafo de un ECG. . . . .	38
6.2. Probabilidad empírica de los estados para los pacientes sanos. . . . .	39
6.3. Probabilidad empírica de los estados para los pacientes con arritmia. . . . .	40
6.4. Probabilidad empírica promedio de los estados para pacientes sanos y pa- cientes con arritmia. . . . .	41
6.5. Producción de entropía promedio para la partición 1. . . . .	43
6.6. Producción de entropía promedio en función de la línea isoeleétrica. . . . .	44



# Capítulo 1

## Introducción

Para caracterizar el equilibrio en sistemas modelados por cadenas de Márkov (CM), el *criterio de Kolmogórov* establece que la probabilidad de visitar los estados  $x_2, x_3, \dots, x_{n-1}, x_n, x_1$  iniciando en  $x_1$  es la misma que la de visitarlos en orden inverso  $x_n, x_{n-1}, \dots, x_2, x_1$  iniciando en  $x_1$ . Por otro lado, en [12], Gaspard estudia los estados estacionarios fuera de equilibrio (EEFE) en el contexto de las cadenas de Márkov. Los EEFE presentan «corrientes» de ciertas cantidades en el tiempo, esto es una manifestación de la irreversibilidad intrínseca del sistema. Así, uno desea cuantificar el grado de irreversibilidad en los sistemas fuera del equilibrio, y precisamente la cantidad definida para este propósito es la producción de entropía. De forma intuitiva, en EEFE existe una dirección preferencial de las trayectorias del sistema, esto es posible detectarlo con la diferencia en la frecuencia de aparición de trayectorias y la frecuencia de su trayectorias invertidas en el tiempo. Por consiguiente la positividad en la producción de entropía es una manifestación del rompimiento de la simetría bajo la inversión temporal [12, 20].

Gaspard propone que la *producción de entropía* en una CM está dada por la diferencia entre la *entropía de Kolmogórov* de la cadena y su *entropía de Kolmogórov* de la cadena reversa. Así pues, a un sistema fuera del equilibrio que puede ser descrito por una cadena de Márkov, se le puede calcular la producción de entropía en función de las probabilidades de tramos de trayectorias. Desde este punto de vista, las trayectorias de una CM están dadas por la representación simbólica del sistema en EEFE, que es una eficiente herramienta para el estudio de sistemas irreversibles [12].

Recientemente la dinámica simbólica se ha usado como una poderosa herramienta para el estudio de sistemas físicos y biológicos [31, 17, 8]. Con este enfoque es posible extraer y clasificar patrones en el análisis series de tiempo biomédicas (vea [1] y las referencias ahí mencionadas). Existen algunos trabajos en los que se ha propuesto diferentes métodos para el análisis de series de tiempo biomédicas empleando herramientas de dinámica simbólica, ver por ejemplo [4], [22] y [1]. En estos trabajos utilizan la entropía del sistema para cuantificar los cambios de complejidad dinámica en las series de tiempo. A saber, la entropía que utilizan, es una estimación de la entropía de Shannon.

Al analizar electrocardiogramas, el particular interés radica en la utilidad de la informa-

ción que proporciona para la detección, diagnóstico y tratamiento de enfermedades cardíacas [2]. Bajo premisa, uno de los trabajos reportados en la literatura, en los que se aplica herramientas de dinámica simbólica al análisis de electrocardiogramas, es [5]; donde Cammarota *et al* estudiaron las propiedades de inversión temporal de series de tiempo biomédicas mediante una codificación ternaria de las series de tiempos; este enfoque simbólico del modelo propuesto por los autores exhibió idoneidad para el modelado no lineal de los latidos normales del corazón. En [3] los autores obtuvieron medidas de complejidad no lineal de la variabilidad de la frecuencia cardíaca, en electrocardiogramas obtenidos de pacientes que sufren esquizofrenia aguda; concluyeron que la complejidad de la modulación de la frecuencia cardíaca se reduce significativamente en la esquizofrenia aguda no tratada, y por lo tanto aumenta el riesgo de sufrir eventos cardiovasculares. En [2] se presenta el enfoque de un modelo escondido de Márkov para la segmentación y clasificación de electrocardiogramas; la detección de latidos de contracción ventricular con una sensibilidad de 99,79% mediante el modelo propuesto, permitió a los autores validar su modelo para la aplicación en datos reales. En [35] los autores modelan la variabilidad del ritmo cardíaco mediante cadenas escondidas de Márkov; con su modelo los autores caracterizaron electrocardiogramas de pacientes con alto riesgo de muerte cardíaca súbita.

Por otra parte, los procesos biológicos, como la actividad eléctrica cardíaca, siguen leyes naturales que explícitamente «rompen» la simetría de inversión del proceso en el tiempo; los ECG exhiben un claro orden en el tiempo. En este sentido si asumimos que un ECG puede ser estudiado mediante una dinámica que tiene comportamiento markoviano, entonces deberíamos ver que la distribución de probabilidad de sus trayectorias difiere de la distribución de probabilidad de las trayectorias invertidas en el tiempo, dado que el sistema es irreversible. Cuantificar el grado de irreversibilidad de ECGs se convierte en un reto importante que puede ser abordado empleando herramientas de teoría de la información y mecánica estadística, mediante la producción de entropía del sistema, como mencionamos al principio. Esto representa el objetivo general de esta tesis; a saber, exhibir evidencia numérica de la diferencia en la producción de entropía promedio entre pacientes sanos y pacientes con arritmia cardíaca. Para lograr esto, asumimos un modelo markoviano, y estimamos las probabilidades de transición entre los estados. A partir de ahí, se puede calcular la producción de entropía directamente del proceso de Márkov estimado. Aunque también se puede estimar la producción de entropía directamente de las trayectorias, usando un estimador de irreversibilidad, que será definido en el capítulo 5.8. Esto representa una contribución clara al estado del arte, ya que hasta nuestro conocimiento, no contamos con antecedentes de trabajos en los que se proponga emplear a la producción de entropía como una alternativa para el análisis de electrocardiogramas.

El contenido de esta tesis está organizado como sigue. En el capítulo 1 se da una breve introducción al trabajo realizado y la motivación del mismo. En el capítulo 2 se presenta los conceptos básicos para el estudio e interpretación de electrocardiogramas. El concepto de cadenas de Márkov y sus principales propiedades se introducen en el capítulo 3. En el capítulo 4 se aborda dos temas esenciales para el desarrollo de la tesis: representación simbólica de series de tiempo reales y la producción de entropía en cadenas de Márkov. El modelo obtenido y el desarrollo de la metodología se describen en el capítulo 5. En el

capítulo 6 se discuten los resultados obtenidos. Finalmente, en el capítulo 7 se discute las principales conclusiones y el trabajo futuro. Además, en los apéndices se puede encontrar los códigos en Matlab usados para los cálculos y simulaciones numéricas realizadas en la tesis.





# Capítulo 2

## Electrocardiogramas

En este capítulo se da una breve introducción a los conceptos básicos para el estudio e interpretación de los Electrocardiogramas (ECG). Se aborda las principales características de un ECG normal, en términos de la duración y amplitud de las ondas que lo componen. En la sección final se define el ritmo sinusal y la arritmia cardíaca. Los temas presentados en éste capítulo pueden ser consultados en [9], [15], [23], [16].

### 2.1. ECG y Derivaciones

Un ECG es el registro de la actividad eléctrica del corazón, que tiene lugar cada vez que éste se contrae. El registro se obtiene colocando electrodos en determinadas zonas del cuerpo de un paciente. Mediante el uso de diversas combinaciones de estos electrodos y empleando un electrocardiógrafo, se observan doce vistas diferentes de la misma actividad eléctrica. Cada vista del corazón se denomina *derivación electrocardiográfica*. En las pruebas de rutina se usa un ECG de doce derivaciones, que consiste en seis Derivaciones del Plano Frontal o Derivaciones de las Extremidades (derivaciones I, II, III, aVR, aVL y aVF) y seis Derivaciones del Plano Horizontal o Derivaciones Precordiales (derivaciones V1, V2, V3, V4, V5 y V6); en la figura 2.1 (fuente [2], [38], [6]) se muestra un esquema general de la obtención de ECG. La actividad eléctrica del corazón se representa en un gráfico bidimensional denominado *electrocardiograma*, colocando en el eje de las abscisas el tiempo (en segundos) y en el eje de las ordenadas el potencial eléctrico (en mV). A las distintas curvaturas hacia arriba y hacia abajo que toma el trazado de un ECG se les denomina *ondas*. Cada una de estas ondas es producto de los potenciales de acción que se producen durante la estimulación cardíaca, y se repiten de forma periódica de un latido a otro, a menos que se presenten alteraciones. Las ondas electrocardiográficas han sido denominadas P, Q, R, S, T, U por ese orden y van unidas entre si por una *línea isoelectrica*, la cual se produce cuando no hay propagación eléctrica del corazón (ver figura 2.1). En las secciones 2.2 y 2.3 se da una descripción más específica del ciclo cardíaco y de cada componente de un ECG normal (para la derivación II), respectivamente.

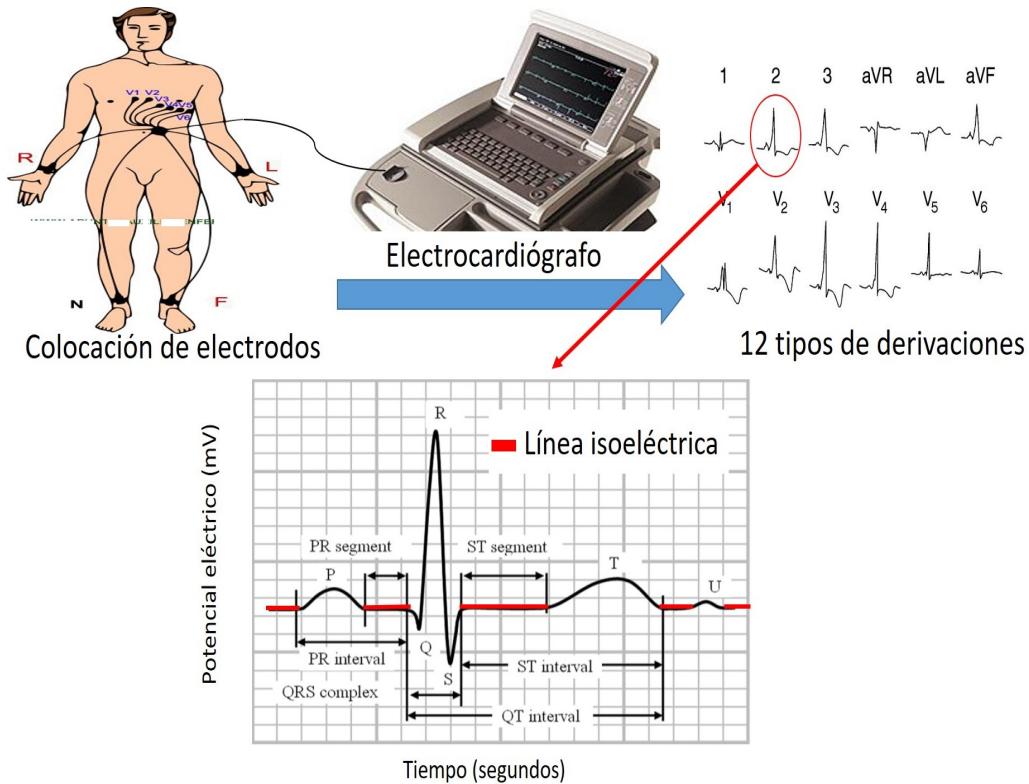


Figura 2.1: Obtención de electrocardiogramas.

## 2.2. ECG normales

Para la derivación II, un ECG normal consiste en una onda P, un complejo QRS, una onda T y una onda U; el complejo QRS está formado por las ondas Q, R y S, mientras que la pequeña onda U normalmente es invisible. En un ciclo cardíaco normal, el complejo QRS aparece como su patrón principal y más dominante. Antes del complejo QRS, existe una onda P, y después del complejo QRS está la onda T, la cual tiene una mayor duración que la onda P. Los segmentos «planos» entre los elementos mencionados anteriormente, son los segmentos PQ, ST y TP (ver figura 2.2, fuente [26]). Un intervalo RR proporcionan información útil para el estudio del ritmo cardíaco. Los parámetros más importantes para los cardiólogos son la duración y las amplitudes de los patrones antes mencionados. Cada una de las ondas en el ECG representa estímulos eléctricos de las aurículas y los ventrículos, como se describe a continuación.

En el corazón, la contracción de cualquier músculo se asocia con cambios eléctricos denominados «despolarización», que puede detectarse mediante electrodos unidos a la superficie corporal. Aunque el corazón tiene cuatro cámaras, desde un punto de vista eléctrico, puede considerarse que sólo tiene dos, porque ambas aurículas se contraen de manera conjunta (despolarización), al igual que ambos ventrículos (ver figura 2.3, fuente [24]). La masa muscular de las aurículas es pequeña en comparación con los de los ventrículos, por lo que la variación eléctrica que acompaña a la contracción auricular es menor que

la variación eléctrica debida a la contracción ventricular. La contracción de las aurículas se asocia con la onda P del ECG. La masa ventricular es grande, por lo que se produce una gran oscilación del ECG cuando los ventrículos se despolarizan: es lo que se denomina complejo QRS. La onda T del ECG se asocia con el retorno de la masa ventricular a su estado de reposo eléctrico («repolarización»). En la sección 2.3 se da una descripción, en términos de intervalos de tiempo y potencial eléctrico, de cada componente del ECG.

La descarga eléctrica necesaria de un ciclo cardíaco normal, se inicia en un área especial de la aurícula derecha denominada «nódulo sinoauricular (SA)» (ver figura 2.4, fuente [36]). A continuación, la despolarización se propaga por las fibras del músculo auricular. Después, la onda de despolarización viaja muy de prisa por un tejido de conducción especializado, denominada «has de His». Finalmente, la despolarización se propaga por el músculo ventricular, al mismo tiempo que se inicia la repolarización auricular. Seguida de la repolarización auricular, y cuando la despolarización ventricular se ha completado, se inicia la repolarización ventricular.

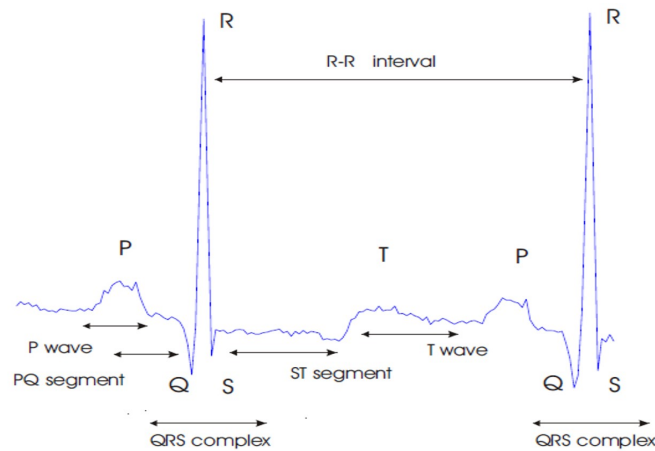


Figura 2.2: ECG normal real.

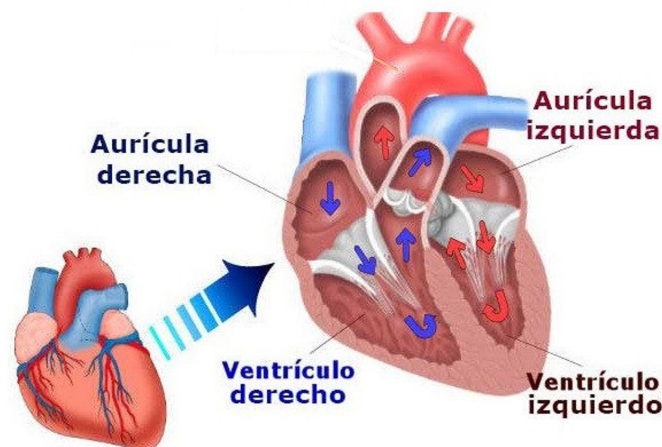


Figura 2.3: Diagrama del corazón.

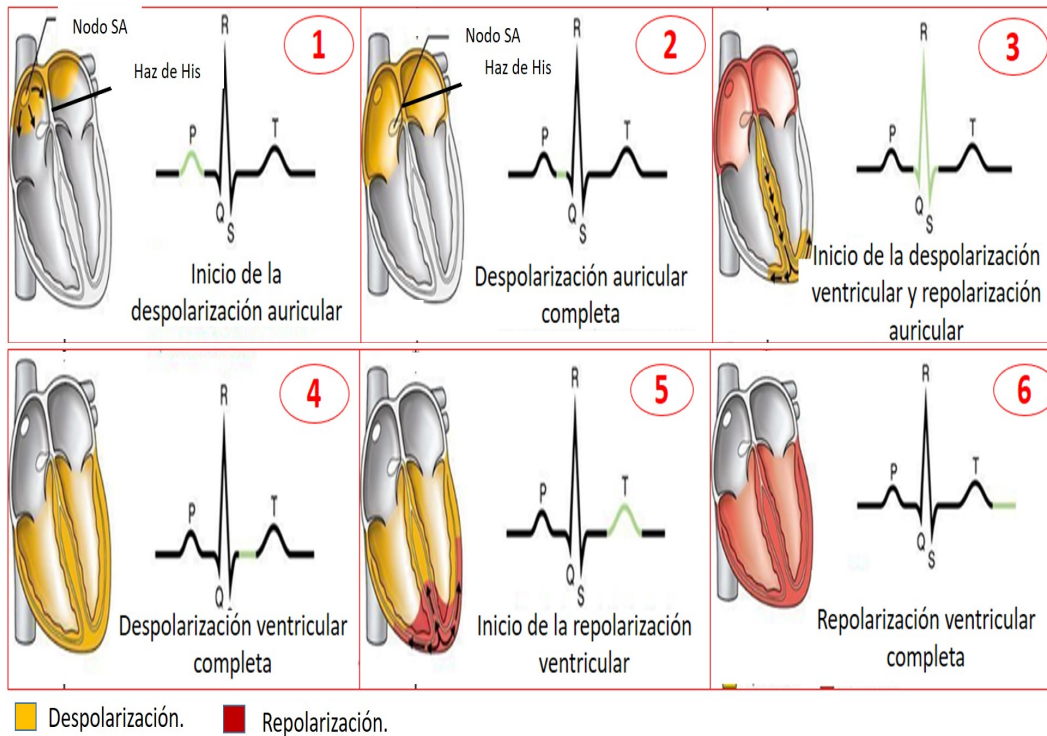


Figura 2.4: Despolarización auricular y ventricular.

### 2.3. Caracterización de ECG

Para que un ECG se considere normal, todas sus partes están bien definidas en ciertos intervalos de tiempo y potencial eléctrico. En esta sección se da una breve descripción de las distintas partes de un electrocardiograma normal, para la derivación II.

#### ■ Onda P

La *onda P* es la primera onda del ciclo cardíaco. Representa la despolarización de las aurículas. Su parte inicial corresponde a la despolarización de la aurícula derecha y su parte final a la de la aurícula izquierda. Para que la onda P se considere normal, debe reunir ciertas características:

- No debe superar los 0.25 mV
- La duración normal de una onda P es menor o igual a 100 ms.
- Tiene que ser redondeada, simétrica y de forma ovalada.

La mejor interpretación de una onda P se obtiene generalmente en las derivaciones II y V2.

#### ■ Onda Q

La *onda Q* es la primera onda del complejo QRS y tiene valores negativos. Las características de una onda Q normal son:

- La duración de la onda Q es menor o igual a 30 ms.
- La profundidad de la onda Q es menor o igual a 0.3 mV, y menor o igual a  $\frac{1}{4}$  de la amplitud máxima de la onda R.

La onda Q normalmente no se presenta en las derivaciones V1-V2.

#### ■ Complejo QRS

El *complejo QRS* aparece después de la onda P. Está formado por un conjunto de ondas que representan la despolarización de los ventrículos; sus ondas tienden a ser muy angostas y en forma de pico, en vez de redondeadas. Las ondas del complejo QRS toman varias morfologías dependiendo de la derivación.

- La duración de un complejo QRS normal varía entre 60 y 100 ms
- El voltaje normal del complejo QRS está entre 0.6 y 3.5 mV

#### ■ Onda T

Representa la repolarización de los ventrículos. En un ECG normal, es positiva en todas las derivaciones excepto en aVR, y es de menor amplitud que el QRS que le precede. La *onda T* normalmente es asimétrica, con una duración inferior a 200 ms y un potencial eléctrico que fluctúa entre 0.2 y 0.3 mV.

#### ■ Onda U

Es una onda habitualmente positiva, de bajo voltaje, que aparece sobre todo en derivaciones precordiales, inmediatamente detrás de la onda T.

#### ■ Intervalos

- Intervalo PQ

Representa la despolarización auricular y el retraso fisiológico que sufre el estímulo a su paso por el nodo Auriculoventricular. Se mide desde el inicio de la onda P hasta el inicio de la onda Q. Su duración normal es  $120ms - 200ms$ .

- Intervalo QT

Representa el conjunto de despolarización y la repolarización de los ventrículos. Su medida varía con la frecuencia cardíaca. Se mide desde el inicio de la onda Q hasta el final de la onda T (no se incluye la onda U cuando está presente). Su duración normal es  $340ms - 440ms$ .

- Intervalo QRS

Mide el tiempo total de despolarización ventricular. Se mide desde el comienzo de onda Q hasta el final de la onda S. Su duración normal es  $60ms - 100ms$ .

Por lo anterior, la duración de un ECG normal fluctúa entre 460 y 640 ms.

- **Intervalo RR**

El *intervalo RR* es la distancia entre dos ondas *R* sucesivas. En el *ritmo sinusal* este intervalo debe ser constante. Este intervalo se mide desde el inicio de una onda *R* hasta el inicio de la onda *R* siguiente, y su duración depende de la frecuencia cardíaca.

## 2.4. Ritmo sinusal y arritmia cardíaca

- **Frecuencia cardíaca**

La *frecuencia cardíaca* se obtiene calculando el número de intervalos RR por minuto y se expresa en latidos por minuto (lpm). La frecuencia cardíaca normal está entre 60 y 100 lpm.

- **Ritmo cardíaco**

Podemos definir el *Ritmo Cardíaco* como la sucesión de los complejos QRS en el tiempo que dura el mismo.

El *Ritmo Sinusal* es el ritmo normal del corazón. Para determinar si un ECG está en el Ritmo Sinusal normal debe tener las siguientes características:

- Onda P positiva en derivaciones inferiores (II, III, y aVF) y precordiales de V2 a V6 y negativa en aVR.
- Cada Onda P debe estar seguida por un *complejo QRS*.
- El intervalo RR debe ser constante.
- El Intervalo PR debe ser igual o mayor de 0.12 segundos.
- La Frecuencia Cardíaca debe estar entre 60 y 100 lpm.

- **Arritmia cardíaca**

Se define como *Arritmia Cardíaca* a la variación de la Ritmo Cardíaco que no se justifica por razones fisiológicas. Se llama **bradicardia** cuando la frecuencia es menor a 60 lpm y **taquicardia** cuando la frecuencia es mayor a 100 lpm.

# Capítulo 3

## Cadenas de Márkov

En este capítulo se da una breve introducción a las Cadenas de Márkov, haciendo énfasis en las principales propiedades que son de utilidad para el desarrollo de los siguientes capítulos. En la segunda parte del capítulo se presenta la ley de los grandes números para cadenas de Márkov, que es un resultado esencial para el estudio de series de tiempo biomédicas finitas, ya que justifica el uso del enfoque empirista usado en esta tesis. Por último, se define la reversibilidad en cadenas de Márkov.

### 3.1. Propiedad de Márkov

Sea  $(\Sigma, \mathcal{F}, \mathbb{P})$  un espacio de probabilidad. Un proceso *estocástico*, a tiempo discreto y con espacio de estados discreto  $\Omega$ , i.e.,  $T \subset \mathbb{N}$  y  $\Omega \subset \mathbb{N}$ , es una función  $X : T \times \Sigma \rightarrow \Omega$ , tal que para cada  $n \in T$ ,  $X(n, \cdot)$  es una variable aleatoria. En general, el parámetro  $n$  se interpreta como el tiempo.

Sea  $\{X_n : n \geq 0\}$  un proceso estocástico. Se dice que  $\{X_n : n \geq 0\}$  es un *proceso de Márkov* si, para toda  $n \geq 1$ , la distribución de probabilidad de  $X_{n+1}$  se determina a partir del estado  $X_n$  del proceso en el tiempo  $n$  y no depende de los valores pasados de  $X_k$  para  $k \leq n - 1$ . A los procesos que satisfacen la *propiedad de Márkov* se les conoce como *Cadenas de Márkov* [28, 34]. Es decir, una CM satisface que, para toda  $n \geq 1$  y toda  $i_0, i_1, \dots, i_n, j \in \Omega$  ([21]),

$$\mathbb{P}(X_{n+1} = j \mid X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j \mid X_n = i_n).$$

Una forma equivalente de expresar esta propiedad es poder calcular probabilidad de la distribución conjunta de las variables  $X_0, X_1, \dots, X_n$  de la siguiente forma:

$$\mathbb{P}(X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \mathbb{P}(X_n = i_n \mid X_{n-1} = i_{n-1}) \dots \mathbb{P}(X_1 = i_1 \mid X_0 = i_0) \mathbb{P}(X_0 = i_0)$$

### 3.2. Matriz de transición

Sean  $i$  y  $j$  dos estados de una cadena de Márkov. La evolución aleatoria de ésta se determina por la probabilidad

$$\mathbb{P}(X_n = j \mid X_0 = i)$$

la cual se denota por  $p_{i,j}^{(n)}$  y representa la *probabilidad de transición* del estado  $i$  al estado  $j$  en  $n$  pasos. Cuando  $p_{i,j}^{(1)}$  se escribe solamente  $p_{i,j}$ ; esta probabilidad se conoce como *probabilidad de transición en un paso*. Cuando  $p_{i,j}(n, n+1)$  no depende de  $n$  se dice que la cadena es *estacionaria*. A la matriz que contiene las *probabilidades de transición en un paso*, tal que  $p_{i,j} \geq 0$  para toda  $i, j \in \Omega$ , y se conoce como *matriz de transición* de la *Cadena de Márkov*:  $[P_{i,j}]_{i,j \in \Omega} = [\mathbb{P}(X_1 = j \mid X_0 = i)]_{i,j \in \Omega}$ . En particular, el estado inicial  $i$  es un número de fila en la matriz, mientras que el estado final  $j$  corresponde a el número de columna. Una *matriz de transición* es una matriz estocástica; los renglones de  $[P_{i,j}]_{i,j \in \Omega}$  satisfacen la condición  $\sum_{j \in \Omega} p_{i,j} = 1$  para toda  $i \in \Omega$ .

### 3.3. Probabilidades de transición de orden superior

Una matriz de transición  $[P_{i,j}]_{i,j \in \Omega}$  describe el comportamiento en un paso,  $\mathbb{P}(X_{n+1} = j \mid X_n = i)$ , de la CM. Más aún, la *matriz de transición* permite analizar el comportamiento a largo plazo de la CM; mediante ésta se puede conocer la *probabilidad de transición* en  $n$  pasos.

**Proposición 1** ([28, 33]). *La probabilidad de transición en  $n$  pasos,  $p_{i,j}^{(n)}$ , está dada por la  $(i, j)$  entrada de la  $n$ -ésima potencia de la matriz  $[P_{i,j}]_{i,j \in \Omega}$ , es decir*

$$p_{i,j}^{(n)} = [P_{i,j}^{(n)}]_{i,j \in \Omega} \quad (3.1)$$

Esta identidad es consecuencia de la ecuación de Chapman-Kolmogorov aplicada  $(n-1)$  veces. La *Ecuación de Chapman-Kolmogorov* establece para cualquier par de números enteros  $r$  y  $n$  tales que  $0 \leq r \leq n$  y para cualesquiera estados  $i, j, k \in \Omega$ , se cumple que  $p_{i,j}^{(n)} = \sum_k p_{i,k}^{(r)} p_{k,j}^{(n-r)}$ . Esta ecuación permite descomponer la probabilidad de pasar del estado  $i$  al  $j$  en  $n$  pasos, en la suma de las probabilidad de las trayectorias que van de  $i$  a  $j$ , y que atraviesan por un estado  $k$  cualquiera en un tiempo intermedio  $r$  ([28, 33]).

### 3.4. Algunas propiedades (irreducibilidad, recurrencia y estacionariedad)

**Definición 1 (Cadenas de Márkov irreducibles).** *Se dice que los estados  $i$  y  $j$  son comunicantes si existen  $m \geq 0$  y  $n \geq 0$  tales que  $p_{i,j}^{(m)} > 0$  y  $p_{j,i}^{(n)} > 0$ . Una cadena de Márkov es irreducible si  $\forall i, j \in \Omega$  existen  $m \geq 0$  y  $n \geq 0$  tales que  $p_{i,j}^{(m)} > 0$  y  $p_{j,i}^{(n)} > 0$ . Es decir, una cadena de Márkov es irreducible si todos los estados se comunican entre sí.*

**Definición 2 (Tiempo de primera visita).** *Sea  $\{X_n : n \geq 0\}$  un proceso de Márkov y  $A \subset \Omega$ . El tiempo de primera visita al conjunto  $A$  es la variable aleatoria*



$$\tau_A = \begin{cases} \text{mín}\{n \geq 1 : X_n \in A\} & \text{si } X_n \in A, \text{ para } n \geq 1 \\ \infty & \text{otro caso.} \end{cases}$$

Cuando el conjunto  $A$  consta de un sólo estado  $j$  y la cadena empieza en el estado  $i$ , el tiempo de primera visita al estado  $j$  se escribe  $\tau_{ij}$ .

**Definición 3 (Recurrencia y transitoridad).** Sea  $f_{ij}(n) = p(\tau_{ij} = n)$ . Se dice que un estado  $i$  es recurrente si  $f_{ii}(n) = 1$ ; es decir, el estado  $i$  es recurrente si y sólo si  $\sum_{n=1}^{\infty} p_{i,i}(n) = \infty$ . Análogamente, un estado  $i$  es transitorio si  $f_{ii}(n) < 1$ ; el estado  $i$  es transitorio si y sólo si  $\sum_{n=1}^{\infty} p_{i,i}(n) < \infty$ .

**Definición 4 (Tiempo medio de recurrencia).** El tiempo medio de recurrencia de un estado recurrente  $j$  a partir de un estado  $i$ , se define como la esperanza de  $\tau_{ij}$ , y se denota por  $\mu_{ij} = \mathbb{E}(\tau_{ij}) = \sum_{n=1}^{\infty} n f_{ij}(n)$ . Se dice que un estado  $i$  es recurrente positivo si  $\mu_i < \infty$ ; es recurrente nulo si  $\mu_i = \infty$ .

**Definición 5 (Periodo).** El periodo de un estado  $i$  es un entero no negativo denotado por  $d(i)$ , definido como  $d(i) = \text{m.c.d.}\{n \geq 1 : p_{i,i}(n) > 0\}$ , en donde m.c.d. significa «máximo común divisor». Cuando  $p_{i,i}(n) = 0$  para toda  $n \geq 1$ , se define  $d(i) = 1$ . En particular, se dice que un estado  $i$  es **aperiódico** si  $d(i) = 1$ . Cuando  $d(i) = k \geq 2$  se dice que  $i$  es periódico de periodo  $k$ .

### 3.5. Ley de los grandes números para cadenas de Márkov

La ley fuerte de los grandes números es un resultado clásico para variables aleatorias independientes e idénticamente distribuidas (vea [33] por ejemplo).

**Teorema 1 (Ley fuerte de los grandes números, [34]).** Sea  $(X_n)_{n \geq 1}$  una sucesión de variables aleatorias independientes idénticamente distribuidas, con  $\mathbb{E}(X_1) < \infty$ . Entonces

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = \mathbb{E}(X_1), \text{ casi seguramente} \quad (3.2)$$

donde  $\bar{X} \rightarrow \mathbb{E}(X_1)$  casi seguramente significa que  $\mathbb{P}(\lim_{n \rightarrow \infty} \bar{X} = \mathbb{E}(X_1)) = 1$

**Definición 6 (Número de visitas).** Considere la función indicadora

$$\mathbb{1}(X_k = j) = \begin{cases} 1 & \text{si } X_k = j \\ 0 & X_k \neq j \end{cases}$$

El número visitas que una cadena realiza sobre el estado  $j$ , empezando en el estado  $X_0 = i$ , en una sucesión de tamaño  $n$ , define la variable aleatoria:

$$N_j(n) := \sum_{k=1}^n \mathbb{1}(X_k = j) \quad (3.3)$$

Para una cadena de Márkov irreducible, el tiempo promedio que la cadena permanece en el estado  $j$  a largo plazo, está dado por

$$\frac{N_j(n)}{n} = \frac{1}{n} \sum_{k=1}^n \mathbb{1}(X_k = j) \quad (3.4)$$

Esta expresión es sólo el promedio de la muestra de una sucesión de variables aleatorias idénticamente distribuidas. Es posible usar la ley fuerte de los grandes números para deducir que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{1}(X_k = j) = \mathbb{E}(\mathbb{1}(X = j)) \quad (3.5)$$

Para una cadena de Márkov, la esperanza de una variable aleatoria es descrita mediante

$$\mathbb{E}(\mathbb{1}(X = j)) = \sum_{x=0}^m \mathbb{1}(x = j) \mathbb{P}(X = x) = \sum_{x=0}^m \mathbb{1}(x = j) \pi_x = \pi_j \quad (3.6)$$

esto es

$$\lim_{n \rightarrow \infty} \frac{N_j(n)}{n} = \pi_j \quad (3.7)$$

La ecuación 3.7 se conoce como **ley de los grandes números para cadenas de Márkov**. Esta ley nos dice que para sucesiones infinitas, el tiempo promedio, en el límite, que la cadena pasa en un estado  $j$ , está dada por la distribución estacionaria  $\pi_j$ .

### 3.6. Distribuciones estacionarias y distribución límite

La matriz de transición de una CM define la evolución de la dinámica markoviana del sistema. Al considerar un espacio de estados  $\Omega$  finito y una distribución de probabilidad inicial  $\pi^0 = (\pi_0^0, \pi_1^0, \dots, \pi_n^0)$ , es posible obtener una sucesión infinita de distribuciones de probabilidad,  $\{\pi^0, \pi^1, \pi^2, \dots, \pi^\infty\}$ , en donde la  $j$ -ésima entrada de cada  $k$ -ésimo vector de distribuciones de probabilidad se obtiene mediante  $\pi_j^k = \sum_{i=0}^n \pi_i^{k-1} p_{i,j}$ . En general, el vector  $\pi^k$  se obtiene a partir de  $\pi^{k-1}$  y de la matriz de transición  $[P_{i,j}]_{i,j \in \Omega}$ , mediante  $\pi^k = \pi^{k-1} P = \pi^0 P^k$ .

La distribución de probabilidad  $\pi = (\pi_0, \pi_1, \dots, \pi_n)$  es *una distribución estacionaria* si  $\pi_j = \sum_i \pi_i p_{i,j}$ . En términos matriciales, la distribución de probabilidad  $\pi$  es estacionaria si  $\pi = \pi P^n$ , para toda  $n \in \mathbb{N}$ . Consecuentemente, para el estudio de la dinámica representada por una CM, es importante saber si existe algún límite para esta sucesión de distribuciones de probabilidad. El hecho más importante sobre una CM es la existencia de una distribución de probabilidad límite[29, 30]; la *reversibilidad* de una CM está en función de la distribución límite[34].

**Definición 7** Considere una cadena de Márkov con matriz de transición  $P$  y distribución de probabilidad inicial  $\pi^0$ . Se llama distribución límite de la cadena al vector  $\pi = \lim_{n \rightarrow \infty} \pi^0 P^n = \lim_{n \rightarrow \infty} \sum_i \pi_i^0 p_{i,j}(n)$ .

El siguiente resultado establece condiciones suficientes para la existencia del límite de las probabilidades de transición cuando el número de pasos crece a infinito, asegurando además que el límite obtenido constituye una distribución de probabilidad estacionaria.

**Teorema 2** ([34]). Considere una cadena de Márkov que es

- irreducible
- recurrente positiva
- aperiódica

Entonces las probabilidades límite  $\pi_j = \lim_{n \rightarrow \infty} p_{i,j}(n)$ , existen, y están dadas por  $\pi_j = \frac{1}{\mu_j}$ , y constituye la única solución al sistema de ecuaciones

$$\pi_j = \sum_{i=0}^N \pi_i p_{i,j} \quad (3.8a)$$

$$\sum_{i=0}^N \pi_j = 1. \quad (3.8b)$$

donde  $\pi_j \geq 0$ . Es decir, la distribución límite  $\pi = (\pi_0, \pi_1, \dots, \pi_N)$  es la solución única no negativa del sistema.

### 3.7. Cadenas reversibles

Sea  $\{X_n : n \geq 0\}$  una CM con probabilidades de transición  $p_{i,j}$ , y sea  $Y_n = X_{m-n}$  para  $0 \leq n \leq m$ . Entonces  $\{Y_n : n = 0, 1, \dots, m\}$ , que es la cadena original pero vista en el sentido inverso en el tiempo, tiene la *propiedad de Márkov*, pero las probabilidades de transición del nuevo proceso no son necesariamente homogéneas, pues para  $0 \leq n < m$  [33, 30, 29],

$$P(Y_{n+1} = j | Y_n = i) = p_{i,j} \frac{P(Y_{n+1} = j)}{P(Y_n = i)}$$

por lo tanto, estas probabilidades dependen de  $n$ . Para que tal dependencia desaparezca se requiere que exista una distribución estacionaria  $\pi$  para  $\{X_n : n \geq 0\}$ . En tal caso, la igualdad anterior se reduce a

$$P(Y_{n+1} = j | Y_n = i) = p_{i,j} \frac{\pi_j}{\pi_i}$$

Bajo esta hipótesis, las probabilidades de transición de la nueva cadena son ahora estacionarias. Esto lleva a la siguiente definición de reversibilidad.

**Definición 8** ([30, 29, 21]). Se dice que una cadena de Márkov con probabilidades de transición  $p_{i,j}$  y con distribución estacionaria  $\pi$  es reversible en el tiempo si para cualesquiera estados  $i, j \in \Omega$ ,

$$\pi_i p_{i,j} = \pi_j p_{j,i} \tag{3.9}$$

Esta ecuación se conoce como *ecuación de balance detallado*.

# Capítulo 4

## Representación simbólica de series de tiempo, entropía y producción de entropía

### 4.1. Representación simbólica de series de tiempo

**Definición 9 (Sistema dinámico).** Dado un espacio  $\Omega$  y una transformación  $f$  definida en  $\Omega$ , al par  $(\Omega, f)$  le llamaremos sistema dinámico. Dado  $\omega_0 \in \Omega$ , el conjunto  $\vartheta_f^+(\omega_0) = \{\omega_0, f(\omega_0), f^2(\omega_0), \dots\}$  recibe el nombre de **órbita** ó **trayectoria** de  $\omega_0$ . Al primer punto de la órbita le llamaremos condición inicial.

**Definición 10 (Sistema dinámico simbólico).** Dado un conjunto finito  $A = \{1, 2, \dots, n\}$ , definimos un espacio de secuencias mediante  $\Sigma_n = A^{\mathbb{N}}$ . Sea  $\omega = (\omega_i)_{i=1}^{\infty} \in \Sigma_n$ , la función  $\sigma : \Sigma_n \rightarrow \Sigma_n$  definida como  $\sigma(x_i) = x_{i+1}$  recibe el nombre de **shift map**. Al par  $(\Sigma_n, \sigma)$  le llamaremos **sistema dinámico simbólico**.

Para analizar de series de tiempo de datos reales, se puede generar una representación simbólica del sistema y hacer uso de las técnicas de estudio de los sistemas dinámicos simbólicos. Para que una representación simbólica sea útil, debe ser más simple de analizar que la serie de tiempo original y al mismo tiempo debe preservar la información de interés [1].

Considérese la dinámica generada por el mapeo de  $f$  aplicada a un elemento  $\omega_0 \in \Omega$ . En la trayectoria de  $\omega_0$ , cada aplicación de  $f$  en el punto  $\omega_n = f(\omega_{n-1}) \in \Omega$  define el estado actual del sistema [1]. Para el problema inverso, cuando la dinámica del sistema es complicada y no se tiene la  $f$  que genera tal dinámica, es posible obtener «imágenes» simplificadas del sistema. Es decir, suponiendo que tenemos una serie de tiempo, deseamos construir la función  $f$  que genera la dinámica. Asumimos que podemos particionar al espacio de estados del sistema un número finito de partes disjuntas  $A_i$ , para  $i = 0, 1, 2, \dots, k-1$ , y se da seguimiento de la trayectoria de  $\omega \in \Omega$  (nuestra condición inicial en los datos), mediante la partición  $\alpha = \{A_0, \dots, A_{k-1}\}$ . Específicamente, a cada  $\omega \in \Omega$  se le asigna una secuencia [1]:

$$\mathbf{S}(\omega) = (S_n(\omega))_{n \in \mathbb{N}} = (i_0, i_1, \dots, i_n, \dots),$$

la  $n$ -ésima entrada  $i_n \in \{0, 1, \dots, N-1\}$  indica en cual elemento de  $\alpha$  se encuentra la iteración  $f^n(\omega)$ .  $(S_n(\omega))_{n \in \mathbb{N}}$  se llama el itinerario de  $\omega$  respecto a la partición  $\alpha$ . Formalmente,  $S_n(\omega) = i_n$  si y sólo si  $f^n(\omega) \in A_{i_n}$ , para  $n \in \mathbb{N}_0$ . Es importante hacer hincapié en el hecho de que  $\mathbf{S}(\omega)$  depende, en general, de la partición elegida  $\alpha$ ; para enfatizar esta dependencia se escribe  $(S_n^\alpha(\omega))_{n \in \mathbb{N}}$ .

## 4.2. Entropía en sistemas dinámicos simbólicos

Dada una partición  $\alpha = \{A_0, A_1, \dots, A_{k-1}\}$  del espacio de estados, y el itinerario  $(S_n(\omega))_{n \in \mathbb{N}}$  generado por  $\alpha$ , el mapeo  $S_n^\alpha: \Omega \rightarrow \{0, 1, \dots, k-1\}$ , define un proceso estocástico estacionario  $(S_n^\alpha)_{n \in \mathbb{N}}$ , denominado representación simbólica de  $f$  con respecto a la partición  $\alpha$ , para la cual, sus elementos  $A_i$ ,  $1 \leq i \leq N-1$ , deben ser medibles [1].

La probabilidad conjunta de observar el sistema en estados sucesivos  $\omega_0 \omega_1 \dots \omega_{n-1}$  está dada por [1, 12]:

$$\mathbb{P}(\omega) = p(\omega_0, \omega_1, \dots, \omega_{n-1}) = \mu(A_{\omega_0} \cap f(A_{\omega_1}) \cap \dots \cap f^{n-1}(A_{\omega_{n-1}})) \quad (4.1)$$

donde  $\mu$  es una medida de probabilidad invariante bajo  $f$ .

Por otro lado, la entropía de Kolmogorov-Sinai de  $f$  con respecto a la partición  $\alpha$ , que está definida por,

$$h_\mu(f, \alpha) = - \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\omega} \mu(A_{\omega_0} \cap f(A_{\omega_1}) \cap \dots \cap f^{n-1}(A_{\omega_{n-1}})) \log \mu(A_{\omega_0} \cap f(A_{\omega_1}) \cap \dots \cap f^{n-1}(A_{\omega_{n-1}}))$$

y la entropía de Shannon del proceso estocástico  $\mathbf{S}^\alpha$ , dada por,

$$h(\mathbf{S}^\alpha) = - \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\omega} p(\omega_0 \omega_1 \dots \omega_{n-1}) \log p(\omega_0 \omega_1 \dots \omega_{n-1});$$

coinciden, i.e.,

$$h_\mu(f, \alpha) = h(\mathbf{S}^\alpha), \quad (4.2)$$

para el caso markoviano [1], debido a la ecuación 4.1.

## 4.3. Producción de entropía en sistemas dinámicos simbólicos

En física, se dice que un sistema está en equilibrio si satisface la condición de balance detallado (ecuación 3.9), la cual describe la relación entre el proceso de transición de un estado a otro y el proceso de transición inverso. En matemáticas, en particular para sistemas markovianos, se tiene el *criterio de Kolmogórov*, el cual establece que la probabilidad de visitar los estados  $x_2, x_3, \dots, x_{n-1}, x_n, x_1$  iniciando en  $x_1$  es la misma que la de visitarlos en orden inverso  $x_n, x_{n-1}, \dots, x_2, x_1$  iniciando en  $x_1$ .

Por otro lado, si la condición de *balance detallado* caracteriza el equilibrio, es natural preguntarse, ¿qué características debe tener un sistema fuera de equilibrio?. En [12], Gaspard describe la irreversibilidad en estados estacionarios fuera de equilibrio (EEFE) en el caso de cadenas de Márkov, mediante la producción de entropía del sistema, como se describe a continuación.

Cuando tenemos una partición  $\alpha$ , naturalmente podemos definir las probabilidades transición  $P(\omega | \omega')$  de una CM, donde  $\omega, \omega' \in \Omega$ . La probabilidad conjunta de las trayectorias  $\omega = \omega_0, \omega_1 \dots \omega_{n-1}$  de una CM está dada por

$$p(\omega) = p(\omega_0, \omega_1 \dots \omega_{n-1}) = p(\omega_0)P(\omega_0 | \omega_1)P(\omega_1 | \omega_2) \dots P(\omega_{n-2} | \omega_{n-1})$$

debido a la propiedad de Márkov. La entropía de Kolmogórov-Sinai de la CM correspondiente está dada por

$$h = \lim_{n \rightarrow \infty} \frac{1}{n} H_n = \lim_{n \rightarrow \infty} -\frac{1}{n} \sum_{\omega \omega'} p(\omega) P(\omega | \omega') \ln P(\omega | \omega')$$

donde  $H_n$  es la entropía de Shannon, ver [12].

Para comparar  $h$  con las propiedades de las trayectorias invertidas en el tiempo, considérese la trayectoria invertida en el tiempo  $\omega^R = \omega_{n-1}, \omega_1 \dots \omega_0$ . En este caso, se está interesado en la probabilidad  $p(\omega^R) = p(\omega_{n-1} \dots \omega_1, \omega_0)$  de ocurrencia de la trayectoria invertida en el tiempo en un proceso estacionario fuera de equilibrio. La entropía invertida en el tiempo de Kolmogórov-Sinai de la correspondiente CM invertida en el tiempo, está dada por

$$h^R = \lim_{n \rightarrow \infty} \frac{1}{n} H_n^R = \lim_{n \rightarrow \infty} -\frac{1}{n} \sum_{\omega \omega'} p(\omega) P(\omega | \omega') \ln P(\omega' | \omega)$$

donde  $H_n^R$  es la entropía de Shannon de la CM invertida en el tiempo, ver [12]. En [12], Gaspard da una expresión para calcular la *producción de entropía* de una CM en estado estacionario en función de las trayectorias, y debido a la *propiedad de Márkov* se puede expresar como:

$$\Delta S = h^R - h = \frac{1}{2} \sum_{\omega \omega'} [p(\omega)P(\omega | \omega') - p(\omega')P(\omega' | \omega)] \ln \frac{p(\omega)P(\omega | \omega')}{p(\omega')P(\omega' | \omega)} \geq 0, \quad (4.3)$$

Dado que en EEFE, la distribución de probabilidad da un «peso» diferente a las «trayectorias hacia adelante» y las «trayectorias hacia atrás», la ecuación (4.3) define la producción de entropía como la diferencia entre la entropía de Kolmogórov-Sinai de las trayectorias de la CM invertida en el tiempo y la entropía de Kolmogórov-Sinai de las trayectorias de la CM.

La entropía de Kolmogórov-Sinai de las trayectorias de la CM cuantifica la «aleatoriedad dinámica» de las trayectorias de un proceso markoviano. Por otro lado, la entropía de

Kolmogórov-Sinai de las trayectorias de la CM invertida en el tiempo cuantifica la «aleatoriedad dinámica» de las «trayectorias hacia atrás» de la CM. Por lo tanto, fuera del equilibrio, la diferencia entre la aleatoriedad dinámica de las trayectorias y la aleatoriedad dinámica de las trayectorias invertidas en el tiempo de una CM, resultan estar relacionadas con la producción de entropía. Estos resultados proveen una interpretación de la producción de entropía como una manifestación del rompimiento de la simetría de la aleatoriedad dinámica de un sistema debido a la inversión en el tiempo [12]. Para dar una mayor interpretación de la cantidad definida por la ecuación 4.3, a continuación incluimos un par de ejemplos, uno de un sistema reversible (cadena Ehrenfest) y otro de un sistema irreversible (ciclo de 3 estados).

### 4.3.1. Ejemplo de proceso reversible: Cadena Ehrenfest

Un ejemplo clásico en de cadenas de Márkov reversibles es la cadena de Ehrenfest (ver figura 4.1, fuente [29]), descrita a continuación.

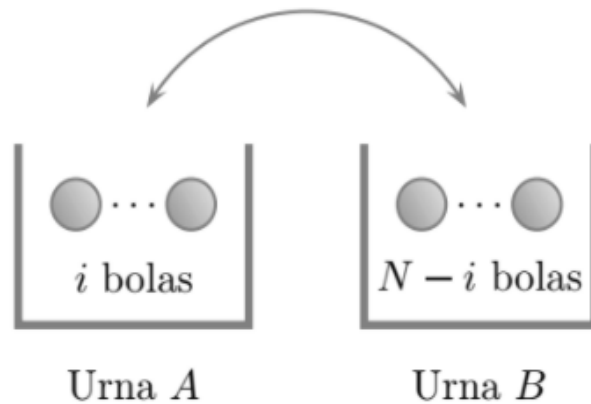


Figura 4.1: Cadena Ehrenfest.

Sean A y B dos urnas dentro de las cuales se encuentran distribuidas un total de  $N$  bolas de acuerdo a cierta configuración inicial. Por ejemplo, en la urna A hay  $i$  bolas y en la urna B hay  $N - i$  bolas. En cada unidad de tiempo se elige una bola al azar y cambia de urna. Sea  $X_n$  el número de bolas en la urna A después de  $n$  extracciones. Entonces  $\{X_n : n = 0, 1, 2, \dots, N\}$  es una cadena de Márkov, con probabilidad de transición

$$\mathbb{P}(X_{n+1} = j \mid X_n = i) = \begin{cases} \frac{i}{N} & \text{si } j = i - 1 \\ \frac{N-i}{N} & \text{si } j = i + 1 \end{cases}$$

y matriz de transición



$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 1/N & 0 & (N-1)/N & 0 & \cdots & 0 & 0 \\ 0 & 2/N & 0 & (N-2)/N & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1/N \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}.$$

A partir de la matriz de transición  $P$ , es posible calcular la distribución de probabilidad invariante  $\pi = (\pi_1, \pi_2, \dots, \pi_N)$  del sistema (ver teorema 2), y haciendo los cálculos de la ecuación (4.3), se obtiene:

$$\Delta S(\text{Ehrenfest}) = 0.$$

Por lo tanto, la cadena Ehrenfest es un proceso markoviano reversible en el tiempo. Intuitivamente, si invertimos en el tiempo la trayectoria de una cadena Ehrenfest, observaríamos el mismo patrón de comportamiento; es decir, no podríamos distinguir entre la trayectoria original y la trayectoria invertida en el tiempo. En la figura 4.2 se muestra la simulación de una cadena Ehrenfest, para 13 estados ( $\Omega = 0, 1, \dots, 12$ ). El tiempo de simulación mostrado es arbitrario, puesto que la cadena Ehrenfest exhibe el mismo comportamiento en cualquier intervalo de tiempo.

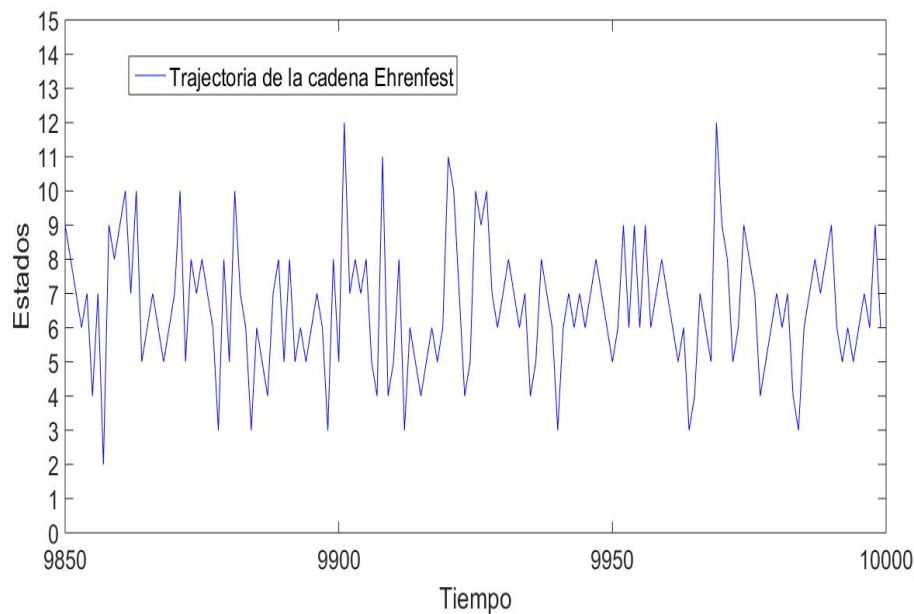


Figura 4.2: Una realización de la Cadena Ehrenfest.

### 4.3.2. Ejemplo de sistema irreversible: Ciclo de 3 estados

El ejemplo más sencillo de un sistema (CM) que es irreversible es el ciclo con tres estados.

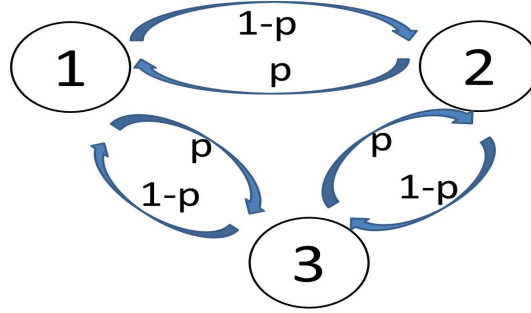


Figura 4.3: Grafo del ciclo de tres estados.

Considere un ciclo de tres estados ( $\Omega = \{1, 2, 3\}$ ), con probabilidad de transición

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

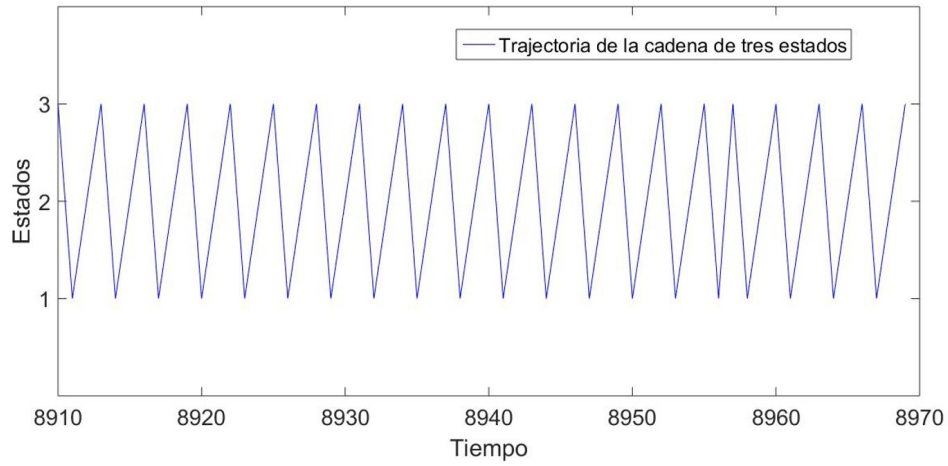
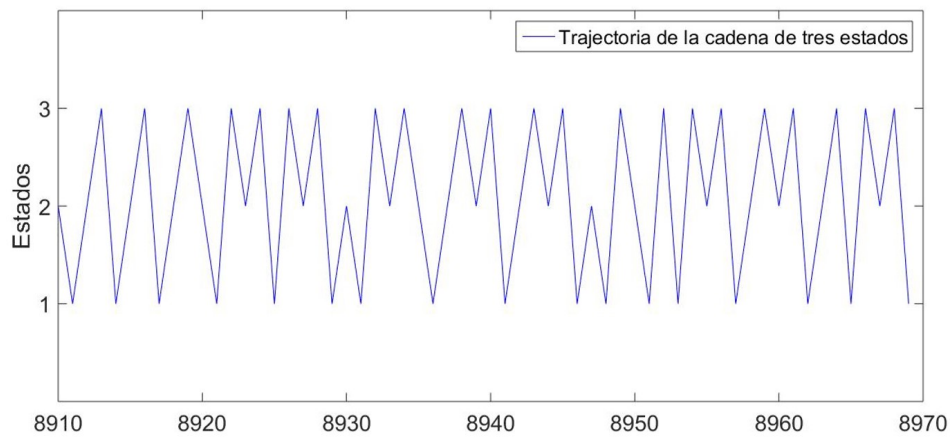
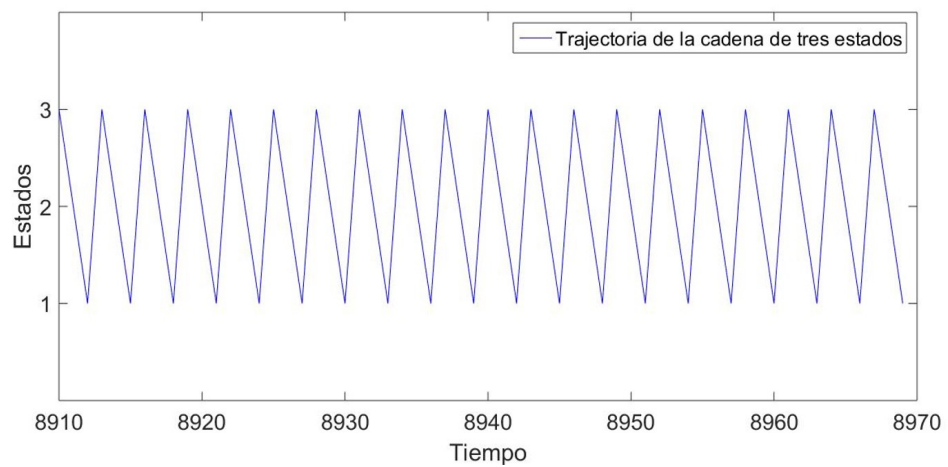
y matriz de transición

$$P = \begin{bmatrix} 0 & 1-p & p \\ p & 0 & 1-p \\ 1-p & p & 0 \end{bmatrix}.$$

A partir de la matriz de transición  $P$ , es posible calcular la distribución de probabilidad invariante  $\pi = (\pi_1, \pi_2, \dots, \pi_N)$  del sistema (ver teorema 2), y haciendo los cálculos de la ecuación (4.3), se obtiene:

$$\Delta S(\text{Ciclo3edos}) = \frac{p}{1-p}, \quad \text{para toda } p \in [0, 1].$$

Por lo tanto, la cadena de tres estados es irreversible para toda  $p \neq \frac{1}{2}$ . En la figura 4.4 se muestra la simulación del ciclo para  $p = 0,02$ ,  $p = 1/2$  y  $p = 0,98$ . Como se mencionó antes, de forma intuitiva, la irreversibilidad de un sistema puede describirse como «el rompimiento de la simetría de la aleatoriedad dinámica debido a la inversión en el tiempo del sistema»; es decir, si al invertir en el tiempo la trayectoria de un sistema, podemos «distinguir el cambio» en las trayectorias, entonces el sistema está en estado irreversible, lo cual, en el caso del ciclo de tres estados, ocurre para  $p \neq 1/2$  (ver figuras 4.4a y 4.4c). Por otro lado, cuando el sistema es reversible en el tiempo, no sería posible distinguir un cambio en la trayectoria del sistema, si esta es vista en el sentido inverso del tiempo (ver figura 4.4b). El tiempo de simulación mostrado en la figura 4.4 es arbitrario, puesto que el ciclo exhibe el mismo comportamiento en cualquier intervalo de tiempo.

(a)  $p=0.02$ (b)  $p=1/2$ (c)  $p=0.98$ Figura 4.4: Realizaciones del ciclo de tres estados para  $p=0.02$ ,  $p=1/2$  y  $p=0.98$ .



# Capítulo 5

## Modelo y Metodología

### 5.1. Modelo

A continuación mostramos las suposiciones del modelo markoviano obtenido:

1. Los electrocardiogramas pueden ser modelados como procesos markovianos. Diferentes autores han obtenido resultados que validan la aplicación de procesos markovianos al análisis de ECG, ver [2], [27], [35] y [18]. En la tabla (5.1) se muestra el valor de los parámetros empleados para las suposiciones del modelo markoviano obtenido.
2. Las cadenas de Márkov obtenidas para cada ECG son estacionarias.
3. Las cadenas de Márkov obtenidas para cada ECG son irreducibles.
4. Los estados de las cadenas de Márkov obtenidas para cada ECG, son recurrentes.
5. Los estados de las cadenas de Márkov obtenidas para cada ECG, son recurrentes positivos.
6. El tamaño de las secuencias simbólicas obtenidas es suficientemente grande, tal que se tiene la convergencia de la ley de los grandes números para cadenas de Márkov.

Suposición	Valor del parámetro
1	
2	
3	$p_{ij} = 0,000001$
4	Verdadero
5	Verdadero
6	$n \geq 450000$

Tabla 5.1: Suposiciones y condiciones del modelo markoviano.

## 5.2. Esquema general

Los procesos biológicos, como la actividad eléctrica cardíaca, siguen leyes naturales que explícitamente «rompen» la simetría de inversión del proceso en el tiempo; los ECG exhiben un claro orden en el tiempo. En este sentido si asumimos que un ECG puede ser estudiado mediante una dinámica que tiene comportamiento markoviano, entonces deberíamos ver que la distribución de probabilidad de su trayectoria difiere de la distribución de probabilidad de la trayectoria invertida en el tiempo. Cuantificar el grado de irreversibilidad de ECGs se convierte en un reto importante, lo cual puede ser abordado empleando herramientas de teoría de la información y mecánica estadística, mediante la producción de entropía del sistema.

Cuantificar el grado de irreversibilidad en ECGs es un proceso integrado por 4 etapas principales, a saber: adquisición de datos, codificación simbólica, construcción del modelo markoviano y estimación de la producción de entropía. En la figura 5.1 se muestra un esquema general de la metodología desarrollada; en las secciones posteriores de este capítulo, se describe de manera sistemática cada una de las etapas que se mencionan en la descripción del esquema. En los ECGs, es decir las series de tiempo biomédicas, asumiendo que pueden ser modeladas por sistemas dinámicos markovianos, es posible encontrar, resolviendo el problema inverso, las secuencias simbólicas que representan al sistema. Con esto se construye el modelo markoviano, como se describe a continuación.

Dado el vector de datos reales  $(x_n)_{n=1}^N$ , se obtiene una muestra  $\mathbf{u}^T$  del ECG. A partir de la muestra  $\mathbf{u}^T$ , se escribe la secuencia simbólica  $\mathbf{w}^{T_2, \alpha} = [\omega_1, \omega_2, \dots, \omega_n]$  del ECG, en función de la frecuencia de muestreo  $T_2$  y la partición  $\alpha = [A_1, A_2, \dots, A_k]$  del espacio de estados. La partición del espacio de estados de un ECG se puede proponer considerando el valor teórico del potencial eléctrico de cada onda en un ECG; no obstante, diferentes autores proponen hacer la partición del espacio de estados a partir de un análisis estadístico. Mediante  $\mathbf{w}^{T_2, \alpha} = [\omega_1, \omega_2, \dots, \omega_n]$ , se construye el modelo markoviano del sistema. La construcción del modelo markoviano implica la estimación de la probabilidad de transición de los estados  $\mathbb{P}(\omega_m = j \mid \omega_{m-1} = i)$ , de la matriz de transición  $[P_{i,j}]_{i,j \in \alpha}$  y de la distribución de probabilidad invariante de los estados  $\mathbb{P}(\omega_j)$ . Finalmente, se calcula producción de entropía del sistema  $\Delta S$ , en función de la matriz de transición y la distribución de probabilidad invariante. Para obtener la evolución en el tiempo de la producción de entropía del sistema, la secuencia simbólica permite calcular la producción de entropía al instante  $n$ .

## 5.3. Adquisición de datos

Los datos se obtuvieron de la base de datos *PhysioBank*, ver [14]. Se seleccionó 2 grupos de pacientes: 40 pacientes sanos y 40 pacientes con arritmia cardíaca. Respecto a los datos de los 40 pacientes sanos, 20 se obtuvieron de «FANTASIA Database» ([19]) y 20 de «Combined Measurement of ECG, Breathing and Seismocardiogram (CEBS Database)» ([11], [10]);

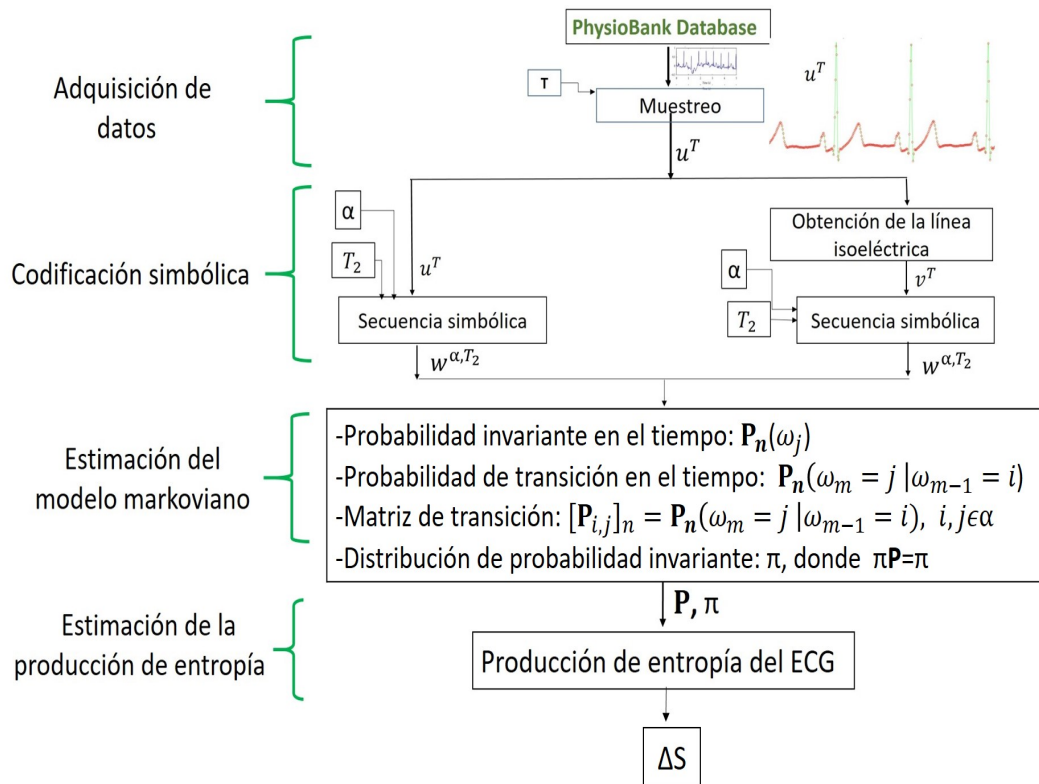


Figura 5.1: Esquema general de la producción de entropía en ECGs.

los datos de los 40 pacientes con arritmia, 10 se obtuvieron de «Massachusetts Institute of Technology and Boston's Beth Israel Hospital Arrhythmia Database (MIT-BIH Arrhythmia Database)» ([25]) y 30 de «The Massachusetts General Hospital/Marquette Foundation Waveform (MGH/MF Database)» ([37]). Para el lector interesado, en cada una de estas bases de datos aparece información completa y específica sobre las condiciones de adquisición de los datos (especificaciones del equipo utilizado para la adquisición de datos, condiciones de los pacientes, diagnóstico médico, etc).

La adquisición de los datos se hizo empleando el entorno de desarrollo "The WFDB Toolbox for MATLAB and Octave", utilizando el comando *rdsamp*, ver [32] y [13]. El «WFDB Toolbox» es una colección de funciones para leer, escribir y procesar señales fisiológicas y series de tiempo en el formato usado por *PhysioBank*.

En un mismo paciente, su ECG puede variar, principalmente, en función de la edad, tipo de derivación, sexo y estado de salud. Para poder comparar la producción de entropía entre pacientes sanos y pacientes con arritmia, se trabajó con ECGs obtenidos bajo las mismas condiciones en función de los factores de variabilidad mostrados en la tabla 5.2. Adicionalmente, como consecuencia de las frecuencias de muestreo empleadas por las distintas bases de datos (en la tabla 5.2 aparecen como «muestras por segundo»), se buscó estandarizar el número de muestras por segundo a analizar, empleando las frecuencias de muestreo  $T_1$  y  $T_2$  mostrada en la misma tabla. También aparecen en la tabla 5.2 el número

de muestras por segundo con las que se trabajó, obtenidas empleando las frecuencias de muestreo  $T_1$  y  $T_2$ .

Factores de variabilidad	Pacientes sanos		Pacientes con arritmia	
	FANTASIA	CEBS	MIT-BIH	MGH/MF
Derivación	II	II	II	II
Edad promedio	76.5	Adultos	65.4	69.45
Muestras por segundo	250	5000	360	360
Tiempo de muestreo (minutos)	30	30	30	30
Sexo	H/M	H/M	H/M	H/M
Frecuencia de muestreo				
$T_1$	2	40	3	3
$T_2$	5	100	6	6
Muestras por segundo				
Muestras obtenidas para $T_1$	125	125	120	120
Muestras obtenidas para $T_2$	50	50	60	60

Tabla 5.2: Factores de variabilidad, frecuencia de muestreo y muestras por segundo.

## 5.4. Partición del espacio de estados

Al trabajar con la representación simbólica de ECGs, se pierde cierta información detallada del sistema, pero la secuencia simbólica obtenida, al preservar la información acerca de los parámetros de interés del sistema (amplitud y frecuencia de las ondas del ECG, descritos en la sección 2.2), permite usar el conocimiento que se tiene acerca de las propiedades de sistemas dinámicos simbólicos en el contexto de las series de tiempo.

Se aplicaron dos enfoques para la partición del espacio de estados; el primero fue trabajar con los ECGs completos y el segundo fue analizar la producción de entropía restringiéndonos de la línea isoeletrica. El criterio para emplear estos dos enfoques de partición del espacio de estados, se basa en los resultados obtenidos en las distintas pruebas realizadas, seleccionando las particiones que más contribuyeron al propósito de exhibir la diferencia en la producción de entropía promedio entre pacientes sanos y pacientes con arritmia cardíaca. De las pruebas realizadas, se seleccionaron dos particiones para el espacio de estados del ECG completo y una para el espacio de estados de la línea isoeletrica.



### 5.4.1. Partición del espacio de estados de los ECG completos

El objetivo de las particiones propuestas fue generar las secuencias simbólicas que permiten caracterizar la variabilidad del ritmo cardíaco a partir de las transiciones entre los estados, con el propósito de distinguir si los ECGs pertenecen a un paciente sano o un paciente con arritmia. Adicionalmente, mediante la probabilidad empírica de cada estado en la partición, se puede observar la convergencia a la probabilidad estacionaria de los estados. La partición del espacio de estados se hizo considerando el valor teórico de potencial eléctrico de cada onda en un ECG, de acuerdo a los descrito en la sección 2.3. En la tabla 5.3 se muestra las dos particiones empleadas para el ECG completo.

La partición 1 hace énfasis en analizar la producción de entropía considerando la frecuencia de los complejos QRS en un intervalo de potencial eléctrico, en función de si los ECGs pertenecen a pacientes sanos o pacientes con arritmia. La partición 2 permitió estudiar la diferencia en la producción de entropía entre pacientes sanos y pacientes con arritmia, analizando sólo las fluctuaciones alrededor de la línea isoeletrica y las transiciones a los estados fuera de este intervalo de potencial eléctrico. En la figura 5.2 se representa, a manera de ejemplo, la representación gráfica de la partición 1.

Partición 1		Partición 2	
Estados	Voltaje (mV)	Estados	Voltaje (mV)
A1	[-5.0,-1.0)	A1	[-5.0,-0.25)
A2	[-1.0,-0.3)	A2	[-0.25,-0.10)
A3	[-0.3,-0.2)	A3	[-0.1,-0.05)
A4	[-0.2,-0.1)	A4	[-0.05,0.0)
A5	[-0.1,0.0)	A5	[0.0,0.05)
A6	[0.0,0.1)	A6	[0.05,0.1)
A7	[0.1,0.2)	A7	[0.1,0.15)
A8	[0.2,0.6)	A8	[0.15,0.20)
A9	[0.6,1.1)	A9	[0.2,10.0)
A10	[1.1,10.0)		

Tabla 5.3: Partición del espacio de estados del ECG completo.

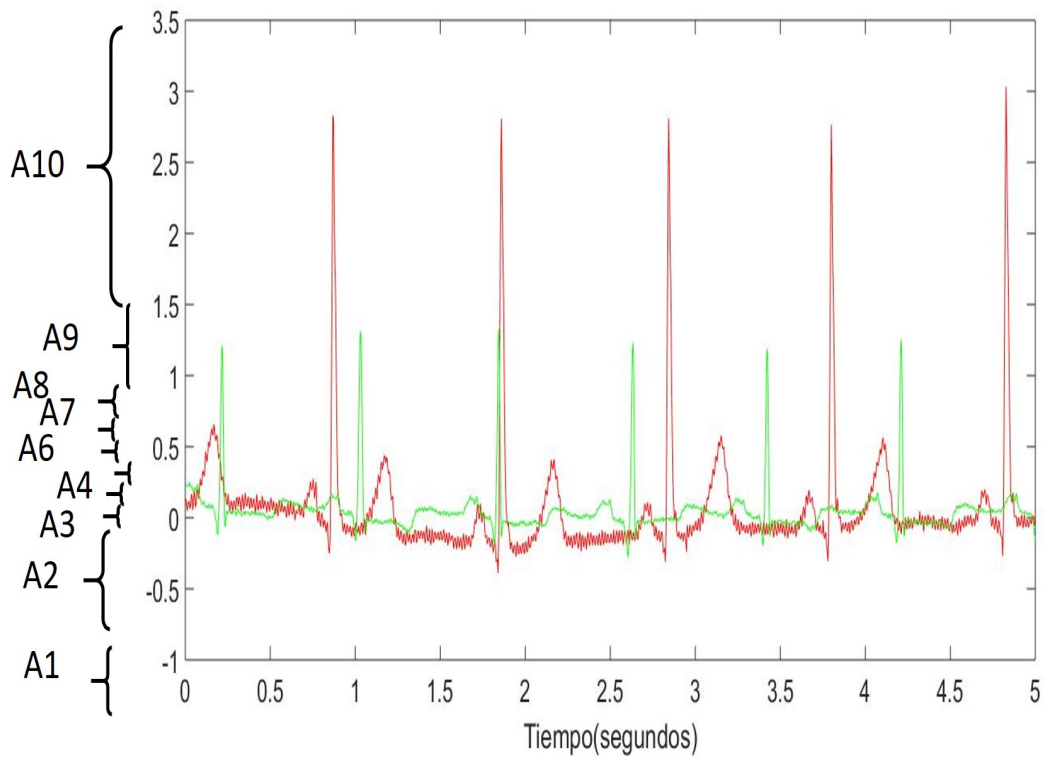
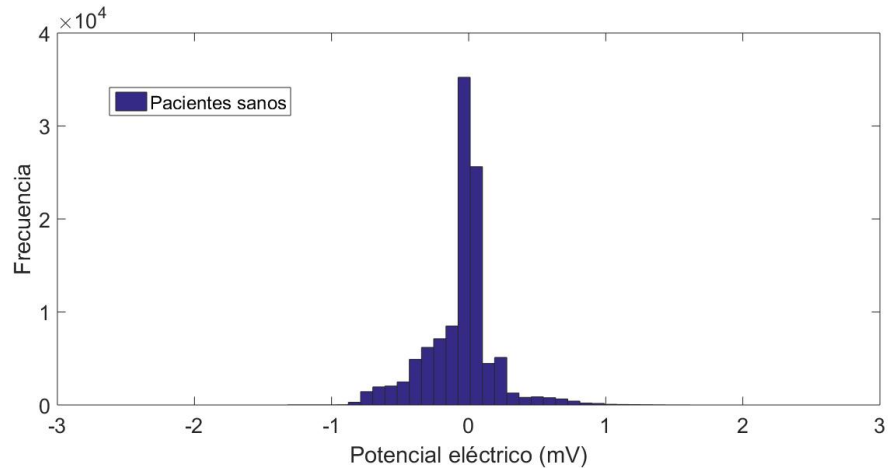


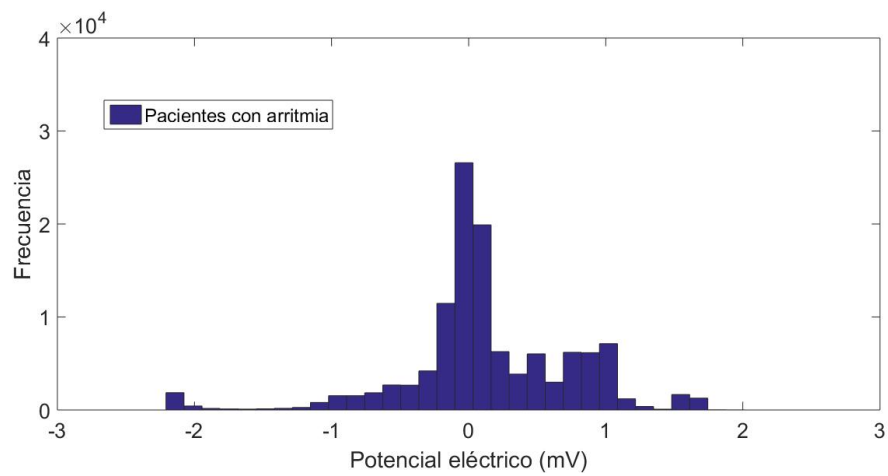
Figura 5.2: Partición del espacio de estados del ECG completo.

#### 5.4.2. Partición del espacio de estados de la línea isoelectrónica

En la figura 5.3 se muestra los histogramas de la línea isoelectrónica de pacientes sanos y pacientes con arritmia, respectivamente. El propósito de estos histogramas es mostrar que la línea isoelectrónica de los pacientes sanos tiene un comportamiento más estable que el de los pacientes arrítmicos. Como se puede apreciar, la dispersión de los pacientes arrítmicos es mayor, esto sugiere que los ECGs de los pacientes sanos son más «ordenados» que el de los arrítmicos, y con ello se espera que los ECG de pacientes sanos presenten una mayor producción de entropía.



(a) Pacientes sanos.



(b) Pacientes con arritmia.

Figura 5.3: Distribución de la línea isoelectrica.

La partición del espacio de estados de la línea isoelectrica, tanto de pacientes sanos como de pacientes con arritmia, se hizo numéricamente, en función de la distribución de la línea isoelectrica alrededor del cero, empleando los histogramas obtenidos. En la tabla 5.4 se muestra los valores de las particiones empleadas para el espacio de estados de la línea isoelectrica; en la figura 5.4 se representa gráficamente la partición.

Estados	Voltaje
A1	[-5.0,-1.0)
A2	[-1.0,-0.3)
A3	[-0.3,-0.2)
A4	[-0.2,-0.1)
A5	[-0.1,0.0)
A6	[0.0,0.1)
A7	[0.1,0.2)
A8	[0.2,0.6)
A9	[0.6,1.1)
A10	[1.1,10.0)

Tabla 5.4: Partición del espacio de estados de la línea isoeletrica.

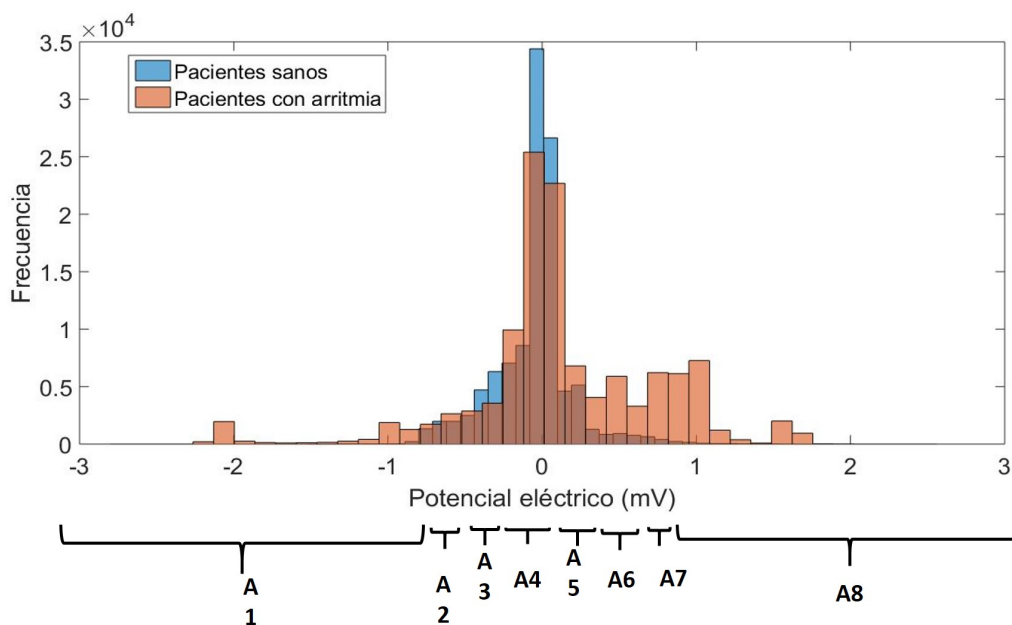


Figura 5.4: Partición del espacio de estados de la línea isoeletrica.

## 5.5. Obtención de la línea isoeletrica

El método empleado para la obtención de la línea isoeletrica de cada ECG, se basa el algoritmo propuesto en [1]. La idea principal fue identificar «puntos» del ECG que corresponden a la línea isoeletrica; se tomaron los puntos de inicio de cada onda del ECG, los cuales, en su conjunto, forman la línea isoeletrica (los puntos verdes en la figura 5.5a). La representación de la línea isoeletrica está dada por vectores reales de patrones ordinales. Dado el vector de la serie de tiempo real  $(x_n)_{n=1}^N$ , es posible obtener un vector representativo  $\mathbf{u}$  de la serie de tiempo original. Este vector representativo se obtiene mediante un

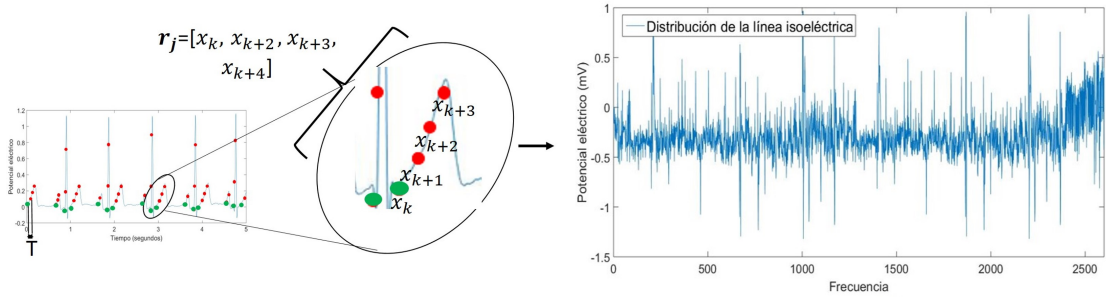
tiempo de muestreo  $T$ , en el vector de la serie de tiempo real  $(x_n)_{n=1}^N$ ; la dependencia de  $\mathbf{u}$  de  $T$  la indicaremos mediante  $\mathbf{u}^T$ . En el apéndice B se muestra el programa desarrollado para la obtención de la línea isoeletrica, cuyo procedimiento se resume a continuación (figura 5.5):

1. Dado el ECG  $(x_n)_{n=1}^N$ , se obtiene la muestra  $\mathbf{u}^T$ .
2. A partir  $\mathbf{u}^T$ , se obtiene un conjunto de vectores de patrones ordinales  $\mathbf{s}^T = \{\mathbf{r}_j \neq \emptyset : \mathbf{r}_j \in o(\mathbf{s}), j = 1, 2, \dots, m\}$ , donde

$$o(\mathbf{s}) = \{\mathbf{r}_j : \mathbf{r}_j = [x_k, x_{k+1}, \dots, x_L], \text{ donde } x_k < x_{k+1} < \dots < x_L, \\ |\mathbf{r}_j(L) - \mathbf{r}_j(1)| \geq c \text{ y } j = 1, 2, \dots, m\}$$

la constante  $c$  es un valor teórico de potencial eléctrico.

3. La *línea isoeletrica* está dada por  $\mathbf{v}^T = [\mathbf{r}_1(1), \mathbf{r}_2(1), \dots, \mathbf{r}_m(1)]$ .



(a) Patrones ordinales en un ECG.

(b) Representación de la línea isoeletrica de un ECG.

Figura 5.5: Obtención de la línea isoeletrica.

## 5.6. Codificación simbólica, matriz de transición y distribución de probabilidad invariante

Sea  $\mathbf{v} = [v_1, v_2, \dots, v_r]$  el vector de datos reales de la línea isoeletrica o de un ECG. A partir de un conjunto de número naturales  $\{1, 2, \dots, k\}$  representando el espacio de estados  $\alpha = [A_1, A_2, \dots, A_k]$ , a cada  $\mathbf{v}$  se le asigna una secuencia simbólica  $\mathbf{w}^{T_2, \alpha} = [\omega_1, \omega_2, \dots, \omega_n]$ , donde  $n = \frac{r}{T_2}$  y  $T_2$  es el tiempo de muestreo en  $\mathbf{v}$ . Decimos que  $\mathbf{v}$  tiene la secuencia  $\mathbf{w}^{T_2, \alpha}$  si  $\mathbf{v}(j) \in \alpha(i)$ .

A partir de la secuencia simbólica  $\mathbf{w}^{T_2, \alpha} = [\omega_1, \omega_2, \dots, \omega_n]$ , es posible estimar la *probabilidad de transición*  $\mathbb{P}(\omega_m = j \mid \omega_{m-1} = i)$  y la probabilidad estacionaria de los estados  $\mathbb{P}(\omega = j)$ , mediante las probabilidades empírica definidas por

$$\hat{\mathbb{P}}_n(\omega_m = j \mid \omega_{m-1} = i) = \frac{1}{n - (n - \delta)} \sum_{m=2}^n \chi_i(\omega_m) * \chi_j(\omega_{m-1}) \quad (5.1)$$

y

$$\hat{\mathbb{P}}_n(\omega = j) = \frac{1}{n} \sum_{m=1}^n \chi_j(\omega_m) \quad (5.2)$$

respectivamente, y donde  $\delta$  es la frecuencia del estado  $i$  y

$$\chi_k(\omega) = \begin{cases} 1 & \text{si } \omega = k \\ 0 & \text{de otra forma.} \end{cases}$$

Sabemos por la ley de los grandes números para cadenas de Márkov que  $\lim_{n \rightarrow \infty} \hat{\mathbb{P}}_n = \mathbb{P}$ .

La entrada  $i, j$  de matriz de transición se estima mediante la probabilidad de transición  $[P_{i,j}]_{i,j \in \alpha} = [\mathbb{P}(\omega_m = j \mid \omega_{m-1} = i)]_{i,j \in \alpha}$ . La probabilidad invariante  $\pi(j)$  del estado  $i$  al instante  $n$ , es la entrada  $j$  del vector propio izquierdo de la matriz de transición construida al instante  $n$ :  $\pi P = \pi$ .

Debido a los distintas frecuencias de muestreo empleadas por cada base de datos (ver tabla 5.2), el número de muestras empleadas por unidad de tiempo para estimar  $\hat{\mathbb{P}}_n(\omega = j)$  y  $\hat{\mathbb{P}}_n(\omega_m = j \mid \omega_{m-1} = i)$ , no es homogéneo; buscando reducir esta diferencia en el número de muestras empleadas por unidad de tiempo, para la primer frecuencia de muestreo, se empleó  $T_1 = 2, T_1 = 40, T_1 = 3$  y  $T_1 = 3$  para las bases de datos FANTASIA, CEBS, MIT-BIH y MGH/ME, respectivamente; para la segunda frecuencia de muestreo, se empleó  $T_2 = 5, T_2 = 100, T_2 = 6$  y  $T_2 = 6$ , para las bases de datos FANTASIA, CEBS, MIT-BIH y MGH/ME, respectivamente (ver tabla 5.2). Emplear estas distintas frecuencias de muestreo permite obtener un número de muestras por unidad de tiempo más homogéneo. En el apéndice C se muestra el programa desarrollado en esta tesis para la estimación de la probabilidad invariante de los estados, y en el apéndice D se encuentra el programa desarrollado para la estimación de la probabilidad de transición de los estados y la construcción de la matriz de transición.

## 5.7. Cálculo de la producción de entropía

Estimar la producción de entropía de un ECG para un determinado instante  $n$ , es un proceso que implica diferentes etapas. A partir de la secuencia  $\mathbf{w}^{T_2, \alpha}$  obtenida al instante  $n$ , se estima la probabilidad de transición  $\mathbb{P}_n(\omega_m = j \mid \omega_{m-1} = i)$ , para cada uno de los estados. A partir de estas probabilidades de transición se estima la matriz de transición  $\hat{P}$  al instante  $n$ . La estimación de la distribución de probabilidad invariante hasta el tiempo  $n$  se obtiene a partir de  $\hat{P}$ . Finalmente, se calcula la producción de entropía al tiempo  $n$ . Este procedimiento se repite para el cálculo de la producción de entropía en cada instante. En el apéndice E se muestra el programa desarrollado para el cálculo de la producción de entropía.

## 5.8. Cálculo de un estimador de irreversibilidad empírico

A partir de la secuencia simbólica  $\mathbf{w}^{T_2, \alpha} = [\omega_1, \omega_2, \dots, \omega_n]$ , es posible dar una prueba de irreversibilidad a través de un estimador. Decimos que  $\mathbf{a}_i = \{a_1^k, a_k^1\}$  es una palabra que puede formarse si  $\mathbf{a}_i \in \mathbf{w}^{T_2, \alpha}$ ,  $a_1^k = (a_1 a_2 \cdots a_{k-1} a_k)$  y  $a_k^1 = (a_k a_{k-1} \cdots a_2 a_1)$ , donde  $k$  es el tamaño de la palabra. El estimador de irreversibilidad que aquí definimos está dado por

$$\langle \hat{S}_k \rangle_{\text{pacientes}} = \frac{1}{|D_k|_{\mathbb{P}(\mathbf{a}_i) > 0}} \sum_{i=1}^{|D_k|} \hat{S}_k(\mathbf{a}_i) \quad (5.3a)$$

donde

$$\hat{S}_k(\mathbf{a}_i) = \log \frac{\hat{\mathbb{P}}(a_1^k)}{\hat{\mathbb{P}}(a_k^1)} \quad (5.3b)$$

En la ecuación 5.3,  $|D_k|_{\mathbb{P}(\mathbf{a}_i) > 0}$  es el número de palabras  $\mathbf{a}_i$ , de tamaño  $k$ , que pueden formarse y  $\mathbb{P}(\mathbf{a}_i) > 0$  si  $\ell_1, \ell_2 \geq 1$ .  $\mathbb{P}(\mathbf{a}_1^k) = \frac{\ell_1}{n-2}$  y  $\mathbb{P}(a_k^1) = \frac{\ell_2}{n-2}$ ;  $\ell_1, \ell_2 \geq 1$  son la frecuencia de  $a_k^1$  y  $a_1^k$ , respectivamente.  $\ell_j$  está dada por  $\ell_j = \sum_{i=1}^{n-k} L_j(a)$ , donde

$$L_j(a) = \begin{cases} 1 & \text{si } a \in \mathbf{w}^{T_2, \alpha} \\ 0 & \text{de otra forma.} \end{cases}$$

El estimador de irreversibilidad definido por la ecuación 5.3 permite estimar la producción de entropía del sistema de forma empírica sin pedir restricciones de modelo a los datos. Lo anterior es de gran utilidad cuando no hay ninguna suposición sobre el modelo. Sin embargo, los estimadores empíricos en general, presentan tiempos lentos de convergencia, lo que representa una desventaja [7].





# Capítulo 6

## Resultados y discusión

En esta tesis nos enfocamos en proponer el principio de una metodología para detectar la diferencia en los ECGs entre pacientes sanos y arrítmicos, usando como indicador a la producción de entropía. Para lograr esto hace falta asumir que los ECGs pueden presentar una naturaleza markoviana. Es importante aclarar que es una suposición viable ya que es evidente que las pulsaciones cardíacas tienen cierto grado de memoria. Por simplicidad, aquí sólo estudiamos el caso de memoria al paso anterior. A partir de esa suposición, usamos directamente las series de tiempo (ECG) provenientes de los pacientes, para generar las secuencias simbólicas del sistema, en función de la partición del espacio de estados. Una vez obtenida la secuencia simbólica representando el ECG, estimamos la probabilidad estacionaria para cada uno de los estados descritos por la partición. Con tal de inferir a partir de los datos un modelo markoviano que mejor aproxima los datos, estimamos las probabilidades de transición, para obtener las matrices de transición del modelo markoviano. Una vez estimados los parámetros de modelo podemos calcular, usando la teoría, la producción de entropía de la estimación empírica de la cadena de Márkov. A continuación describiremos los resultados.

### 6.1. Modelo

El modelo markoviano de un sistema está representado por su matriz de transición de estados. En esta sección se presenta, a manera de ejemplo, la estimación  $\hat{P}$  de la matriz de transición de estados de un paciente sano, para la partición 1 del espacio de estados y el tiempo de muestreo  $T_2$ ; para esta partición, se obtuvieron 10 estados del sistema (ver tabla 5.3). Por la suposición 3 del modelo (cadenas de Márkov irreducibles), se asignó una probabilidad de transición arbitraria  $p_{i,j} = 0,00001$  a las entradas  $p_{i,j} = 0$  del modelo original, lo cual permite una conexión completa entre los estados. Este modelo fue obtenido para 30 minutos de muestreo del ECG correspondiente. En la figura 6.1 se representa las transiciones permitidas por el modelo. Por claridad, en el grafo sólo se señala las probabilidades de transición correspondientes a los bucles.

$$\hat{P} = \begin{bmatrix} 0,9750 & 0,0250 & 0,0001 & 0,0001 & 0,0001 & 0,0001 & 0,0001 & 0,0001 & 0,0001 & 0,0001 \\ 0,0031 & 0,2954 & 0,1908 & 0,0677 & 0,0246 & 0,0215 & 0,0001 & 0,0092 & 0,0062 & 0,3815 \\ 0,0001 & 0,0293 & 0,4564 & 0,3251 & 0,0355 & 0,0058 & 0,0021 & 0,0021 & 0,0131 & 0,1307 \\ 0,0001 & 0,0090 & 0,0524 & 0,6544 & 0,2462 & 0,0109 & 0,0032 & 0,0022 & 0,0063 & 0,0154 \\ 0,0001 & 0,0012 & 0,0077 & 0,0991 & 0,7247 & 0,1501 & 0,0081 & 0,0019 & 0,0024 & 0,0048 \\ 0,0001 & 0,0002 & 0,0027 & 0,0138 & 0,2361 & 0,5941 & 0,1149 & 0,0308 & 0,0026 & 0,0047 \\ 0,0001 & 0,0004 & 0,0019 & 0,0132 & 0,0960 & 0,2925 & 0,2678 & 0,2102 & 0,1051 & 0,0130 \\ 0,0001 & 0,0013 & 0,0026 & 0,0074 & 0,0431 & 0,1888 & 0,2621 & 0,1804 & 0,2982 & 0,0161 \\ 0,0001 & 0,0010 & 0,0019 & 0,0095 & 0,0196 & 0,0159 & 0,0867 & 0,1538 & 0,6768 & 0,0348 \\ 0,0001 & 0,0092 & 0,0390 & 0,2041 & 0,2378 & 0,1139 & 0,0321 & 0,0313 & 0,1659 & 0,1667 \end{bmatrix}$$

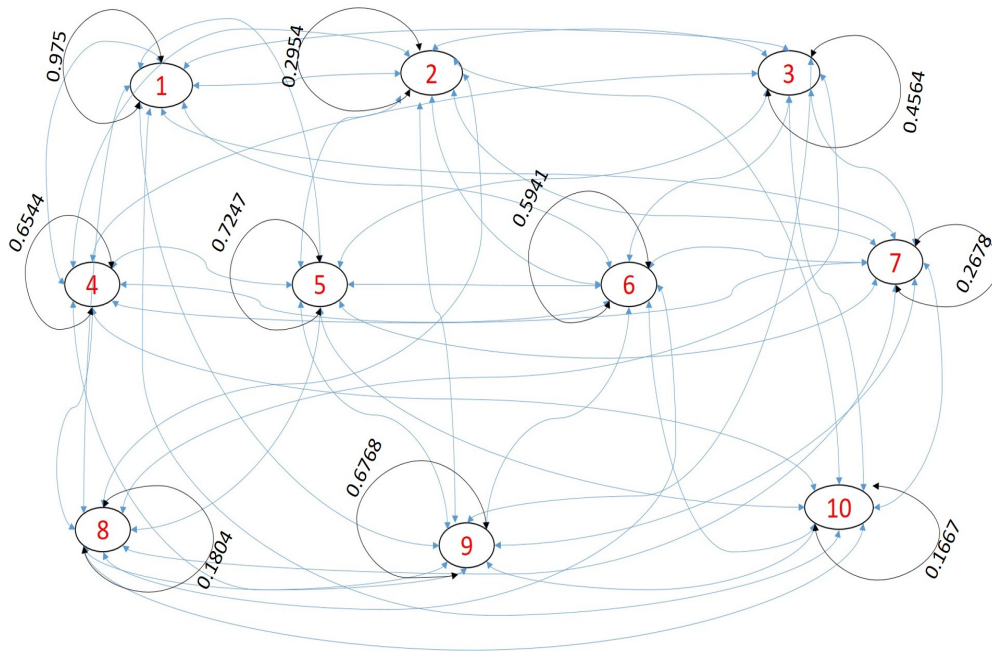


Figura 6.1: Grafo de un ECG.

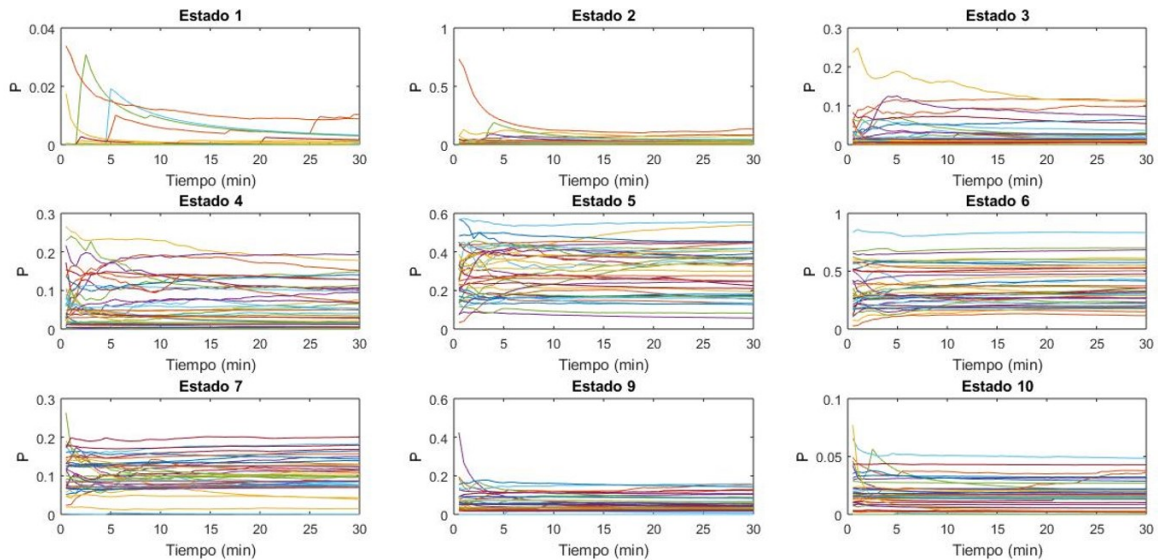
## 6.2. Probabilidad invariante

Los resultados mostrados en esta sección son, a manera de ejemplo, para una de las condiciones descritas en la metodología, ya que, para todas las condiciones, la probabilidad empírica exhibió un comportamiento de convergencia similar. Los resultados mostrados en la presente sección corresponden al segundo tiempo de muestreo ( $T_2$ ) y la primera partición del espacio de estados (ver tabla 5.2 y sección 5.4).

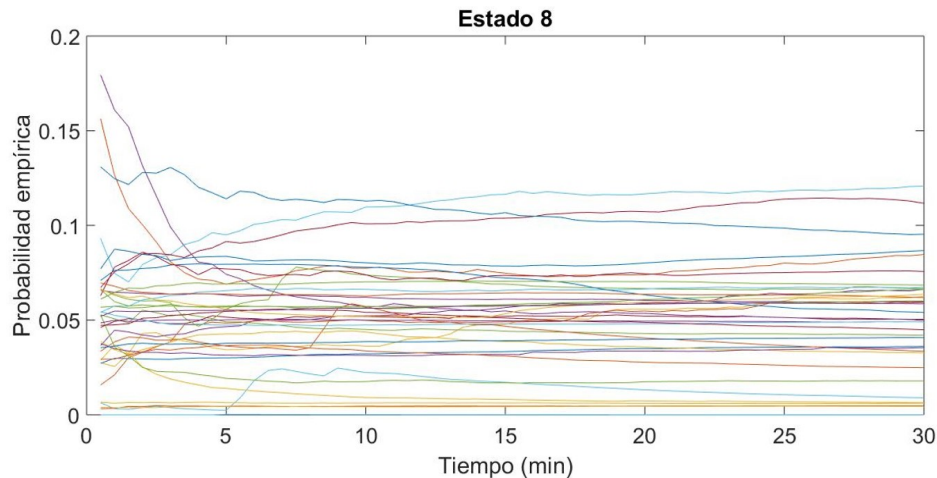
En la figura 6.2a se muestra la estimación de la probabilidad empírica de los estados al tiempo  $n$  (en minutos), para cada uno de los estados, para los pacientes sanos. En la figura 6.2b se observa que, estadísticamente, hay una clara convergencia de la probabilidad

empírica de los estados, para el 83% de los pacientes (33 de los 40 pacientes), en un tiempo promedio de 15 minutos. El valor al que convergen es el que suponemos como la probabilidad estacionaria de dicho estado, para cada paciente. Respecto a los pacientes que no parecen exhibir una clara convergencia en el intervalo de tiempo estudiado, esto puede atribuirse al ruido presente durante la adquisición de los datos o diversas variaciones en la actividad del paciente durante la adquisición de datos.

En la figura 6.3 se muestra la estimación de la probabilidad empírica de los estados al



(a) Probabilidad estacionaria de cada estado.



(b) Probabilidad estacionaria del estado 8.

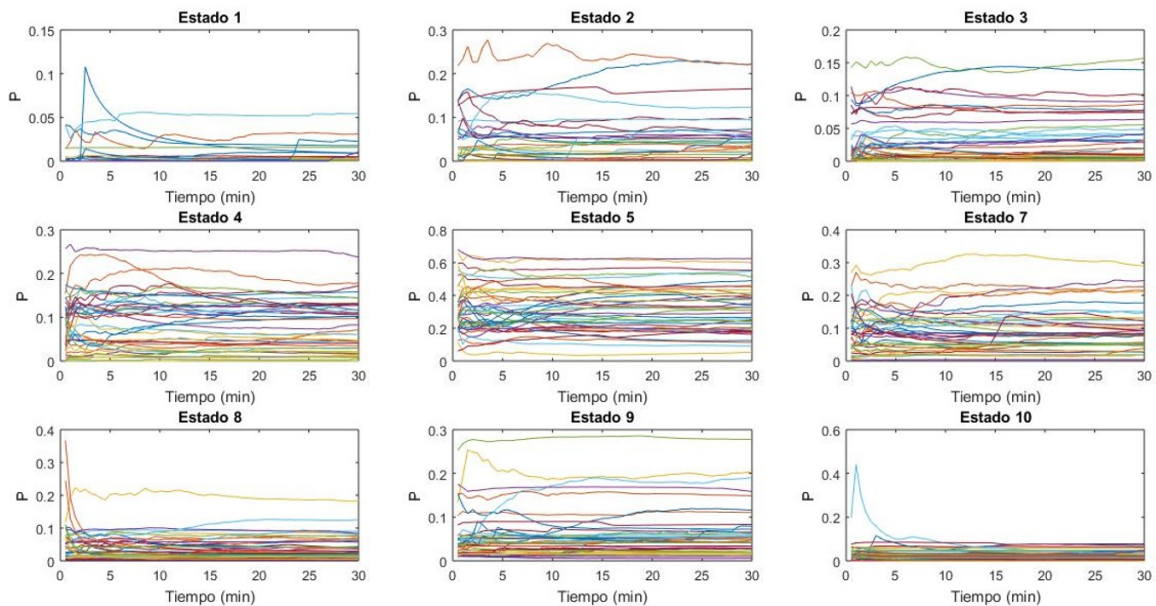
Figura 6.2: Probabilidad empírica de los estados para los pacientes sanos.

tiempo  $n$  (en minutos), para cada uno de los estados, para los pacientes con arritmia. En la figura 6.2b se observa que, al igual que con los pacientes sanos, estadísticamente hay una convergencia de la probabilidad empírica para el 80% de los pacientes con arritmia (32 de 40 pacientes), en un tiempo promedio de 15 minutos. El valor al que convergen es

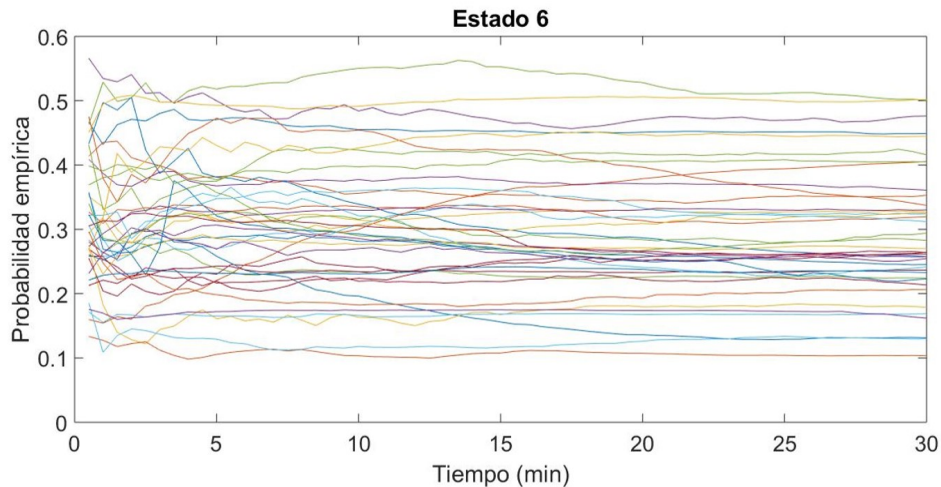
el que suponemos como la probabilidad estacionaria de dicho estado, para cada paciente.

Los resultados de convergencia a la probabilidad estacionaria de los estados, tanto para pacientes sanos como pacientes con arritmia, dan sentido a la suposición inicial del modelo, al trabajar con cadenas de Márkov estacionarias.

En la figura 6.4 se muestra la probabilidad empíricas promedio y su barra de desviación



(a) Probabilidad estacionaria de cada estado.



(b) Probabilidad estacionaria del estado 6.

Figura 6.3: Probabilidad empírica de los estados para los pacientes con arritmia.

estándar, para pacientes sanos y pacientes con arritmia. Se observa que en pacientes con arritmia, para todos los estados, la desviación estándar es mayor, exhiben una mayor fluctuación que los pacientes sanos. Vemos que los estados 10 y 5 presentan una mayor dife-

rencia en la probabilidad empírica entre pacientes sanos y pacientes con arritmia. En el caso del estado 5, esta diferencia de probabilidad de estar alrededor de la línea isoeletrica, llevó a analizar la producción de entropía del ECG, pero en función sólo de la línea isoeletrica, como se explica en la sección 6.3.

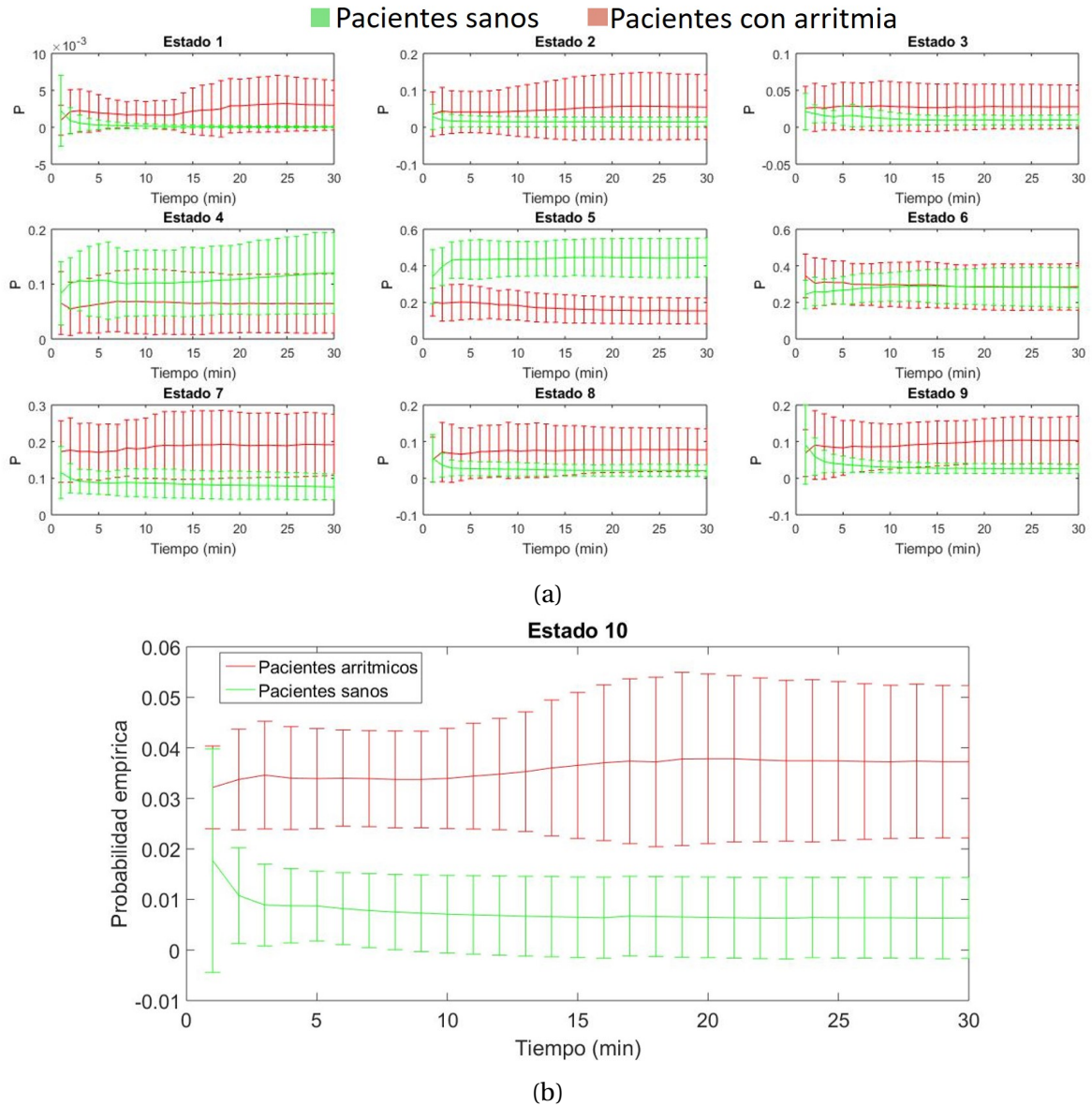


Figura 6.4: Probabilidad empírica promedio de los estados para pacientes sanos y pacientes con arritmia.



### 6.3. Producción de entropía: modelo markoviano

En la figura 6.5 se muestra la estimación de la producción de entropía promedio hasta el tiempo  $n$  para pacientes sanos y pacientes con arritmia, en función de la partición 1. La producción de entropía en las figuras 6.5a y 6.5b corresponden a los tiempos de muestreo  $T_1$  y  $T_2$ , respectivamente. A diferencia del tiempo de muestreo  $T_1$ , para el tiempo de muestreo  $T_2$ , se observa diferencia en la producción de entropía promedio entre ambos grupos de pacientes. Para determinar si esta diferencia es significativa, hace falta validar estadísticamente el estudio, empleando un mayor número de datos.

En la figura 6.6 se muestra la evolución en el tiempo de la producción de entropía promedio para pacientes sanos y pacientes con arritmia, en función de la línea isoelectrica. Al igual que para el tiempo de muestreo  $T_2$ , se observa diferencia en la producción de entropía entre ambos grupos de pacientes. La mayor producción de entropía en pacientes sanos es consistente con el hecho de que su distribución alrededor del cero exhiba mayor «orden» respecto a los pacientes con arritmia (ver figura 5.3).

El objetivo de esta tesis es exhibir evidencia numérica de la diferencia en producción entropía entre pacientes sanos y pacientes con arritmia cardíaca. Los resultados numéricos sugieren que existe una diferencia en la producción de entropía promedio entre estos dos grupos de pacientes, pero también reflejan que hay una intersección entre la desviación de estándar de ambos grupos. Estos resultados dan argumentos de plausibilidad de que la producción de entropía puede servir como un indicador de condiciones biomédicas importantes en los ECGs. Sin embargo, hace falta analizar los datos bajo condiciones más homogéneas en cada grupo y entre ambos grupos. Por ejemplo, en la tabla 5.2 vemos que la base de datos «CEBS» no reporta edad promedio de los pacientes estudiados, y sólo los reporta como «adultos». En la misma tabla vemos también que las muestras de los datos se obtuvieron para ambos sexos. Adicionalmente, cada base de datos presenta un tiempo de muestreo diferente. Al no haber homogeneidad en el tiempo de muestreo, el cálculo de las probabilidades estacionarias y de las probabilidades de transición por unidad de tiempo se ve «afectada», ya que se emplea un número de muestras diferente por unidad de tiempo para cada base de datos, en función de la frecuencia de muestreo.

### 6.4. Producción de entropía: estimador de irreversibilidad

En la tabla 6.1 se muestra la producción de entropía promedio y su desviación estándar respectiva, calculadas a partir del modelo markoviano y a partir del estimador de irreversibilidad, para el tiempo de muestreo  $T_2$ . Al igual que los resultados obtenidos a partir del modelo markoviano, la producción de entropía obtenida a partir del estimador de irreversibilidad definido en (5.3a), es mayor en pacientes sanos que pacientes con arritmia, tanto para la partición 1 como para la línea isoelectrica. El estimador de irreversibilidad evidencia una diferencia más clara en la producción de entropía promedio, ya que para la

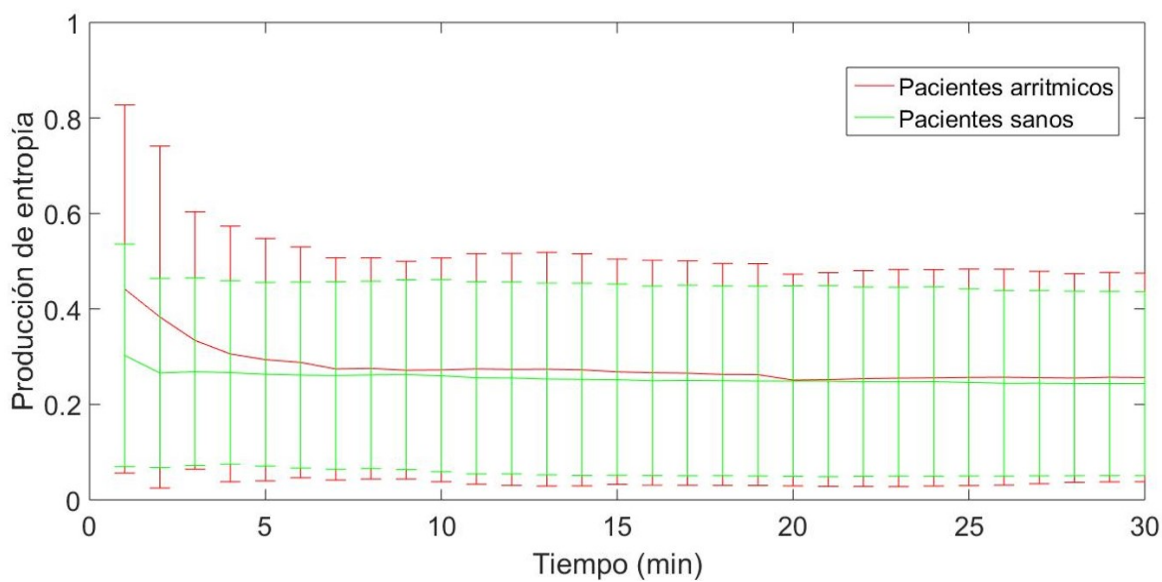
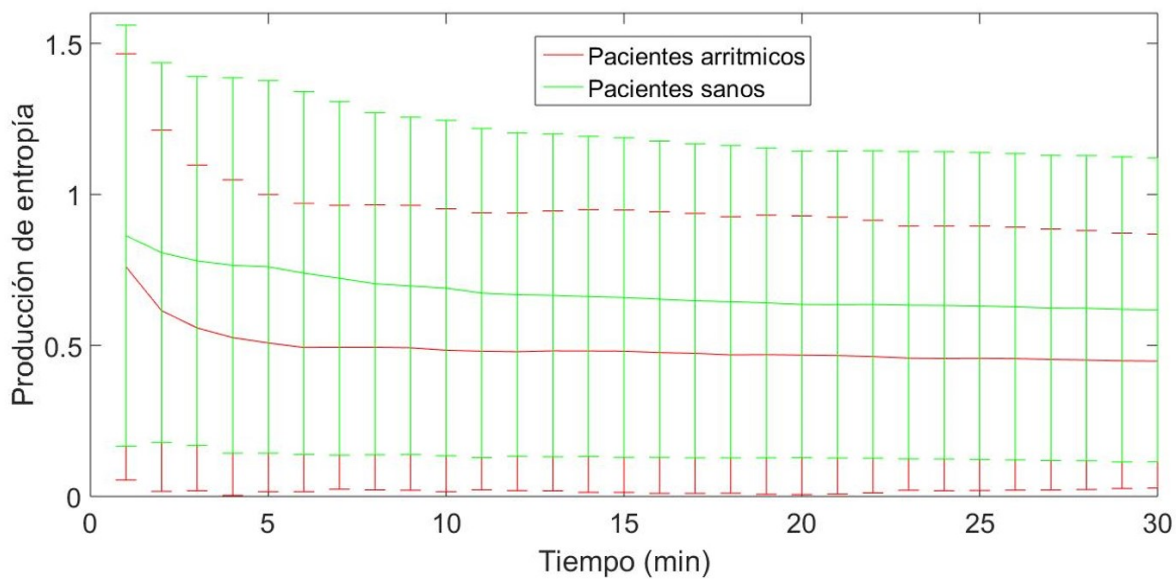
(a) Tiempo de muestreo  $T_1$ .(b) Tiempo de muestreo  $T_2$ .

Figura 6.5: Producción de entropía promedio para la partición 1.

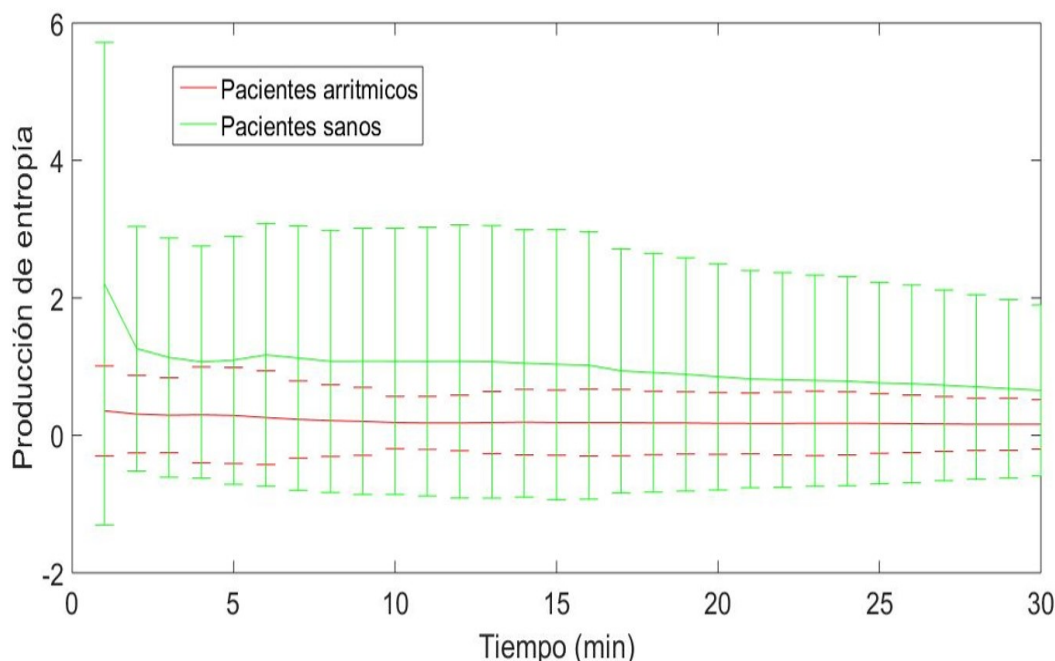


Figura 6.6: Producción de entropía promedio en función de la línea isoelectrica.

línea isoelectrica, la producción de entropía en pacientes sanos es 5 veces mayor que en pacientes con arritmia; para la partición 1, la producción de entropía en pacientes sanos es 1.5 veces mayor que en pacientes con arritmia.

La consistencia en los resultados obtenidos en la producción de entropía promedio mediante el enfoque del modelo markoviano y el enfoque del estimador de irreversibilidad, sugiere que es posible exhibir una producción de entropía típica en pacientes sanos y en pacientes con arritmia, si se encuentra la representación simbólica adecuada. En la tabla 6.1, PS es una abreviatura para «pacientes sanos» y PA para «pacientes arritmicos».

	Partición 1		Línea isoelectrica	
	PS	PA	PS	PA
Modelo markoviano	0.6179±0.5032	0.4480±0.4205	0.6545±1.241	0.1603±0.3615
Estimador empírico	0.00071±0.0005	0.00044±0.0003	0.0081±0.0142	0.0014±0.0037

Tabla 6.1: Producción de entropía promedio.



# Capítulo 7

## Conclusiones y trabajo futuro

### 7.1. Conclusiones

En general, la probabilidad empírica de los estados converge en un tiempo promedio de 15 minutos, para todos los factores de variabilidad y con los dos grupos de pacientes. La probabilidad estacionaria de pacientes con arritmia cardíaca mostró mayor fluctuación en relación con los pacientes sanos.

Los resultados de producción de entropía son consistentes en dos sentidos. Primero, al comparar la producción de entropía entre pacientes sanos y pacientes con arritmia, los resultados son consistentes con los datos reales. Es decir, si se analiza gráficamente los ECG de pacientes sanos y pacientes con arritmia, vemos que los ECG correspondientes a pacientes con arritmia exhiben un mayor desorden, son más simétricos respecto a la inversión temporal, y por lo tanto se esperaría una menor producción de entropía. Segundo, existe consistencia entre la producción de entropía calculada a partir del enfoque markoviano y la producción de entropía obtenida a partir del enfoque del estimador de irreversibilidad; en ambos casos se tiene una mayor producción en pacientes sanos que en pacientes con arritmia cardíaca. Este comportamiento de la producción de entropía para pacientes sanos y pacientes con arritmia se tuvo para los dos tiempos de muestreo empleados, para todas las particiones y para la línea isoeletrica. Sin embargo, para dar resultados más contundentes respecto a las fluctuaciones típicas de la producción de entropía en pacientes sanos y pacientes con arritmia, es necesario un estudio más exhaustivo y detallado, aumentando las estadísticas y restringiendo más las condiciones.

### 7.2. Trabajo futuro

- Establecer una metodología sistemática para la codificación simbólica que capture adecuadamente la información de interés del sistema.
- Refinar el estudio usando un mayor número de datos y validándolos estadísticamente.

- Generalizar el trabajo usando cadenas escondidas de Márkov y/o medidas de Gibbs.
- Proponer estimadores de irreversibilidad.
- Proponer la producción de entropía como un indicador útil de condiciones médicas de interés.

# Bibliografía

- [1] José M Amigó, Karsten Keller, and Valentina A Unakafova. Ordinal symbolic analysis and its application to biomedical recordings. *Phil. Trans. R. Soc. A*, 373(2034):20140091, 2015.
- [2] Rodrigo Varejão Andreão, Bernadette Dorizzi, and Jérôme Boudy. Ecg signal analysis through hidden markov models. *IEEE Transactions on Biomedical engineering*, 53(8):1541–1549, 2006.
- [3] Karl-Jürgen Bär, Michael Karl Boettger, Mandy Koschke, Steffen Schulz, Pratap Chokka, Vikram K Yeragani, and Andreas Voss. Non-linear complexity measures of heart rate variability in acute schizophrenia. *Clinical neurophysiology*, 118(9):2009–2015, 2007.
- [4] Angela A Bruzzo, Benno Gesierich, Maurizio Santi, Carlo Alberto Tassinari, Niels Birbaumer, and Guido Rubboli. Permutation entropy to detect vigilance changes and preictal states from scalp eeg in epileptic patients. a preliminary study. *Neurological sciences*, 29(1):3–9, 2008.
- [5] Camillo Cammarota and Enrico Rogora. Time reversal, symbolic series and irreversibility of human heartbeat. *Chaos, Solitons & Fractals*, 32(5):1649–1654, 2007.
- [6] Juan Osvaldo Balbuena Carrillo. Sistema cardiovascular: las derivaciones del electrocardiograma. [urlhttp://fisiologiahumanajuan.blogspot.com/2013/02/sistema-cardiovascular-las-derivaciones.html](http://fisiologiahumanajuan.blogspot.com/2013/02/sistema-cardiovascular-las-derivaciones.html), 2013.
- [7] Jean-René Chazottes and F Redig. Testing the irreversibility of a gibbsian process via hitting and return times. *Nonlinearity*, 18(6):2477, 2005.
- [8] R. Cofré and C. Maldonado. Information entropy production of maximum entropy markov chains from spike trains. *Entropy*, 2018(20(1):34).
- [9] Dale Davis. *Interpretación del ECG: su dominio rápido y exacto*. Ed. Médica Panamericana, 2007.
- [10] MA Garcia-Gonzalez, A Argelagós, Mireya Fernández-Chimeno, and J Ramos-Castro. Differences in qrs locations due to ecg lead: relationship with breathing. In *XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013*, pages 962–964. Springer, 2014.

- [11] Miguel A García-González, Ariadna Argelagós-Palau, Mireya Fernández-Chimeno, and Juan Ramos-Castro. A comparison of heartbeat detectors for the seismocardiogram. In *Computing in Cardiology Conference (CinC), 2013*, pages 461–464. IEEE, 2013.
- [12] Pierre Gaspard. Time-reversed dynamical entropy and irreversibility in markovian random processes. *Journal of statistical physics*, 117(3-4):599–615, 2004.
- [13] AL Goldberger, LAN Amaral, and L Glass. 1. m. hausdorff, pc ivanov, rg mark, je mietus, gb moody, c.-k. peng, he stanley, physiobank, physiokit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation*, 101:23, 2000.
- [14] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. Physiobank, physiokit, and physionet. *Circulation*, 101(23):e215–e220, 2000.
- [15] Christian W Hamm and Stephan Willems. *El electrocardiograma: su interpretación práctica; 32 cuadros*. Ed. Médica Panamericana, 2010.
- [16] John Hampton. *ECG fácil*. Elsevier Health Sciences, 2014.
- [17] Andrew Hart, Servet Martínez, and Felipe Olmos. A Gibbs approach to Chargaff's second parity rule. *J. Stat. Phys.*, 146(2):408–422, 2012.
- [18] Nicholas P Hughes, Lionel Tarassenko, and Stephen J Roberts. Markov models for automated ecg interval analysis. In *Advances in Neural Information Processing Systems*, pages 611–618, 2004.
- [19] Morin R Goldberger AL Lipsitz LA. Iyengar N, Peng C-K. Fantasia database. *Am J Physiol*, 271(23):e1078–e1084, 1996.
- [20] Da-Quan Jiang, Min Qian, and Ming-Ping Qian. *Mathematical theory of nonequilibrium steady states: on the frontier of probability and dynamical systems*. Springer, 2003.
- [21] Gregory F Lawler. *Introduction to stochastic processes*. CRC Press, 2006.
- [22] Xiaoli Li, Gaoxian Ouyang, and Douglas A Richards. Predictability analysis of absence seizures with permutation entropy. *Epilepsy research*, 77(1):70–74, 2007.
- [23] Jorge Hernán López-Ramírez. *La alegría de leer el electrocardiograma*. Bogotá: Médica Celsus, 2006.
- [24] Oscar Castellero Mimenza. Las 13 partes del corazón humano (y sus funciones). [urlhttps://psicologiymente.com/salud/partes-del-corazon](https://psicologiymente.com/salud/partes-del-corazon), 2013.
- [25] George B Moody and Roger G Mark. The impact of the mit-bih arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine*, 20(3):45–50, 2001.

- [26] Daniel Novák. *Electrocardiogram Signal Processing using Hidden Markov Models*. PhD thesis, PhD thesis, Czech Technical University in Prague, Faculty of Electrical Engineering, Technicka 2, 166 27 Prague 6, Czech Republic, 2003. 5.1, 5.2, 5.2, 2003.
- [27] Daniel Novák. *Electrocardiogram Signal Processing using Hidden Markov Models*. PhD thesis, PhD thesis, Czech Technical University in Prague, Faculty of Electrical Engineering, Technicka 2, 166 27 Prague 6, Czech Republic, 2003. 5.1, 5.2, 5.2, 2003.
- [28] Nicolas Privault. *Understanding Markov chains: examples and applications*. Springer Science & Business Media, 2013.
- [29] Luis Rincón. Introducción a los procesos estocásticos. *Departamento de Matemáticas, Facultad de Ciencias UNAM*, 2012.
- [30] Sheldon M Ross. *Introduction to probability models*. Academic press, 2014.
- [31] R. Salgado-García and Cesar Maldonado. Normal-to-anomalous diffusion transition in disordered correlated potentials: From the central limit theorem to stable laws. *Phys. Rev. E*, 88:062143, Dec 2013.
- [32] Ikaro Silva and George B Moody. An open-source toolbox for analysing and processing physionet databases in matlab and octave. *Journal of open research software*, 2(1), 2014.
- [33] David Stirzaker. *Stochastic processes and models*. Oxford University Press, 2005.
- [34] Howard M Taylor and Samuel Karlin. *An introduction to stochastic modeling*. Academic press, 2014.
- [35] M Vallverdu, M Palacios, and P Caminal. Modeling the dynamics of the heart rate variability by hidden markov models. In *Computers in Cardiology, 2003*, pages 461–464. IEEE, 2003.
- [36] Medical Videos. Ekg explanation. url<https://www.medicalvideos.com/>, 2007.
- [37] J Welch, P Ford, R Teplick, and R Rubsamen. The massachusetts general hospital-marquette foundation hemodynamic and electrocardiographic database—comprehensive collection of critical care waveforms. *Clinical Monitoring*, 7(1):96–97, 1991.
- [38] Wikipedia. Aparato de electrocardiógrafo. url<https://es.wikipedia.org/wiki/Electrocardiograma>, 2012.



# Apéndice A

## Espacio de probabilidad, variable aleatoria y entropía

### A.1. Espacio de probabilidad

**Definición 11** ( $\sigma$ -álgebras). Sea  $\Omega$  un espacio muestral. Una colección  $\mathcal{F} \subset \Omega$  es una  $\sigma$ -álgebra si satisface:

1.  $\Omega \in \mathcal{F}$ .
2. Si  $A \in \mathcal{F}$ , entonces  $A^c \in \mathcal{F}$ .
3. Si  $A_1, A_2, \dots, A_n \in \mathcal{F}$ , entonces  $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ .

A la dupla  $(\Omega, \mathcal{F})$  se le llama espacio medible y a los elementos de  $\mathcal{F}$  se les llama eventos o conjuntos medibles.

**Definición 12** (Medida de probabilidad). Sea  $(\Omega, \mathcal{F})$  un espacio medible. Una medida de probabilidad es una función  $P : \mathcal{F} \rightarrow [0, 1]$  que satisface

1.  $P(\Omega) = 1$ .
2.  $P(A) \geq 0$ , para cualquier  $A \in \mathcal{F}$ .
3. Si  $A_1, A_2, \dots, A_n \in \mathcal{F}$  son ajenos dos a dos, esto es,  $A_n \cap A_m = \emptyset$  para  $n \neq m$ , entonces  $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$ .

**Definición 13** (Espacio de probabilidad). Un espacio de probabilidad es la terna  $(\Omega, \mathcal{F}, P)$ , en donde  $\Omega$  es un espacio muestral,  $\mathcal{F}$  es una  $\sigma$ -álgebra de subconjuntos de  $\Omega$  y  $P$  es una medida de probabilidad definida sobre  $\mathcal{F}$ .

### A.2. Variable aleatoria

**Definición 14** ( $\sigma$ -álgebra generada). Sea  $\mathcal{L}$  una colección de subconjuntos de  $\Omega$ . La  $\sigma$ -álgebra generada por  $\mathcal{L}$ , denotada por  $\sigma(\mathcal{L})$ , es la colección

$$\sigma(\mathcal{L}) = \bigcap \{ \mathcal{F} : \mathcal{F} \text{ es } \sigma\text{-álgebra y } \mathcal{L} \subset \mathcal{F} \}$$

**Definición 15** ( $\sigma$ -álgebra de Borel de  $\mathbb{R}$ ). *Considérese una colección de intervalos abiertos  $(a, b)$  de  $\mathbb{R}$ , en donde  $a \leq b$ . A la  $\sigma$ -álgebra generada por esta colección se le llama  $\sigma$ -álgebra de Borel de  $\mathbb{R}$  y se le denota por  $\mathcal{B}(\mathbb{R})$ :*

$$\mathcal{B}(\mathbb{R}) = \sigma\{(a, b) \subseteq \mathbb{R} : a \leq b\}$$

A los elementos de  $\mathcal{B}(\mathbb{R})$  se les llama conjuntos de Borel.

**Definición 16** (Variable aleatoria). *Una variable aleatoria real es una función  $X : \Omega \rightarrow \mathbb{R}$  tal que para cualquier conjunto Boreliano  $B$ , se cumple que el conjunto  $X^{-1}B$  es un elemento de  $\mathcal{F}$*



## Apéndice B

### Obtención de la línea isoelectrica

```
function[vector_L_isoelectrica]=Line_Iso(a,Pw)
mo=1;
if a<=20
    mf=450000;
    step=2;
elseif a>=21
    mf=9000000;
    step=40; end
v_datos=Nread_sanos_40(a,mo,mf); %Adquisición de datos
n=0; n2=0; n3=0; n_identf=0; m=0; %inicialización
conta=0; r=0; r2=0; g=0; %inicialización
for k_id=1:step:mf %ciclo para obtención de muestra
    n_identf=n_identf + 1;
    c_identf(n_identf)=k_id;
    ident_ondas(n_identf)=v_datos(k_id);
end
v=ident_ondas; %"v" contiene la muestra
for i=1:length(v)-1 %Ciclo para índices de patrón ordinal
    if v(i)~=v(i+1) && v(i)<v(i+1)
        n=n+1;
        wave(n)=v(i); %patrón ordinal x_1<v_2...x_i<x_{i+1}
        c(n)=i; end
        if v(i)==v(i+1) || v(i)>v(i+1)
            continue;
        end
    end
end
end
```

---

```

%Este ciclo encuentra los índices donde
%termina el patrón ordinal
for j=1:1:n-1
    dif=c(j+1)-c(j);
    if abs(dif)>1
        conta=conta+1;
        z1(conta)=c(j);
        z2(conta)=c(j+1);
    end
end

%Este ciclo separa los vectores con patrones
%ordinales para las distintas ondas del ECG,
%en función del potencial eléctrico
for k=1:1:length(z1)+1
    if k==1
        vector_ondas=v(c(1):z1(k));
        M=c(1)*step;
        [max1,max2]=max(vector_ondas);
        pre_Max=c(1)+max2-1;
        Max=step*pre_Max;
    elseif k>=2 && k<=length(z1)
        vector_ondas=v(z2(k-1):z1(k));
        M=z2(k-1)*step;
        [max1,max2]=max(vector_ondas);
        pre_Max=z2(k-1) + max2 -1;
        Max=step*pre_Max;
    elseif k>length(z1)
        vector_ondas=v(z2(k-1):c(end));
        M=z2(k-1)*step;
        [max1,max2]=max(vector_ondas);
        pre_Max=z2(k-1) + max2 -1;
        Max=step*pre_Max;
    end
end

```

---

```
%Encuentra el voltaje total de cada onda
dif_volt=max(vector_ondas)-min(vector_ondas);

if dif_volt>=0 && dif_volt<=PR
    continue;

elseif abs(dif_volt)>PR && abs(dif_volt)<3

    r2=r2+1;
    SampleR=Max;
    SampleRR(r2)=Max;
    vector_instantsR(r2)=SampleR*(1/360);
    vector_pts_ini_B(r2)=vector_ondas(1);
    vector_t_iniR(r2)=M;

elseif dif_volt>=3
    continue;
end
end
first_val=vector_pts_ini_B(1);
if first_val>0
    vector_pts_ini=vector_pts_ini_B-abs(first_val);
elseif first_val<0
    vector_pts_ini=vector_pts_ini_B+abs(first_val);
elseif first_val==0
    vector_pts_ini=vector_pts_ini_B;
end
%regresa el vector con la línea isoeléctrica
%alrededor del cero
vector_L_isoelectrica=vector_pts_ini;
```



# Apéndice C

## Probabilidad estacionaria.

---

```
limit_inf=15000*(1/15000); R1=1; R2=40;
for t=0.5:0.5:30 % ciclo para tiempo (en minutos)
    limit_sup=(15000/5)*(t); %es entre 5 por step2
    for i=1:1:40 %pacientes
        sample=i;
        symbolic_sec_s_pre=Step2_secu_simbol_2_sanos(sample);
        symbolic_sec_s=symbolic_sec_s_pre(limit_inf:limit_sup);
        X_stateS=0;
        for j=1:1:10
            for k=1:1:length(symbolic_sec_s)
                if j==symbolic_sec_s(k)
                    X_stateS=X_stateS+1;
                end
                prob=(sum(X_stateS))/length(symbolic_sec_s);
            end
            Probabilidades(j)=prob;
            X_stateS=0;
        end
        Pacientes(:,i)=Probabilidades;
        j=1; k=1;
    end
end
r=40; R1=R1+r; R2=R2+r;
end
```



# Apéndice D

## Probabilidad y matriz de transición

---

```
function[A]=step2_P_M_Transicion_ArrSec2(C,C2,C3)
limit_inf=21600*(C2);      %frecuencia de muestreo:360
limit_sup=(21600/6)*(C3); %frecuencia de muestreo:360
sample=C;

%Adquisición de datos
states_pre=Step2_secu_simbol_2_arritmia(sample);
states=states_pre(limit_inf:limit_sup);
%Inicialización de algunas variables
n1=0; n2=0; n3=0; n4=0; n5=0; n6=0; n7=0;
n8=0; n9=0; n10=0; i2=0;
N1=0;

for k1=1:1:10      %son los estados

    for k2=1:1:length(states)
        if k1==states(k2)
            N1=N1+1;
        end
    end
    freq_states(k1)=N1;
    N1=0;
end
```

---

```

for i=1:1:10
    %i2=i2+1;
    v=[0 0 0 0 0 0 0 0 0 0];          %0 para cada estado
        for j=2:1:length(states)

            if i==1
                if states(j-1)==i && states(j)==i
                    n1=n1+1;
                    v(1,1)=n1;
                elseif states(j-1)==i && states(j)==i+1
                    n2=n2+1;
                    v(1,2)=n2;
                elseif states(j-1)==i && states(j)==i+2
                    n3=n3+1;
                    v(1,3)=n3;
                elseif states(j-1)==i && states(j)==i+3
                    n4=n4+1;
                    v(1,4)=n4;
                elseif states(j-1)==i && states(j)==i+4
                    n5=n5+1;
                    v(1,5)=n5;
                elseif states(j-1)==i && states(j)==i+5
                    n6=n6+1;
                    v(1,6)=n6;
                elseif states(j-1)==i && states(j)==i+6
                    n7=n7+1;
                    v(1,7)=n7;
                elseif states(j-1)==i && states(j)==i+7
                    n8=n8+1;
                    v(1,8)=n8;
                elseif states(j-1)==i && states(j)==i+8
                    n9=n9+1;
                    v(1,9)=n9;
                elseif states(j-1)==i && states(j)==i+9
                    n10=n10+1;
                    v(1,10)=n10;
                end
            end
        end
    end
end

```



---

```
if i==2
    if states(j-1)==i && states(j)==i-1
        n1=n1+1;
        v(1,1)=n1;
    elseif states(j-1)==i && states(j)==i
        n2=n2+1;
        v(1,2)=n2;
    elseif states(j-1)==i && states(j)==i+1
        n3=n3+1;
        v(1,3)=n3;
    elseif states(j-1)==i && states(j)==i+2
        n4=n4+1;
        v(1,4)=n4;
    elseif states(j-1)==i && states(j)==i+3
        n5=n5+1;
        v(1,5)=n5;
    elseif states(j-1)==i && states(j)==i+4
        n6=n6+1;
        v(1,6)=n6;
    elseif states(j-1)==i && states(j)==i+5
        n7=n7+1;
        v(1,7)=n7;
    elseif states(j-1)==i && states(j)==i+6
        n8=n8+1;
        v(1,8)=n8;
    elseif states(j-1)==i && states(j)==i+7
        n9=n9+1;
        v(1,9)=n9;
    elseif states(j-1)==i && states(j)==i+8
        n10=n10+1;
        v(1,10)=n10;
    end
end
o
o
o
```

```
o
o
o
  if i==10
    if states(j-1)==i && states(j)==i-9
      n1=n1+1;
      v(1,1)=n1;
    elseif states(j-1)==i && states(j)==i-8
      n2=n2+1;
      v(1,2)=n2;
    elseif states(j-1)==i && states(j)==i-7
      n3=n3+1;
      v(1,3)=n3;
    elseif states(j-1)==i && states(j)==i-6
      n4=n4+1;
      v(1,4)=n4;
    elseif states(j-1)==i && states(j)==i-5
      n5=n5+1;
      v(1,5)=n5;
    elseif states(j-1)==i && states(j)==i-4
      n6=n6+1;
      v(1,6)=n6;
    elseif states(j-1)==i && states(j)==i-3
      n7=n7+1;
      v(1,7)=n7;
    elseif states(j-1)==i && states(j)==i-2
      n8=n8+1;
      v(1,8)=n8;
    elseif states(j-1)==i && states(j)==i-1
      n9=n9+1;
      v(1,9)=n9;
    elseif states(j-1)==i && states(j)==i
      n10=n10+1;
      v(1,10)=n10;
    end
  end
end

end
```

```
for c=1:1:10
    if frec_states(i)~=0
        A(i,c)=v(c)/frec_states(i);
    elseif frec_states(i)==0
        A(i,c)=0;
    end
end
%Reinicialización de variables
n1=0; n2=0; n3=0; n4=0; n5=0; n6=0; n7=0; n8=0; n9=0; n10=0;
end

end
```



# Apéndice E

## Estimación de la producción de entropía

```
C2_A=1/21600; C2_S=1/15000;
time=1:1:30;
for N=1:1:30 %tiempo (en minutos)
for k=1:1:40 %pacientes
%Obtención de la matriz de transición para el paciente k al instane N
P0_A=step2_P_M_Transicion_ArrSec2(k,C2_A,N);
P0_S=step2_P_M_Transicion_SanosSec2(k,C2_S,N);
P=P0_A;
PS=P0_S;
[s1 s2]=size(P); %% Aquí empiezan los pacientes arrítmicos
[U V W]=eig(P); %obtención de vector de probabilidad invariante
zz0=diag(V);
[zz2,z]=max(zz0);
dis_lim_pre=abs(W(:,z));
dis_lim=transpose(dis_lim_pre);
for M1=1:1:s1 %Asigna 0.00001 a los ceros de la matriz P
for M2=1:1:s2
if P(M1,M2)==0
P(M1,M2)=0.00001;
end
end
end
end %hasta aquí arritmia%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

---

```

%%Pacientes sanos%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
[Us,Vs,Ws]=eig(PS);

zzs=diag(Vs);
[zz2s,zs]=max(zzs);

dis_lim_pre_s=abs(Ws(:,zs));
dis_lim_s=transpose(dis_lim_pre_s);

for ms=1:1:s1
    if dis_lim_s(ms)==0
        dis_lim_s(ms)=0.000001;
    end
end
%este ciclo le asigna 0.00001 a los ceros de la matriz PS
for M1s=1:1:s1
    for M2s=1:1:s2
        if PS(M1s,M2s)==0
            PS(M1s,M2s)=0.00001;
        end
    end
end
end
%%hasta aquí sanos%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Calculo de la producción de entropía para el k-ésimo paciente
%al instante N
for j=1:1:s1
    for i=1:1:s2
        %arritmicos
e=(dis_lim(i)*P(i,j)-dis_lim(j)*P(j,i))*log((dis_lim(i)*P(i,j)) /
(dis_lim(j)*P(j,i)));
e2(i)=e;
        %sanos
es=(dis_lim_s(i)*PS(i,j)-dis_lim_s(j)*PS(j,i))*log((dis_lim_s(i)*PS(i,j)) /
(dis_lim_s(j)*PS(j,i)));
e2s(i)=es;
    end
    e3(j)=sum(e2);
    e2(1:i)=0;

    e3s(j)=sum(e2s);
    e2s(1:i)=0;
end
Conta_pacientes(k)=k;

```

---

```
%Guarda la producción de entropía para cada k paciente
%al instante N
entropy_production(k)=sum(e3);
e3(1:j)=0;

entropy_production_s(k)=sum(e3s);
e3s(1:j)=0;
end
zz2(1:k)=0;
zz2s(1:k)=0;

%Obtiene valor promedio y desviación estándar de la
%producción de entropía para todos los pacientesal instanter N
promedio_A(N)=mean(entropy_production);
desv_std_A(N)=std(entropy_production);
MPES2Arr(:,N)=entropy_production;
entropy_production(1:k)=0;

promedio_S(N)=mean(entropy_production_s);
desv_std_S(N)=std(entropy_production_s);
MPES2Sanos(:,N)=entropy_production_s;
entropy_production_s(1:k)=0;
end

%grafico con barra de error
figure(1)
errorbar(time,promedio_A,desv_std_A,'r');
hold on
errorbar(time,promedio_S,desv_std_S,'g');
title('Producción de entropía')
xlabel('Tiempo (min)')
ylabel('Producción de entropía')
legend('Pacientes arritmicos','Pacientes sanos')
```





# Apéndice F

## Estimador de irreversibilidad

---

```
%PARTE 1: generar las permutaciones
s=[1 2 3 4 5 6 7 8];
z=combnk(s,3); %z es de de 56 x 3: 8C3=56
[c1,c2]=size(z);
r1=1; r2=3; %inicialización de algunas variables
for i=1:1:c1
    c=z(i,:); %obtiene la combinación del renglón i
    A=perms(c); %obtiene todas las permutaciones de c,
    A2=A((1:3),:);
    A_permu((r1:r2),:)=A2;
    r=3;
    r1=r1+r; r2=r2+r;
end
[w1,w2]=size(A_permu);
m5=0; %%PARTE 2: Cálculo del estimador de irreversibilidad
for paciente=21:1:40
    limit_inf=1;
    limit_sup=30;
    %obtención de la secuencia
    sec=Secu_simbol(paciente,limit_inf,limit_sup);
    m3=0; m4=0; %inicialización de algunas variables
    for j=1:1:w1 %llega hasta el número de renglones de A_permu
        a=A_permu(j,:); %toma la permutación del renglón i de A_permu.
        ar=[a(3),a(2),a(1)]; %genera el reverso de a.
        m1=0; m2=0;
```

---

```

for k=1:1:length(sec)-2
v=[sec(k),sec(k+1),sec(k+2)];
m1=m1+1; m2=m2+1;
compare_a(m1)=isequal(v,a);
compare_ar(m2)=isequal(v,ar);
end

frec_a=sum(compare_a); %frecuencia de a
frec_ar=sum(compare_ar); %frecuencia de ar
if frec_a>=1 && frec_ar>=1
    m3=m3+1;
    prob_a=frec_a/length(sec)-2;
    prob_ar=frec_ar/length(sec)-2;
    S_k_pre(m3)=abs(log(prob_a/prob_ar));
end
if frec_a<1 || frec_ar< 1
    % m4=m4+1;
    % S_k_pre(m4)=1000;
    continue;
end

compare_a(1:k)=0;
compare_ar(1:k)=0;

end

if m3==0
    continue;
end

if m3>=1
m5=m5+1;
S_k=S_k_pre(1:m3);
Sk_promedio_paciente_k=mean(S_k);
S_k_PACIENTES(m5)=Sk_promedio_paciente_k;
M=length(S_k_pre);
S_k_pre(1:M)=0;
end
end

```