

# Sobre la Noción de Distancia en Física Estadística y Mecánica Cuántica

por Ana Paula Majtey

Presentado ante la Facultad de Matemática, Astronomía y Física como parte de los requerimientos para la obtención del grado de Doctora en Física de la

UNIVERSIDAD NACIONAL DE CORDOBA

septiembre de 2006

©FaMAF - UNC 2006

Director: Pedro Walter Lamberti

*A mis padres que me enseñaron a andar...*

*a Damián que recorre conmigo.*

# Resumen

Estudiamos el concepto de distancia entre distribuciones de probabilidad y sus aplicaciones en problemas fundamentales de la Física y la Biología. El eje central de este trabajo de tesis lo constituye la divergencia de Jensen-Shannon (DJS), una medida de distancia entre distribuciones de probabilidad proveniente de la Teoría de la Información. Proponemos dos extensiones de esta cantidad: una en el contexto de la Mecánica Estadística no Extensiva y la otra en el marco de la Mecánica Cuántica.

A partir de la generalización de la DJS en el marco de la Mecánica Estadística no Extensiva, definimos una familia monoparamétrica de métricas asociada al parámetro de deformación  $q$ . Aplicamos esta familia de métricas al estudio estadístico de secuencias simbólicas, analizando las principales propiedades en este contexto.

Partiendo de un proceso de segmentación entrópico basado en la DJS, definimos una medida de complejidad para secuencias simbólicas y la aplicamos a secuencias de ADN. Con esta medida estudiamos la presencia de correlaciones de largo alcance en secuencias genómicas y su relación con aspectos evolutivos.

Partiendo de una medida de distancia entre secuencias genéticas basada en la complejidad de Kolmogorov estudiamos un método combinado con el proceso de segmentación previamente mencionado para realizar estudios filogenéticos y lo aplicamos a una familia de mamíferos.

Finalmente, en el contexto mecánico cuántico, estudiamos en detalle el problema de distinguibilidad entre estados por medio de la DJS tanto para estados puros como para estados mezclas. Establecemos relaciones entre la DJS y algunas medidas de distancia generalmente utilizadas en Mecánica Cuántica. Estudiamos aspectos de esta medida en lo que respecta a la cuestión de la accesibilidad de información en Teoría de la Información Cuántica. A partir de la DJS proponemos una nueva medida de entrelazamiento cuántico y una medida de fidelidad.

PACS: 05.20.-y, 03.65.-w, 03.67.-a, 87.10.+e, 89.70+c, 64.60.Cn, 05.45.+b



# Abstract

In this work we study the notion of distance between probability distributions and their applications to fundamental problems of Physics and Biology. We deal with a particular distance between probability distributions, the Jensen Shannon divergence (JSD). This distance arises in Information Theory and has several interesting properties. Here we propose two extensions of the JSD: the first one in the context of the non-extensive Statistical Mechanics; the second one in the realm of Quantum Theory.

From the non-extensive extension we can introduce a monoperametric family of true metrics for Statistical Mechanics. We focus on the applications of this family to the analysis of symbolic sequences.

Starting from a segmentation procedure we introduce a complexity measure for symbolic sequences. We apply it to the analysis of DNA sequences. By combining this segmentation scheme and a distance measure introduced by M. Li and collaborators, we can make an approach to phylogenetics studies.

Lately we use the JSD to the fundamental problem of distinguishability between quantum states. We investigate it for pure and mixed states. We show that the JSD works as a unifying distance measure with many interesting properties. In particular we show that the proposed measure allows to introduce a new fidelity and entanglement and can be physically interpreted as an upper bound for the accessible information in a quantum communication protocol.



# Índice general

<b>1. Introducción</b>	<b>11</b>
1.1. Distancias en Mecánica Estadística . . . . .	12
1.2. Distancias Genéticas . . . . .	16
1.3. Distancias en Física Cuántica . . . . .	19
1.4. Mapa de esta Tesis . . . . .	21
<b>2. La Divergencia de Jensen-Shannon</b>	<b>23</b>
2.1. Información y Física . . . . .	23
2.1.1. Un enfoque probabilístico . . . . .	23
2.1.2. El enfoque físico . . . . .	24
2.2. Derivación de la DJS . . . . .	24
2.3. Interpretaciones de la DJS . . . . .	26
2.3.1. La DJS como entropía intensiva de la mezcla . . . . .	26
2.3.2. La DJS como información mutua . . . . .	27
<b>3. La <math>q</math>-Divergencia de Jensen-Shannon</b>	<b>29</b>
3.1. Mecánica estadística no-extensiva . . . . .	29
3.2. Definición y propiedades . . . . .	31
3.3. La $q$ -DJS aplicada al análisis de secuencias - Segmentación . . . . .	34
3.3.1. Propiedades Estadísticas de la $q$ -DJS: . . . . .	35
3.3.2. Nivel de significancia . . . . .	38
3.4. Propiedad de Ramificación . . . . .	40
3.5. El carácter métrico de la $q$ -DJS . . . . .	41
3.6. Comentarios Finales . . . . .	45
<b>4. Complejidad de Secuencias Simbólicas</b>	<b>47</b>
4.1. Introducción . . . . .	47
4.2. Sistemas complejos . . . . .	49
4.3. Una medida de complejidad . . . . .	50
4.3.1. Método de segmentación . . . . .	51
4.3.2. Definición de la complejidad . . . . .	52
4.4. Aplicaciones y Resultados . . . . .	55
4.5. Comentarios Finales . . . . .	60

<b>5. Distancias y Filogénesis</b>	<b>63</b>
5.1. Ideas para el análisis filogenético . . . . .	64
5.1.1. Manipulación de las secuencias . . . . .	64
5.1.2. El ADN mitocondrial . . . . .	65
5.2. Contrucción de los árboles . . . . .	67
5.2.1. Esquema de cálculo . . . . .	67
5.2.2. Motivación . . . . .	67
5.2.3. Materiales y métodos . . . . .	69
5.3. Comentarios Finales . . . . .	70
<b>6. La Divergencia de Jensen-Shannon Cuántica</b>	<b>75</b>
6.1. Estados y operaciones en Física Cuántica . . . . .	76
6.2. Parte I: Estados puros . . . . .	78
6.2.1. Introducción . . . . .	78
6.2.2. Distancia Wootters . . . . .	79
6.2.3. Distancia Hellinger . . . . .	81
6.2.4. Distancia Bhattacharyya . . . . .	82
6.2.5. Métrica Fubini-Study . . . . .	83
6.2.6. Un nuevo criterio de distinguibilidad basado en la DJS . . . . .	84
6.3. Parte II: Estados mezcla . . . . .	89
6.3.1. La entropía relativa . . . . .	90
6.3.2. Una alternativa a la entropía relativa: La DJS Cuántica . . . . .	92
6.3.3. La DJSC y su relación con otras medidas de distancia . . . . .	95
6.3.4. Ejemplo de evaluación de la DJSC . . . . .	96
6.3.5. La DJSC y la TI cuántica . . . . .	96
6.3.6. Accesibilidad de la información cuántica . . . . .	97
6.3.7. Entrelazamiento . . . . .	98
6.3.8. Fidelidad . . . . .	100
6.4. Comentarios finales . . . . .	102
<b>A modo de conclusiones ...</b>	<b>103</b>
¿Qué queda por hacer? . . . . .	103
<b>Bibliografía</b>	<b>105</b>
<b>Apéndices</b>	<b>114</b>
<b>A. Palabras en secuencias</b>	<b>115</b>
<b>B. Complejidad máxima</b>	<b>117</b>
<b>Índice de Figuras</b>	<b>120</b>
<b>Publicaciones</b>	<b>122</b>







# Capítulo 1

## Introducción

*Los que olvidaban su deber de calibrar el patrón de longitud las noches de luna llena, se enfrentaban a la pena de muerte. Ese era el peligro que corría el grupo de arquitectos reales, responsables de la construcción de templos y pirámides en el antiguo Egipto de los Faraones, 3000 años a.C. El primer codo real fue definido como la longitud del antebrazo del Faraón, desde el codo hasta el extremo del dedo medio, teniendo la mano extendida, más la anchura de su mano. La medida original fue transferida y materializada en granito negro. En los lugares de construcción, los trabajadores poseían copias en granito o madera, siendo responsabilidad de los arquitectos su mantenimiento.*

A lo largo de los últimos ciento cincuenta años ha habido una marcada tendencia a la formulación geométrica de las distintas teorías físicas. Esto tuvo su clímax durante el siglo XX con la formulación de la Teoría General de la Relatividad de Einstein y con la descripción de la microfísica en el marco de la teoría de los espacios de Hilbert. Desde los trabajos pioneros de Gauss y Riemann quedó en claro la relación existente entre las propiedades geométricas de un espacio y la noción de distancia entre los puntos de dicho espacio. La idea de geometrización permeó finalmente en prácticamente todos los ámbitos de la Física lo cual tuvo, como una consecuencia natural, la introducción de medidas de distancias en los más variados contextos temáticos de esta disciplina científica. Precisamente, el presente trabajo de tesis tiene como propósito el estudio del rol del concepto de distancia en el marco de la Mecánica Estadística y de la Mecánica Cuántica. Es de destacar que estas teorías logran su significación física por medio del concepto matemático de probabilidad. Por ello, el *leit motiv* del presente trabajo es la noción de distancia entre distribuciones de probabilidad.

En este capítulo discutiremos algunos ejemplos en donde la noción de distancia entre distribuciones de probabilidad resulta de importancia. Antes de ello, definiremos lo que entenderemos por una distancia definida sobre un espacio arbitrario.

Sea  $\mathcal{X}$  un conjunto abstracto. Una función

$$d : \mathcal{X} \times \mathcal{X} \rightarrow \mathfrak{R}$$

es una distancia definida sobre el conjuntos  $\mathcal{X}$ , si cumple con las siguientes propiedades:

1.  $d(x, y) > 0$  para  $x \neq y$  (1.1)
2.  $d(x, x) = 0$
3.  $d(x, y) = d(y, x)$

Si además  $d$  satisface la desigualdad triangular:

$$4. \quad d(x, y) \leq d(x, z) + d(z, y) \quad (1.2)$$

diremos que  $d$  es una métrica sobre el espacio  $\mathcal{X}$ .

Al estudiar distancias sobre distribuciones de probabilidad el espacio  $\mathcal{X}$  es precisamente el conjunto de distribuciones de probabilidad,  $M_+^1(N) = \{p_i \in \mathfrak{R} / p_i \geq 0 \text{ y } \sum_i p_i = 1\}^1$ . Sin embargo, también trabajaremos con distancias definidas sobre los *espacios de configuración del sistema*. Por ejemplo, en el capítulo dedicado al estudio de la noción de distancia en Mecánica Cuántica, además del conjunto  $M_+^1(N)$ , tendremos el espacio de Hilbert de los estados del sistema físico y la correspondiente distancia entre ellos, la cual no coincidirá, en general, con la inducida por el producto interno del espacio de Hilbert.

A continuación realizaremos una revisión de distintas medidas de distancias introducidas en diversos contextos de interés. Las medidas aquí comentadas no agotan, ni mucho menos, el listado completo de todas las distancias propuestas. Sólo nos referiremos a las que tienen más relevancia para los contextos temáticos analizados en los capítulos siguientes, y que son de fundamental interés para nuestro trabajo.

En este trabajo usaremos  $\ln$  para indicar el logaritmo natural y  $\log$  para indicar el logaritmo en base 2.

## 1.1. Distancias en Mecánica Estadística

- *La Noción de Estabilidad:* Lesche destacó una propiedad interesante de la entropía de Boltzmann-Gibbs, conocida como *estabilidad* [1]. La entropía de Boltzmann-Gibbs está definida por la expresión

$$H(P) = -k_B \sum_j p_j \log p_j, \quad (1.3)$$

---

<sup>1</sup>En general el espacio de distribuciones de probabilidad puede representarse por  $M_+^1(A)$ , donde  $A$  es un conjunto provisto con alguna  $\sigma$ -álgebra.

donde  $p_j$  es la probabilidad del estado  $j$  –ésimo y  $k_B$  es la constante de Boltzmann. Se dice que un funcional acotado de una distribución de probabilidad  $P = \{p_j\}$ ,  $\Gamma(P)$ , es estable si y sólo si, para cualquier  $\epsilon > 0$  existe  $\delta_\epsilon > 0$  tal que

$$d(P, P') < \delta_\epsilon \Rightarrow \frac{\Gamma(\{p_j\}) - \Gamma(\{p'_j\})}{\Gamma_{max}} < \epsilon,$$

donde  $d(P, P')$  es la distancia  $d(P, P') = \sum_i |p_i - p'_i|$  y  $\Gamma_{max}$  es el valor máximo que toma el funcional  $\Gamma$ .

En este contexto, Lesche mostró que la entropía de Boltzmann-Gibbs es estable. Sin embargo la entropía de Renyi

$$H_R^\alpha = \log \sum_j \frac{p_j^\alpha}{\alpha - 1}$$

no lo es para  $\alpha \neq 1$ .

Un punto que ha sido investigado recientemente es la dependencia del concepto de estabilidad con la distancia  $d(P, P')$  utilizada en su definición. También se ha investigado el concepto de estabilidad para otras entropías no extensivas [2].

- *Definición del concepto de Complejidad:* La medida de complejidad estadística introducida por Lopez-Ruiz, Mancini y Calvet, se basa en la noción de *desequilibrio* [3]. Esta medida de complejidad estadística está dada por:

$$C_{LMC} = QH,$$

donde  $Q$  es el desequilibrio,

$$Q(\{p_j\}) = Q_0 D(\{p_j\}, \{p_j^{(e)}\}),$$

con  $D$  una medida apropiada de distancia entre la distribución  $P = \{p_j\}$  y la distribución de equilibrio  $P^{(e)} = \{p_j^{(e)}\}$ , la cual, para un sistema con  $N$  estados posibles vale  $P^{(e)} = (1/N, \dots, 1/N)$ .  $Q_0$  es una constante de normalización y  $H$  es la entropía (1.3).

En el trabajo original de Lopez-Ruiz, Mancini y Calvet, el desequilibrio  $D$  se toma como el cuadrado de la distancia euclídea, es decir

$$D(P, P^{(e)}) = \sum_{j=1}^N (p_j - 1/N)^2.$$

Recientemente se ha investigado el uso de una medida alternativa de desequilibrio, en la cual se modifica la distancia  $D$  [4, 5].

- *Estudio Estadístico de Secuencias Simbólicas*: Un problema de gran importancia teórica y práctica es el de la comparación de secuencias simbólicas. Áreas tan dispares como las comunicaciones digitales y los estudios filogenéticos, requieren de tal comparación. En la sección siguiente haremos una descripción detallada de varias de las distancias introducidas en este último contexto (distancias genéticas). Sin embargo, y a los fines puramente enumerativos, describiremos aquí dos de las distancias introducidas originalmente en el marco de la comparación de secuencias binarias.

Una de las maneras de cuantificar las “diferencias” entre secuencias de bits es por medio de la distancia Hamming [6]. Esta distancia se define como el número de posiciones para las cuales los símbolos son diferentes, y mide el número de sustituciones que son necesarias para cambiar una secuencia en la otra. Obviamente esta distancia sólo es válida para secuencias de la misma longitud. La distancia Hamming fue introducida en el contexto de la detección y la corrección de errores y se utiliza en telecomunicaciones para contar el número de bits alterados en una palabra binaria de longitud fija. Para cada longitud la distancia Hamming es una métrica sobre el espacio de vectores de palabras de esa longitud. Sin embargo, para comparar secuencias de longitudes diferentes, o cuando se consideran además de las sustituciones, las inserciones o el borrado de símbolos, es necesario usar una distancia más sofisticada como por ejemplo la métrica Levenshtein [7]. Ésta está dada por el número de operaciones necesarias para transformar una cadena en la otra de la misma forma que la distancia Hamming, pero en este caso la comparación permite agregar, quitar, o sustituir caracteres.

- *Teoría de la Información Clásica*: Una de las grandes creaciones intelectuales del siglo XX fue, sin lugar a dudas, la Teoría de la Información (TI) desarrollada por Claude Shannon [8]. Si bien originalmente su desarrollo estuvo relacionado con la formulación matemática de la teoría de comunicaciones, rápidamente sus ideas se incorporaron a otros ámbitos del conocimiento. Tempranamente las principales ideas de la TI fueron incorporadas a la Física, y en particular a la Mecánica Estadística, a finales de la década de 1950, en los trabajos de E.T. Jaynes [9, 10].

Un concepto central en TI es el de la entropía. La entropía de una variable aleatoria  $X$ , con distribución de probabilidad  $p(x)$ , está definida por

$$H(X) = - \sum_x p(x) \log p(x) \quad (1.4)$$

Ésta es una medida de la incerteza promedio en la variable aleatoria;  $H(X)$  es el número promedio de bits necesarios para describir la variable  $X$ .<sup>2</sup>

Supongamos que cierta variable aleatoria  $X$  (asociada quizás con algún fenómeno físico) tiene una distribución de probabilidades “verdadera”,  $p(x)$ , y que por determinado experimento, se infiere (erróneamente) que la distribución de probabilidad

---

<sup>2</sup>La unidad bit es la que resulta de usar el logaritmo en base 2.

para esa variable es  $q(x)$ . Podemos preguntarnos entonces sobre el modo de cuantificar la ineficiencia cometida en la descripción de la variable  $X$  al suponer la distribución  $q(x)$  en lugar de la distribución correcta  $p(x)$ . La respuesta a esta pregunta está precisamente en la diferencia:

$$\begin{aligned} K(P, Q) &= -\sum p(x) \log q(x) - \left(-\sum p(x) \log p(x)\right) \\ &= \sum p(x) \log \frac{p(x)}{q(x)} \end{aligned} \quad (1.5)$$

Esta cantidad es conocida como la divergencia de Kullback-Leibler y sólo cumple con las condiciones 1) y 2) de la definición de distancia. Volveremos a esta cantidad en el capítulo siguiente.

- *Inferencia Estadística:* Desde los trabajos de R.A. Fisher [11] de comienzo del siglo pasado, el problema de la inferencia estadística, es decir, la obtención de la ley de distribución de probabilidad que sigue determinado fenómeno aleatorio, ha sido encarado con diversos tratamientos formales. Muchas veces, la determinación de esa ley requiere de un gran refinamiento en la capacidad de discriminación entre dos distribuciones de probabilidad próximas. Estos desarrollos han conducido a la introducción de diversas distancias entre distribuciones de probabilidad. Fue el propio Fisher quien, al definir una manera de discriminar entre una distribución de probabilidad continua  $p(x)$  y su versión desplazada  $p(x+\delta)$ , introdujo la cantidad (conocida en la actualidad como información de Fisher):

$$I[p(x)] = \int \frac{[dp(x)/dx]^2}{p(x)} dx \quad (1.6)$$

Otro ejemplo emergente del problema de discriminación entre distribuciones de probabilidad es la *distancia traza*, que se define por la ecuación:

$$d_{Tr}(P^{(1)}, P^{(2)}) = \frac{1}{2} \sum_i |p_i^{(1)} - p_i^{(2)}|. \quad (1.7)$$

Esta cantidad es conocida como distancia  $L_1$  o distancia Kolmogorov y es una métrica sobre las distribuciones de probabilidad.

Otra medida de distancia puede derivarse del concepto de *fidelidad*:

$$F(P^{(1)}, P^{(2)}) = \sum_i \sqrt{p_i^{(1)} p_i^{(2)}}. \quad (1.8)$$

Cuando las distribuciones  $P^{(1)}$  y  $P^{(2)}$  son idénticas,  $F(P^{(1)}, P^{(2)}) = \sum_i p_i^{(1)} = 1$ . Naturalmente la cantidad  $d_F \equiv 1 - F$  cumple con las propiedades de una distancia. La interpretación geométrica de la fidelidad es muy simple: es el producto interno de vectores sobre la esfera unidad con componentes  $\sqrt{p_i^{(1)}}$  y  $\sqrt{p_i^{(2)}}$ .

En este trabajo de tesis se hace una aplicación sistemática del concepto de distancia entre distribuciones de probabilidad, tanto al estudio de secuencias simbólicas, en particular secuencias de ADN, como al problema de distinguibilidad entre estados cuánticos. Es por ello que dedicaremos las siguientes dos secciones de este capítulo a realizar una revisión de las diferentes distancias introducidas en estos contextos. Un desarrollo más exhaustivo se realizará en los siguientes capítulos.

## 1.2. Distancias Genéticas

En los últimos años ha habido un creciente interés en el estudio estadístico de las secuencias simbólicas, particularmente de las secuencias de ADN [12]. Muchos de los resultados emergentes de estos estudios han sido aplicados a otros campos, como por ejemplo, el de la lingüística [13]

Con muchos genomas ya secuenciados, entre ellos el genoma humano [14, 15] y la finalización próxima de otros proyectos de secuenciación [16], surge una pregunta fundamental en la biología moderna: ¿Cómo comparar dos genomas? Varias técnicas han sido propuestas para encarar este problema [17], [18], involucrando muchas de ellas la introducción de distancias entre secuencias simbólicas.

Como se indicó más arriba, algunas de las medidas de distancias definidas entre secuencias se derivan examinando pares de secuencias **alineadas** símbolo por símbolo (i.e., nucleótido por nucleótido en el caso del ADN) y estimando el número de sustituciones necesarias para pasar de una secuencia a otra. Los métodos que involucran alineamientos tienen algunos inconvenientes, como por ejemplo, se limitan al estudio de segmentos homólogos (i.e., segmentos que se repiten en cada una de las secuencias analizadas) y son, en general, de un alto costo computacional.

Un enfoque alternativo, es el desarrollo de medidas que no requieren del alineamiento previo de las secuencias. Estos métodos *libres de alineamiento*, se pueden clasificar a su vez, en dos categorías: los basados en un *ranking* o clasificación de las frecuencias de las palabras, o más en general, de  $n$ -*tuplas*, y los que no requieren del estudio de las secuencias en segmentos de longitudes fijas (*libre de resolución*).

### Métodos basados en las frecuencias de las palabras

Es posible asignar un vector a una secuencia, haciendo que cada componente de ese vector corresponda al número de veces que aparece cada una de las  $n$ -*tuplas* presentes en ella. Estas  $n$ -*tuplas* se definen previamente y su longitud determina la resolución con la que se estudia la secuencia. Este proceso de vectorización permite aplicar las herramientas del álgebra lineal y de la estadística, al estudio de las secuencias. Los vectores representan la secuencia original con una resolución fija dada por el tamaño de la palabra considerada,  $n$ . Ver Apéndice A.

La idea básica detrás de la comparación de secuencias, es que secuencias similares compartirán muchas palabras, es decir, tendrán una composición similar de palabras. Esto



se puede cuantificar por medio de varias técnicas. Si bien estos son métodos libres de alineamiento, dependen de la resolución, ya que los resultados cambian con el largo de las palabras consideradas. Existen métodos para derivar distancias combinadas que tienen en cuenta la información a diferentes resoluciones, con lo cual se logra independencia del largo de los segmentos que se utilizan para la vectorización.

Una forma alternativa para cuantificar la disimilaridad entre secuencias, es modelarlas como cadenas de Markov [19]. Recordemos que una cadena de Markov es una serie de eventos, en la cual la probabilidad de que ocurra un evento depende del evento inmediato anterior, y esta probabilidad está dada por la matriz de transición que caracteriza a cada cadena. La diferencia entre las secuencias se cuantifica evaluando la distancia euclídea entre sus matrices de transición. En este formalismo la matriz de transición se puede identificar con todas las  $n$ -*tuplas*, para cada resolución o longitud de palabra  $n$ . Entonces el cuadrado de la distancia euclídea entre dos secuencias  $\mathcal{S}_1$  y  $\mathcal{S}_2$ , está dado:

$$d_n^E(\mathcal{S}_1, \mathcal{S}_2) = \sum_i \left( C_{n,i}^{(1)} - C_{n,i}^{(2)} \right)^2, \quad (1.9)$$

donde  $C_{n,i}^{(k)}$  es la cantidad de  $n$ -*tuplas* del tipo  $i$  que hay en la secuencia  $k$ , ( $k = 1, 2$ ).

La posibilidad de que la frecuencia de diferentes palabras pueda tener un impacto distinto en la distancia euclídea, llevó a la derivación de medidas pesadas. Las distancias pesadas se combinan sumando con diferentes pesos diferentes resoluciones. Por ejemplo si se consideran desde  $m$ -*tuplas* hasta  $u$ -*tuplas*, se tiene:

$$d^P(\mathcal{S}_1, \mathcal{S}_2) = \sum_{n=m}^u \sum_i \rho_n \left( C_{n,i}^{(1)} - C_{n,i}^{(2)} \right)^2 \quad (1.10)$$

Obviamente una vez que se establece la conversión de secuencias en vectores con el número de  $n$ -*tuplas*, es posible trabajar con distintas distancias, siendo la euclídea sólo un ejemplo. Entre las alternativas a ésta se destaca la distancia Mahalanobis [20].

Los métodos comentados previamente se basan en medidas de distancias estadísticas entre vectores frecuencia. Este formalismo de  $n$ -*tuplas* se puede usar también aplicando medidas de distancia que se originan en TI [21]. Por ejemplo, si se usa la divergencia Kullback-Leibler (1.5), y se denota por  $f_{n,i}^{(1)}$  y  $f_{n,i}^{(2)}$  a las frecuencias de ocurrencia de la  $n$ -*tupla*  $i$  en las secuencia  $\mathcal{S}_1$  y  $\mathcal{S}_2$  respectivamente, tal discrepancia se escribe

$$d_n^{KL}(\mathcal{S}_1, \mathcal{S}_2) = \sum_i f_{n,i}^{(1)} \cdot \log \left( \frac{f_{n,i}^{(1)}}{f_{n,i}^{(2)}} \right) \quad (1.11)$$

Un enfoque similar lo dan Yang *et al.* [22] quienes confeccionan una clasificación determinando la frecuencia de cada  $n$ -*tupla* por medio de una ventana que se desliza a través de la secuencia completa y luego ordenándola en forma decreciente. La diferencia entre dos secuencias, o textos, se puede visualizar graficando el número que ocupa cada palabra compartida, en la clasificación de cada texto. Las palabras centradas en la diagonal principal indican que el número en la clasificación es muy similar en ambos textos. Por el contrario,

si se comparan textos muy diferentes, las palabras se encuentran más dispersas y alejadas de la diagonal. Esto permite, por ejemplo, reconocer dos obras literarias o composiciones musicales, del mismo autor. La distancia entre dos textos se puede cuantificar midiendo la dispersión de estos puntos de la línea diagonal en el gráfico del orden en la clasificación. La disimilitud entre dos secuencias usando  $n$ -*tuplas* de símbolos se define como:

$$D_n(\mathcal{S}_1, \mathcal{S}_2) = \frac{1}{N-1} \sum_{k=1}^N |R_1(w_k) - R_2(w_k)| \frac{H_1(w_k) + H_2(w_k)}{\sum_{k=1}^N [H_1(w_k) + H_2(w_k)]}, \quad (1.12)$$

donde  $R_1(w_k)$  y  $R_2(w_k)$  representan la posición en la clasificación de una  $n$ -*tupla* específica  $w_k$  en las secuencias  $\mathcal{S}_1$  y  $\mathcal{S}_2$  respectivamente.  $H_{1(2)}$  es la entropía de Shannon (1.4) para  $w_k$  en las secuencias  $\mathcal{S}_1$  y  $\mathcal{S}_2$  y mide la información de cada  $n$ -*tupla* en cada secuencia. Con esto, la  $n$ -*tupla* más frecuente, contribuye más a la medida de similitud;  $H$  funciona en este caso como un factor de peso. Particularizando a secuencias de ADN,  $N = 4^n$  es el número de diferentes  $n$ -*tuplas*, pues para secuencias de nucleótidos se tiene un alfabeto de tamaño 4 ( $\{A, C, G, T\}$ ).

La distancia (1.12) ha sido utilizada en contextos genéticos [22], fisiológicos [23] y lingüísticos [24].

### Métodos libres de resolución

Entre las medidas de distancia que pertenecen al segundo grupo, es decir, las que no requieren la asignación de la secuencia a vectores, vamos a destacar una medida que se basa en la complejidad Kolmogorov [25]. Kolmogorov definió la complejidad algorítmica descriptiva de un objeto como la longitud del programa más corto que describe al objeto. Formalmente, la complejidad Kolmogorov,  $C_K$ , de una cadena  $\mathcal{S}$ , respecto a una computadora universal  $U$ , está definida como:

$$C_K(\mathcal{S}) = \min_{p:U(p)=\mathcal{S}} l(p), \quad (1.13)$$

donde  $l(p)$  representa la longitud del programa  $p$  que genera la secuencia  $\mathcal{S}$ .

La medida de distancia basada en  $C_K$  fue introducida por M. Li y colaboradores [26], y tiene la siguiente representación: dadas dos secuencias  $\mathcal{S}_1$  y  $\mathcal{S}_2$

$$d_K(\mathcal{S}_1, \mathcal{S}_2) = 1 - \frac{C_K(\mathcal{S}_1) - C_K(\mathcal{S}_1|\mathcal{S}_2)}{C_K(\mathcal{S}_1\mathcal{S}_2)}, \quad (1.14)$$

donde  $C_K(\mathcal{S}_1|\mathcal{S}_2)$  es la complejidad Kolmogorov de la secuencia  $\mathcal{S}_1$  dada  $\mathcal{S}_2$ , y se define como la longitud del programa más corto que hace que una computadora estándar dé como resultado una salida  $\mathcal{S}_1$  a partir de la entrada  $\mathcal{S}_2$ .  $C_K(\mathcal{S}_1)$  está definida como  $C_K(\mathcal{S}_1|\epsilon)$  donde  $\epsilon$  es una cadena vacía. El numerador es la cantidad de información que  $\mathcal{S}_2$  tiene de  $\mathcal{S}_1$  y es igual a la información que  $\mathcal{S}_1$  tiene de  $\mathcal{S}_2$  (simetría) [25]; el denominador es la información de la cadena resultante de la concatenación de  $\mathcal{S}_1$  con  $\mathcal{S}_2$  y sirve como un factor de normalización, tal que  $d_K(\mathcal{S}_1, \mathcal{S}_2)$  toma valores entre  $[0, 1]$ .

La complejidad  $C_K$  se puede pensar como el límite inferior de todas las medidas de información y en el caso general no se puede calcular [25]. Desde un punto de vista operativo, la distancia  $d_K$  se puede aproximar estimando la complejidad de cada secuencia como la longitud de la secuencia comprimida. Para realizar esta aproximación se utilizan programas de compresión. La distancia (1.14) ha sido aplicada a estudios filogenéticos. Volveremos a este punto en el capítulo 5.

Es importante destacar que esta medida de distancia es una herramienta general, universal y no se restringe a un área en particular. Otras aplicaciones de ella se pueden encontrar, por ejemplo, en estudios de relaciones de lenguajes y análisis de cadenas de cartas [27].

### 1.3. Distancias en Física Cuántica

El desarrollo práctico de los campos de la información y de la computación cuántica requiere disponer de técnicas experimentales que permitan una manipulación fina de los estados de un sistema. La descripción formal de este manejo preciso tiene, como uno de sus ejes principales, a la noción de distinguibilidad entre estados cuánticos. Por ejemplo, una manera de medir la calidad de un protocolo de teleportación cuántica, es evaluando la diferencia entre el estado original y el teletransportado. Idéntica situación se da en muchos otros procesos originados en información cuántica: la clonación, la reconstrucción de estados cuánticos y el diseño operativo de puertas cuánticas, son algunos ejemplos claves [28].

Precisamente en este contexto, se ha reconocido que el concepto de distinguibilidad es básico para manipular información: mientras mejor podamos distinguir un estado del otro, más información tendremos sobre ellos. Es a través de la noción de distinguibilidad que se puede construir un puente entre la Teoría Cuántica y la TI [29].

El concepto de distinguibilidad entre estados se puede encontrar en otras áreas de la teoría cuántica: en el estudio de estados coherentes congelados (squeezed), en el análisis de la calidad de diferentes aproximaciones, en la detección de los cambios producidos en los estados de un sistema por una modificación ligera de su dinámica [30], etc.

Usualmente los criterios de distinguibilidad entre estados cuánticos se definen por medio de una distancia entre ellos. A continuación enumeraremos algunas de las distancias propuestas en el ámbito de la teoría cuántica.

- La norma de Hilbert-Schmidt de un operador,  $\|A\|_2$ , induce una distancia entre dos operadores densidad:

$$d_{HS}[\rho_1, \rho_2] = \sqrt{\text{Tr}[(\rho_1 - \rho_2)^2]}.$$

Esta distancia ha sido utilizada recientemente para describir la dinámica del campo en el modelo de Jaynes-Cummings [31], y para evaluar la distancia entre ciertos estados que se utilizan en óptica cuántica [32].

- Otra distancia comúnmente usada es la generada por la norma traza  $\|A\|_1$  [33]. La norma traza es la versión cuántica de la ecuación (1.7):

$$d_{Tr}[\rho_1, \rho_2] = \text{Tr}|\rho_1 - \rho_2|.$$

con  $\rho_1$  y  $\rho_2$  operadores densidad.

- El concepto de distancia estadística fue introducido por Wootters [34] en el contexto de mediciones que discriminan, de una manera óptima, los estados puros vecinos y tiene una interesante interpretación a la luz de la métrica de Fubini-Study en el espacio proyectivo complejo [35].

La distancia de Wootters fue generalizada por Braunstein y Caves para el caso de matrices densidad [36] y se puede mostrar que esta generalización es proporcional a la métrica de Bures [37]. La forma explícita de la distancia de Bures fue encontrada por Uhlmann [38] y Hübner [39].

- Otro ejemplo destacable es la contrapartida cuántica de la fidelidad (1.8),

$$F[\rho_1, \rho_2] = \text{Tr} \sqrt{(\rho_1)^{1/2} \rho_2 (\rho_1)^{1/2}}.$$

Si los operadores densidad conmutan, se reobtiene la ecuación clásica. En el caso de estados puros  $\Psi$  y  $\Phi$ ,  $F = |\langle \Psi | \Phi \rangle|^2$ , o sea la fidelidad está dada por la probabilidad de transición entre los estados [40]. También es posible interpretarla indirectamente en términos de la distancia estadística.

- Otra medida de distancia de uso generalizado en información cuántica, es la entropía relativa, la cual es una extensión cuántica de la expresión (1.5) [41].
- La distancia Monge [42] entre las correspondientes distribuciones de Husimi [43] se introdujo por la búsqueda de exponentes de Lyapunov cuánticos; esto requirió establecer una relación entre distancias en el espacio de Hilbert y en el espacio de fase clásico. La distancia Monge entre estados coherentes es igual a la distancia euclídea entre los puntos en el espacio de fase clásico.

En el capítulo 5 volveremos a analizar más detalladamente algunas de las distancias aquí mencionadas.

Hasta aquí hemos hecho una enumeración de diversas distancias utilizadas en distintos contextos de la Física Estadística y de la Mecánica Cuántica. A lo largo de los próximos capítulos nos centraremos en el estudio de una distancia entre distribuciones de probabilidad, conocida como la divergencia de Jensen-Shannon (DJS).

### Los principales aportes del presente trabajo de tesis son:

- Definición de una familia monoparamétrica de métricas, la cual emerge de una generalización de la DJS, y estudio de sus principales propiedades.
- Aplicación de esta familia de métricas al estudio estadístico de secuencias simbólicas.

- Estudio filogenético a través de la medida introducida por Li y la utilización de un método de segmentación basado en la DJS.
- Definición de una medida de complejidad para secuencias simbólicas, con aplicaciones a las secuencias de ADN.
- Estudio del problema de distinguibilidad entre estados cuánticos por medio de la DJS tanto para estados puros como para estados mixtos.

## 1.4. Mapa de esta Tesis

En el capítulo 2 se introduce la divergencia de Jensen-Shannon, se analizan posibles interpretaciones y se estudian sus propiedades como medida de distancia para el espacio  $M_+^1(N)$ .

En el capítulo 3 se propone una generalización de la divergencia de Jensen-Shannon en el marco de la Mecánica Estadística no extensiva. Se introduce de este modo una familia monoparamétrica de métricas asociada con el parámetro  $q$ .

En el capítulo 4 se estudian aplicaciones de la divergencia de Jensen-Shannon al análisis de secuencias simbólicas. En particular se estudia un método de segmentación de secuencias y se lo aplica a la segmentación de secuencias genómicas. Se propone una definición de complejidad basada en las longitudes de los segmentos resultantes de la segmentación con esta medida. Se investiga un método de segmentación de secuencias por medio de la versión de la DJS generalizada.

En el capítulo 5 se estudia el problema de la filogénesis en un grupo de mamíferos, buscando sectores en las secuencias que contienen la información relevante.

En el capítulo 6 se investigan aplicaciones de la divergencia de Jensen-Shannon a la Mecánica Cuántica. En la primera parte del capítulo se estudia su relación con varias medidas de distancia, en particular su relación con la distancia Wootters, limitando esta parte al caso de *estados puros*. Se analizan sus propiedades como una medida unificante de distinguibilidad entre estados cuánticos. En la segunda parte se define la divergencia de Jensen-Shannon para operadores densidad, extendiendo su definición para el caso de *estados mezcla*. Se realiza un análisis exhaustivo de sus propiedades, relaciones con otras distancias en el espacio de Hilbert y se la propone como una nueva medida de entrelazamiento y de fidelidad. También se la relaciona con las cotas de la información cuántica accesible.

Finalmente se presentan las conclusiones y se establecen posibles desarrollos futuros en estas líneas de trabajo.



# Capítulo 2

## La Divergencia de Jensen-Shannon

Este capítulo está dedicado a la definición y al estudio de las principales propiedades de una distancia entre distribuciones de probabilidad, conocida como la *divergencia de Jensen-Shannon*, y originada en TI. El estudio de esta cantidad constituye el eje central del presente trabajo de tesis.

### 2.1. Información y Física

Desde hace tiempo se sabe que la Física y la *información* están íntimamente relacionadas. Las ecuaciones diferenciales de la Física se pueden pensar, en muchas situaciones, como algoritmos para procesar la información contenida en las condiciones iniciales. Los datos que se obtienen de la experimentación u observación contienen la información que es la base de nuestro entendimiento de la Naturaleza [44].

Como sucede con muchos conceptos de la Física, la noción de información adquiere un significado preciso, y por lo tanto científicamente aceptable, cuando se transforma en una cantidad “medible”. Los primeros intentos para dar una medida de la información fueron realizados por H. Nyquist [45] en 1924 y luego por R. Hartley [46] en 1928. Sin embargo una resolución satisfactoria a este problema tuvo que esperar hasta los trabajos de C. Shannon de finales de la década del 40 del siglo pasado [8].

#### 2.1.1. Un enfoque probabilístico

La Teoría de la Información moderna tuvo su origen en la teoría de las comunicaciones. Wiener [47], influenciado por von Neumann introdujo la cantidad

$$\int_{-\infty}^{\infty} p(x) \log p(x) dx \quad (2.1)$$

como una medida de la información asociada con función densidad de probabilidad  $p(x)$ . Wiener advirtió la similitud de esta expresión con la *entropía* de la Mecánica Estadística.

Simultáneamente Shannon, partiendo de que el problema de enviar y recibir mensajes era básicamente estadístico, extendió las ideas de Hartley a mensajes no igualmente probables. Si los mensajes corresponden a un alfabeto  $\mathcal{A}$  compuesto por  $n$  símbolos con probabilidades  $(p_1, p_2, \dots, p_n)$ , la cantidad de información contenida en el mensaje está definida como,

$$H(\mathcal{A}) = -K \sum_{i=1}^n p_i \log p_i, \quad (2.2)$$

donde  $K$  es una constante positiva dependiente de las unidades<sup>1</sup>.

Si los símbolos son igualmente probables se reobtiene el resultado de máxima información de Hartley. Si en cambio un símbolo se trasmite con probabilidad 1,  $H(\mathcal{A}) = 0$  pues no hay información en un mensaje cuyo contenido se conoce previamente.

### 2.1.2. El enfoque físico

Más de un siglo antes de Wiener y Shannon, la búsqueda de una expresión teórica para la entropía de R. Clausius  $S = \oint dQ/dT$ , llevó a L. Boltzmann a relacionar entropía con probabilidad, sugiriendo en 1877 la conocida expresión

$$H = k_B \log W, \quad (2.3)$$

donde  $W$  es el número de estados microscópicos del sistema, compatibles con su estado termodinámico. Esta entropía provee una medida de la *falta* de información del estado microscópico del sistema.

La conexión rigurosa entre la TI y la Física, se debe a E. Jaynes [9, 10] con sus trabajos de 1957, en los que enuncia su principio de Máxima Entropía. Este autor reformula la Mecánica Estadística de acuerdo a las ideas de la TI. Jaynes reinterpreta a  $H$  dada por la ecuación (2.3) como una medida cuantitativa de la cantidad de información que se requiere para remover la incerteza en una distribución de probabilidad.

## 2.2. Derivación de la DJS

Como se mencionó en la Introducción, muchas de las medidas de distancia entre distribuciones de probabilidad provienen de la TI y sus aplicaciones son diversas [48, 49, 50, 51]. Entre las medidas propuestas que tienen su origen en la TI, una de las más usadas es la entropía relativa o divergencia de Kullback-Leibler [52].

Sea  $X$  una variable aleatoria y sean  $P^{(1)}$  y  $P^{(2)}$  dos distribuciones de probabilidad para  $X$ . La divergencia de Kullback-Leibler entre estas dos distribuciones de probabilidad

<sup>1</sup>Se se usa logaritmo en base 2,  $K = 1$ , y  $H$  se mide en bits.



está dada por:

$$K[P^{(1)}, P^{(2)}] = \sum_j p_j^{(1)} \log \frac{p_j^{(1)}}{p_j^{(2)}}, \quad (2.4)$$

donde  $p_j^{(i)}$  es la probabilidad de ocurrencia del valor  $X = x_j$  de acuerdo con la distribución de probabilidad  $P^{(i)}$ ,  $i = 1, 2$ . Obviamente  $K[P^{(1)}, P^{(2)}] = 0$  si y sólo si  $P^{(1)} = P^{(2)}$ .

La desigualdad de Jensen afirma que si  $\sum_j \alpha_j = 1$  y  $f''(x) > 0$ , entonces

$$\sum_j \alpha_j f(x_j) \geq f\left(\sum_j \alpha_j x_j\right), \quad (2.5)$$

A partir de esta desigualdad, se puede ver que  $K[P^1, P^2] > 0$  si  $P^{(1)} \neq P^{(2)}$ . Sin embargo,  $K$  no es una distancia, pues no es simétrica:

$$K[P^{(1)}, P^{(2)}] \neq K[P^{(2)}, P^{(1)}].$$

Tampoco satisface la desigualdad triangular.

La divergencia  $K[P^{(1)}, P^{(2)}]$  no es acotada, y sólo está definida si las distribuciones de probabilidad involucradas satisfacen la condición de continuidad absoluta [53], esto es, se debe cumplir que  $p_j^{(1)} = 0$  siempre que  $p_j^{(2)} = 0$ . Este requerimiento es fuertemente restrictivo, como se verá en los próximos capítulos.

Para superar las falencias de esta medida, C. Rao [54] y J. Lin [55] introdujeron una medida de divergencia alternativa, conocida como la divergencia de Jensen-Shannon.

A continuación, y siguiendo a Lin, justificaremos su definición. Sea

$$I[P^{(1)}, P^{(2)}] = \sum_j p_j^{(1)} \log \frac{p_j^{(1)}}{\frac{1}{2}p_j^{(1)} + \frac{1}{2}p_j^{(2)}}, \quad (2.6)$$

que en términos de la divergencia  $K$  se puede escribir en la forma:

$$I[P^{(1)}, P^{(2)}] = K\left[P^{(1)}, \frac{1}{2}P^{(1)} + \frac{1}{2}P^{(2)}\right] \quad (2.7)$$

Es claro que para (2.6) se relaja la condición de continuidad absoluta, ya que si el numerador del cociente en el logaritmo no se anula, el denominador tampoco. A su vez, la simetrización de  $I$ , conduce a la cantidad:

$$JS[P^{(1)}, P^{(2)}] = \frac{1}{2}\{I[P^{(1)}, P^{(2)}] + I[P^{(2)}, P^{(1)}]\}. \quad (2.8)$$

Esta es la definición de la divergencia de Jensen-Shannon (DJS). Esta medida de divergencia se puede expresar en términos de la entropía de Shannon, ec. (2.2), de la forma:

$$JS[P^{(1)}, P^{(2)}] = H\left[\frac{P^{(1)} + P^{(2)}}{2}\right] - \frac{1}{2}H[P^{(1)}] - \frac{1}{2}H[P^{(2)}]. \quad (2.9)$$

$JS$  es definida positiva (lo cual puede verificarse por la convexidad de la entropía de Shannon y por la desigualdad de Jensen), simétrica (por construcción) y nula si y sólo si las distribuciones de probabilidad son iguales.

Un aspecto que hace a esta cantidad sumamente interesante es el hecho de que es posible asignar a las distribuciones de probabilidad  $P^{(1)}$  y  $P^{(2)}$  distintos pesos. En efecto, sean  $\pi^{(1)}$  y  $\pi^{(2)}$  dos números reales no negativos (pesos), tales que  $\pi^{(1)} + \pi^{(2)} = 1$ . Entonces, una generalización natural de la cantidad (2.9) es:

$$\begin{aligned} JS^{(\pi^{(1)}, \pi^{(2)})}[P^{(1)}, P^{(2)}] &= H[\pi^{(1)}P^{(1)} + \pi^{(2)}P^{(2)}] \\ &- \pi^{(1)}H[P^{(1)}] - \pi^{(2)}H[P^{(2)}]. \end{aligned} \quad (2.10)$$

Seguiremos llamando a esta cantidad la divergencia de Jensen-Shannon (DJS) entre  $P^{(1)}$  y  $P^{(2)}$ . Si  $\pi^{(1)} \neq \pi^{(2)}$ ,  $JS^{(\pi^{(1)}, \pi^{(2)})}[P^{(1)}, P^{(2)}]$  deja de ser simétrica, pero las otras propiedades siguen siendo válidas. La extensión recién sugerida puede llevarse más allá, y tomar a  $JS$  como una medida de “distancia” entre más de dos distribuciones de probabilidad: Sean  $P^{(1)}, P^{(2)}, \dots, P^{(m)}$ ,  $m$  distribuciones de probabilidad, con pesos  $\pi^{(1)}, \pi^{(2)}, \dots, \pi^{(m)}$  ( $\pi^{(i)} > 0$ ;  $\sum_{i=1}^m \pi^{(i)} = 1$ ). Entonces la DJS entre las distribuciones  $P^{(i)}$  se define como:

$$JS^{(\pi^{(1)}, \dots, \pi^{(m)})}[P^{(1)}, P^{(2)}, \dots, P^{(m)}] = H \left[ \sum_{i=1}^m \pi^{(i)} P^{(i)} \right] - \sum_{i=1}^m \pi^{(i)} H[P^{(i)}]. \quad (2.11)$$

Otra propiedad notable de la DJS es su carácter métrico: para el caso de dos distribuciones de probabilidad y de pesos iguales,  $\pi^{(1)} = \pi^{(2)} = \frac{1}{2}$ , se ha demostrado que la raíz cuadrada  $\sqrt{JS^{(\frac{1}{2}, \frac{1}{2})}}$  verifica la desigualdad triangular [56, 57]. Volveremos a este punto en el próximo capítulo.

## 2.3. Interpretaciones de la DJS

La DJS tiene interesantes interpretaciones en el marco de la Mecánica Estadística y en TI clásica y cuántica. Pospondremos esta última para el capítulo 6 y nos limitaremos ahora al marco clásico [58].

### 2.3.1. La DJS como entropía intensiva de la mezcla

Consideremos  $m$  recipientes que contienen una mezcla de  $k$  gases ideales. Sea  $F^{(j)} \equiv (f_1^{(j)}, f_2^{(j)}, \dots, f_k^{(j)})$  el vector de las fracciones molares de los  $k$  gases en el  $j$ -ésimo recipiente, para  $j = 1, 2, \dots, m$ , y sea  $n^j$  el número total de moléculas en el recipiente  $j$ . De la segunda ley de la termodinámica se sabe que la suma de las entropías de Boltzmann de los  $m$  recipientes separados es menor o igual que la entropía del gas contenido en el recipiente

de la mezcla. Fácilmente se puede ver que la diferencia de las entropías antes y después de mezclar los gases, es decir la entropía de mezcla  $H_{mezcla}$ , es igual a

$$H_{mezcla} = k_B \ln 2 \left( NH[F] - \sum_{j=1}^m n^j H[f^j] \right), \quad (2.12)$$

donde  $N \equiv \sum_{j=1}^m n^j$  es el número total de partículas de gas ideal en los  $m$  recipientes, y  $F \equiv \sum_{j=1}^m \frac{n^j}{N} f^j$ . Entonces,

$$H_{mezcla} = Nk_B \ln 2 JS^{(n^1/N, n^2/N, \dots, n^m/N)}, \quad (2.13)$$

Así  $JS$  se puede pensar como la entropía intensiva de la mezcla medida en unidades de  $k_B \ln 2$ .

### 2.3.2. La DJS como información mutua

La DJS ha sido utilizada en diversos contextos como una herramienta útil para el estudio estadístico de secuencias simbólicas. Nosotros haremos un estudio detallado de estas aplicaciones en el próximo capítulo. Sin embargo, en esta sección repasaremos una interesante interpretación de la DJS en el marco de la Teoría de la Información aplicada al estudio de secuencias simbólicas.

Sea  $\mathcal{S}$  una secuencia con  $L$  símbolos tomados de un alfabeto  $\mathcal{A} = \{a_1, \dots, a_k\}$ , y supongamos que  $p_i$  es la probabilidad de encontrar el símbolo  $a_i$ ,  $i = 1, 2, \dots, k$ , en una posición arbitraria pero fija en la secuencia  $\mathcal{S}$ . Supongamos que la secuencia  $\mathcal{S}$  se divide en  $m$  subsecuencias  $\mathcal{S}^{(1)}, \mathcal{S}^{(2)}, \dots, \mathcal{S}^{(m)}$  de longitudes  $n^1, n^2, \dots, n^m$ , respectivamente. Denotemos por  $p_i^{(j)}$  la probabilidad de encontrar el símbolo  $a_i$  en una posición arbitraria pero fija en la subsecuencia  $\mathcal{S}^{(j)}$ , para  $i = 1, 2, \dots, k$  y  $j = 1, 2, \dots, m$ .

Introducimos ahora dos variables aleatorias:  $a \in \mathcal{A}$  y  $s \in \{\mathcal{S}^{(1)}, \mathcal{S}^{(2)}, \dots, \mathcal{S}^{(m)}\}$ . La variable aleatoria  $a$  puede tomar los valores  $a_1, a_2, \dots, a_k$  con probabilidades  $p_1, p_2, \dots, p_k$  mientras que la variable aleatoria  $s$  puede tomar alguno de los valores  $\mathcal{S}^{(1)}, \mathcal{S}^{(2)}, \dots, \mathcal{S}^{(m)}$  con probabilidades  $\pi^1 = n^1/L, \pi^2 = n^2/L, \dots, \pi^m = n^m/L$ .

Supongamos que se toma un símbolo  $a$  de la secuencia  $\mathcal{S}$ , sin saber a qué subsecuencia pertenece y pensemos que nuestra tarea es “adivinar” a cuál subsecuencia corresponde ese símbolo. La TI provee una respuesta a la pregunta de cuánta información se puede obtener de la identidad del símbolo  $a$  a partir de la identificación de la subsecuencia  $s$  de la cual fue tomado. Esa cantidad es la información mutua,  $IM$ , de  $a$  sobre  $s$  [52]:

$$IM \equiv \sum_{i=1}^k \sum_{j=1}^m p_{ij} \log \frac{p_{ij}}{\pi^j p_i}, \quad (2.14)$$

donde  $p_{ij}$  es la probabilidad conjunta de  $a$  y  $s$ , i.e. la probabilidad de encontrar el símbolo  $a_i$  en una posición arbitraria pero fija de la secuencia  $\mathcal{S}^{(j)}$ . Entonces, teniendo en cuenta que  $p_i^{(j)}$  denota la probabilidad condicional de encontrar el símbolo  $a_i$  en una posición

arbitraria pero fija en una dada secuencia (fija)  $\mathcal{S}^{(j)}$ ,  $p_{ij} = \pi^j p_i^{(j)}$ , la ecuación (2.14) se puede reescribir como:

$$IM \equiv \sum_{i=1}^k \sum_{j=1}^m \pi^j p_i^{(j)} \log \frac{p_i^{(j)}}{p_i}. \quad (2.15)$$

o equivalentemente:

$$IM = \sum_{j=1}^m \pi^j \sum_{i=1}^k p_i^{(j)} \log p_i^{(j)} - \sum_{i=1}^k \left( \sum_{j=1}^m \pi^j p_i^{(j)} \right) \log p_i. \quad (2.16)$$

Como  $p_i = \sum_{j=1}^m \pi^j p_i^{(j)}$  define la probabilidad de encontrar el símbolo  $a_i$  en la secuencia, se tiene

$$IM = JS^{(\pi^1, \dots, \pi^m)}[P^{(1)}, P^{(2)}, \dots, P^{(m)}] \quad (2.17)$$

con  $P^{(j)} = \{p_i^{(j)}\}$ . Para el caso en que  $P^{(1)} = P^{(2)} = \dots = P^{(m)}$ , es claro que conocer la identidad del símbolo  $a$  no dice nada sobre la identidad de la subsecuencia  $s$  a la cual  $a$  pertenece. De la misma forma, en este caso, conocer la subsecuencia  $s$  de la cual se tomó  $a$  no dice nada respecto de la identidad de  $a$ . Entonces es intuitivamente aceptable que la información mutua de  $a$  sobre  $s$  (la información mutua de  $s$  sobre  $a$ ) es igual a cero, y obviamente en este caso la divergencia  $JS$  es igual a cero.

# Capítulo 3

## La $q$ -Divergencia de Jensen-Shannon

*Há de se distinguir claramente una proposta de generalizar a forma funcional da entropia BG, de um empreendimento consideravelmente maior, qual seja uma proposta de generalizar a própria mecânica estadística de BG, e de sua conexão com a termodinâmica.*<sup>1</sup>

Este capítulo está dedicado al estudio de una generalización de la divergencia de Jensen-Shannon en el marco de la mecánica estadística no-extensiva de Tsallis [59]. Presentamos aquí las principales propiedades de esta generalización, entre las que destaca su carácter métrico. Dado que la estadística de Tsallis tiene un parámetro libre, usualmente denotado por  $q$  y actualmente referenciado como parámetro de deformación, la generalización propuesta conduce a una familia monoparamétrica de métricas. Investigaremos esta medida en el contexto del análisis estadístico de secuencias simbólicas. Antes de ir a la definición de la divergencia generalizada, haremos un breve repaso de la estadística no-extensiva de Tsallis.

### 3.1. Mecánica estadística no-extensiva

El punto de partida de la estadística no-extensiva de Tsallis (de ahora en más referenciada como qME) es la  $q$ -entropía, definida a través de la expresión

$$H_q[P] = -k_B \frac{1 - \sum_{j=1}^n p_j^q}{1 - q}; \quad \left( \sum_{j=1}^n p_j = 1 \right) \quad (3.1)$$

donde  $q$  es un número real, y como antes,  $p_j$  representa la probabilidad del estado  $j$ -ésimo ( $j = 1 \dots n$ ). Para  $q = 1$  se reobtiene la expresión de Boltzmann-Gibbs (BG) para la entropía (1.3).

A lo largo de las últimas dos décadas esta generalización de la Mecánica Estadística ha sido aplicada a una gran cantidad de problemas de interés en Física, Biología, Astronomía,

---

<sup>1</sup>C. Tsallis. Algumas reflexões sobre a natureza das teorias físicas em geral e da mecânica estatística em particular, 2003

Economía y otras áreas del conocimiento. Entre ellos se destacan los estudios sobre el plegamiento de proteínas [60], los rayos cósmicos [61], el fenómeno de turbulencia [62, 63, 64], las finanzas [65], la aniquilación electrón-positrón [66], los movimientos de las células de la *Hydra* [67], la epilepsia [68], astrofísica [69], las estructuras de las aglomeraciones urbanas [70], las características de la web [71], etc. Un listado bibliográfico exhaustivo y actualizado sobre la qME puede encontrarse en [72].

Desde el punto de vista de los fundamentos, podemos decir que la Mecánica Estadística de Boltzmann-Gibbs (MEBG) luce como una buena aproximación para estudiar sistemas que involucran correlaciones de corto alcance, sistemas markovianos y en general, para sistemas que involucran caos (fuerte) en su dinámica microscópica, i.e. exponentes de Lyapunov positivos y ocupación ergódica del espacio de las fases. Sin embargo, esta descripción estándar falla para sistemas cuya dinámica microscópica es más compleja, i.e. sistemas que exhiben caos débil con un espectro de Lyapunov que se anula. Para estos sistemas es esperable que la ocupación del espacio de las fases tenga una estructura menos trivial con una geometría (multi)fractal o jerárquica. En este sentido la qME es una propuesta de generalización de, y no una alternativa a, la MEBG, y se presenta para extender el dominio de aplicabilidad de la Mecánica Estadística a sistemas con estas características [73, 74].

Dentro del formalismo matemático de la qME, la *función q-exponencial*

$$e_q(x) \equiv [1 + (1 - q)x]^{1/(1-q)} \quad (q, x \in R) \quad (3.2)$$

juega un rol fundamental. Esta función tiene las siguientes propiedades:

- (i) En el límite  $q \rightarrow 1$  se reobtiene la función exponencial usual:  $e_1(x) = e^x (\forall x)$
- (ii) para  $q > 1$ ,  $e_q(x)$  decae en forma de ley de potencia cuando  $x \rightarrow -\infty$  y diverge en  $x = 1/(q - 1)$
- (iii) para  $q < 1$ , y para valores de  $x$  menores que  $-1/(1 - q)$ , la exponencial  $e_q(x)$  resulta ser idénticamente igual a cero.

La función inversa de la q-exponencial es el *q-logaritmo*, el cual está definido por:

$$\log_q(x) \equiv \frac{x^{1-q} - 1}{1 - q} \quad (q \in R). \quad (3.3)$$

Nuevamente en el caso  $q = 1$  se tiene  $\log_1(x) = \ln x (\forall x)$ .

En términos del  $\log_q(x)$ , la entropía  $H_q$  puede expresarse en la forma:

$$H_q[P] = -k_B \sum_{j=1}^n p_j^q \log_q(p_j) \quad (3.4)$$

El carácter no-extensivo de la qME emerge de la siguiente propiedad de la entropía  $H_q$ :

$$H_q[A + B] = H_q[A] + H_q[B] + (1 - q)H_q[A].H_q[B], \quad (3.5)$$

donde  $A$  y  $B$  representan a dos sistemas físicos independientes, en el sentido de que la distribución de probabilidad del sistema compuesto  $A + B$  se puede factorizar como el producto de las distribuciones de probabilidad de  $A$  y de  $B$ :

$$p_{ij}^{A+B} = p_i^A p_j^B \quad \forall (i, j).$$

Uno de los resultados más destacados de la qME es la deducción de las leyes de distribución de Levy por medio de la aplicación de un principio variacional a la entropía  $H_q$  [75]. Esta vinculación permite establecer una conexión entre el índice  $q$  y los exponentes de las distribuciones de Levy. Es importante, en este contexto, tener presente la relación existente entre las distribuciones de Levy y el fenómeno de difusión anómala, el cual se ha observado en una gran variedad de sistemas físicos y biológicos. Un ejemplo particularmente significativo, lo proveen las secuencias de ADN, pues es conocido que ciertos sectores de estas secuencias exhiben correlaciones de largo alcance con propiedades estadísticas análogas a las de la difusión anómala [76]. Motivados en estos hechos, estudiaremos la generalización de la DJS propuesta, al estudio de secuencias simbólicas.

## 3.2. Definición y propiedades

La extensión natural de la divergencia de Kullbak-Leibler entre las distribuciones de probabilidad  $P^{(1)}$  y  $P^{(2)}$ , en el marco de la qME es [77]:

$$K_q[P^{(1)}, P^{(2)}] = - \sum_{j=1}^n p_j^{(1)} \log_q \left( \frac{p_j^{(2)}}{p_j^{(1)}} \right) \quad (3.6)$$

Se puede ver fácilmente que  $K_q \geq 0$  para  $q > 0$  y  $K_q = 0$  para  $q = 0$  [78]. En efecto, la siguiente desigualdad es válida para  $x > 0$  y  $q > 0$ :

$$\frac{x^{q-1} - 1}{q - 1} \geq 1 - \frac{1}{x} \quad (3.7)$$

Entonces, a partir de la definición (3.6), resulta:

$$\begin{aligned} K_q[P^{(1)}, P^{(2)}] &= - \sum_{j=1}^n p_j^{(1)} \frac{\left( \frac{p_j^{(2)}}{p_j^{(1)}} \right)^{1-q} - 1}{1 - q} \\ &= \sum_{j=1}^n p_j^{(1)} \frac{\left( \frac{p_j^{(1)}}{p_j^{(2)}} \right)^{q-1} - 1}{q - 1} \\ &\geq \sum_{j=1}^n p_j^{(1)} \left[ 1 - \frac{p_j^{(2)}}{p_j^{(1)}} \right] = 0 \end{aligned} \quad (3.8)$$

pues ambas distribuciones  $P^{(1)}$  y  $P^{(2)}$  están normalizadas.

Ya munidos de la q-divergencia de Kullback-Leibler, podemos seguir los pasos que nos condujeron, en el capítulo anterior, a la definición de la divergencia de Jensen-Shannon. En efecto, la cantidad (2.8) se puede generalizar, para  $q \neq 1$ , en la forma:

$$L_q[P^{(1)}, P^{(2)}] = K_q \left[ P^{(1)}, \frac{P^{(1)} + P^{(2)}}{2} \right] + K_q \left[ P^{(2)}, \frac{P^{(1)} + P^{(2)}}{2} \right] \quad (3.9)$$

lo cual, en términos del  $\log_q$ , se puede reescribir como:

$$\begin{aligned} L_q[P^{(1)}, P^{(2)}] = & - \sum_{j=1}^n \left[ \left( p_j^{(1)} \right)^q + \left( p_j^{(2)} \right)^q \right] \log_q \left( \frac{p_j^{(1)} + p_j^{(2)}}{2} \right) \\ & + \sum_{j=1}^n \left( p_j^{(1)} \right)^q \log_q p_j^{(1)} + \sum_{j=1}^n \left( p_j^{(2)} \right)^q \log_q p_j^{(2)} \end{aligned} \quad (3.10)$$

Tomando la suma simetrizada de esta cantidad, llegamos a la generalización buscada para la DJS en el marco de la qME:

$$\begin{aligned} JS_q[P^{(1)}, P^{(2)}] = & - \sum_{j=1}^n \frac{1}{2} \left[ \left( p_j^{(1)} \right)^q + \left( p_j^{(2)} \right)^q \right] \log_q \left( \frac{p_j^{(1)} + p_j^{(2)}}{2} \right) \\ & + \frac{1}{2} \sum_{j=1}^n \left( p_j^{(1)} \right)^q \log_q p_j^{(1)} + \frac{1}{2} \sum_{j=1}^n \left( p_j^{(2)} \right)^q \log_q p_j^{(2)} \end{aligned} \quad (3.11)$$

o, en general, si se asignan pesos  $\pi^{(1)}$  y  $\pi^{(2)}$  a las distribuciones  $P^{(1)}$  y  $P^{(2)}$ , respectivamente, con  $\pi^{(1)} + \pi^{(2)} = 1$ , se llega a:

$$\begin{aligned} JS_q^{(\pi^{(1)}, \pi^{(2)})}[P^{(1)}, P^{(2)}] = & - \sum_{j=1}^n \left[ \pi^{(1)} \left( p_j^{(1)} \right)^q + \pi^{(2)} \left( p_j^{(2)} \right)^q \right] \log_q \left( \pi^{(1)} p_j^{(1)} + \pi^{(2)} p_j^{(2)} \right) \\ & + \pi^{(1)} \sum_{j=1}^n \left( p_j^{(1)} \right)^q \log_q p_j^{(1)} + \pi^{(2)} \sum_{j=1}^n \left( p_j^{(2)} \right)^q \log_q p_j^{(2)} \end{aligned} \quad (3.12)$$

Esta última expresión es nuestra definición de la q-DJS. En el límite  $q \rightarrow 1$  se recupera la DJS definida en el capítulo anterior.

La extensión natural de la definición (3.12) a un número arbitrario  $m$  de distribuciones de probabilidad está dada por



$$\begin{aligned}
JS_q^{(\pi^{(1)}, \dots, \pi^{(m)})}[P^{(1)}, \dots, P^{(m)}] &= - \sum_{j=1}^n \left[ \pi^{(1)} \left( p_j^{(1)} \right)^q + \dots + \pi^{(m)} \left( p_j^{(m)} \right)^q \right] \\
&\quad \times \log_q \left[ \pi^{(1)} p_j^{(1)} + \dots + \pi^{(m)} p_j^{(m)} \right] \\
&\quad + \pi^{(1)} \sum_{j=1}^n \left( p_j^{(1)} \right)^q \log_q p_j^{(1)} + \dots + \pi^{(m)} \sum_{j=1}^n \left( p_j^{(m)} \right)^q \log_q p_j^{(m)}.
\end{aligned} \tag{3.13}$$

Es importante notar que nuestra generalización no consiste en el reemplazo directo de  $H$  por  $H_q$ , en la expresión (2.11)<sup>2</sup>. Es más, fácilmente se puede verificar que, para  $q \neq 1$ ,

$$JS_q^{(\pi^{(1)}, \dots, \pi^{(m)})}(P^{(1)}, \dots, P^{(m)}) \neq \Delta H_q \equiv H_q(\pi^{(1)} P^{(1)} + \dots + \pi^{(m)} P^{(m)}) - \sum_{j=1}^m \pi^{(j)} H_q(P^{(j)});$$

en particular se cumple la desigualdad

$$\begin{aligned}
JS_q - \Delta H_q &= \sum_{i=1}^k \sum_{j=1}^m \pi^{(j)} \left( p_i^{(j)} \right)^q \log_q(p_i) - \sum_{i=1}^k (p_i)^q \log_q(p_i) \\
&= \sum_{i=1}^k \left[ \sum_{j=1}^m \pi^{(j)} \left( p_i^{(j)} \right)^q - \left( \sum_{j=1}^m \pi^{(j)} p_i^{(j)} \right)^q \right] \log_q(p_i) \\
&\geq 0
\end{aligned} \tag{3.14}$$

para  $q > 1$ .

$JS_q$  satisface las siguientes propiedades:

- (i)  $JS_q[P^{(1)}, P^{(2)}] \geq 0$  para  $q > 0$ ,  $JS_q[P^{(1)}, P^{(2)}] = 0$  para  $q = 0$  y  $JS_q[P^{(1)}, P^{(2)}] \leq 0$  para  $q < 0$ . En adelante trabajaremos con valores de  $q$  mayores que cero.
- (ii)  $JS_q[P^{(1)}, P^{(2)}] = 0$  si y sólo si  $P^{(1)} = P^{(2)}$ .
- (iii)  $JS_q$  es simétrica si los pesos  $\pi^{(1)} = \pi^{(2)} = 1/2$ .
- (iv) La q-DJS es una función que crece monótonamente con  $q$ .

La divergencia  $JS_q$  puede interpretarse como la información mutua no extensiva.

En la figura 3.1 se muestra el crecimiento monótono de  $JS_q$  como función del parámetro  $q$ . En este caso se eligió un alfabeto binario y distribuciones de probabilidad  $P^{(1)} = (t, 1-t)$  y  $P^{(2)} = (1-t, t)$ ,  $t \in [0, 1]$ .

<sup>2</sup>La definición propuesta para la qDJS es la única compatible con la q divergencia de Kullback-Leibler. La sustitución de  $H$  por  $H_q$ , aún cuando conduce a una divergencia que satisface algunas de las propiedades deseables para una distancia entre distribuciones de probabilidad, modifica su significado en el contexto de una teoría de la información compatible con los postulados de la qME

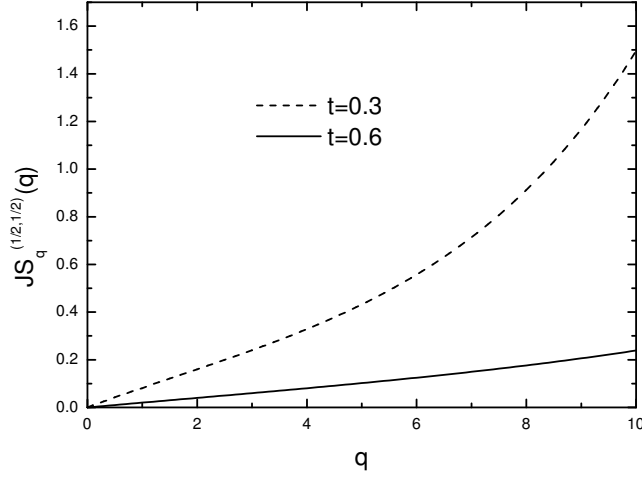


Figura 3.1:  $JS_q$  en función de  $q$ , para distribuciones de probabilidad correspondientes a un alfabeto binario  $P^{(1)} = (t, 1 - t)$  y  $P^{(2)} = (1 - t, t)$ , para  $t = 0.3, 0.6$ . Los pesos se tomaron  $\pi^{(1)} = \pi^{(2)} = 1/2$ .

### 3.3. La q-DJS aplicada al análisis de secuencias - Segmentación

Consideremos un alfabeto de  $k$  letras,  $\mathcal{A} = \{a_1, \dots, a_k\}$  y sea  $\mathcal{S}$  una secuencia de  $L$  símbolos tomados del alfabeto  $\mathcal{A}$ ,

$$\mathcal{S} = a_{n1}a_{n2}\dots a_{nL}.$$

Dada la secuencia  $\mathcal{S}$ , es posible evaluar las frecuencias de ocurrencia de los distintos símbolos:

$$f_j = \frac{N_j}{L}, \quad (3.15)$$

donde  $N_j$  representa el número de símbolos  $a_j$  en la secuencia  $\mathcal{S}$ . En las situaciones de interés, las secuencias simbólicas no tienen una composición homogénea en toda su extensión (son no estacionarias), sino que esa homogeneidad composicional se encuentra en parches o dominios. Así, resulta de gran relevancia práctica y formal, el desarrollo de métodos que permitan determinar esos distintos dominios de homogeneidad composicional. En lo que sigue de esta sección, usaremos a la divergencia q-JSD como una herramienta de detección de esos dominios de homogeneidad.

Supongamos que la secuencia  $\mathcal{S}$  está formada por dos o más subsecuencias  $\mathcal{S}^{(i)}$ , cada una de ellas construidas a partir de diferentes distribuciones de probabilidad de ocurrencia de cada uno de los símbolos. El problema de detección de los puntos de fusión de estas

subsecuencias se conoce como el problema de la segmentación de la secuencia  $\mathcal{S}$ , y resulta ser de gran interés práctico. Más adelante consideraremos contextos en donde este problema se presenta. Por ahora, sólo haremos un tratamiento formal del mismo.

La utilización de la divergencia  $JS_q$  en el tratamiento del problema de segmentación está motivada en las propiedades de la divergencia  $JS_q$  que a continuación exponemos:

### 3.3.1. Propiedades Estadísticas de la q-DJS:

Formalmente, la divergencia  $JS_q$  es una función de las distribuciones de probabilidad. Cuando se estudian secuencias simbólicas las distribuciones de probabilidad se pueden estimar a partir de las distribuciones de frecuencia de ocurrencia de los símbolos. Las fluctuaciones estadísticas asociadas con esta estimación repercutirá en la distribución de los valores de  $JS_q$ .

Otro aspecto importante de la divergencia  $JS_q$  es la libertad que permite en la elección de los pesos  $\pi^{(j)}$ . Para poner de manifiesto el comportamiento de la  $JS_q$  ante cambios en la elección de los pesos, y dentro del análisis de secuencias, procederemos a la realización de algunas simulaciones.

Consideremos un conjunto de 2000 secuencias binarias, ( $k = 2$ ) de  $L = 1500$  símbolos cada una. Cada una de estas secuencias se obtiene de concatenar dos subsecuencias ( $m=2$ ) de la siguiente forma: una de las subsecuencias, de tamaño  $L_1 = 500$  símbolos se construye tomando símbolos de manera aleatoria, con una probabilidad  $P^{(1)} = (t, 1 - t)$ , donde  $t$  es algún número arbitrario entre cero y uno, mientras que la otra subsecuencia de longitud  $L - L_1 = 1000$  símbolos con distribución de probabilidad  $P^{(2)} = (s, 1 - s)$ , con  $s$  otro número arbitrario entre cero y uno. Definimos sobre la secuencia completa un cursor móvil, el cual separa a esta en dos subsecuencias, una a la izquierda y otra a la derecha del cursor. Para cada una de estas subsecuencias calculamos las frecuencias de ocurrencia de los símbolos y evaluamos la divergencia  $JS_q$  entre ambas distribuciones de frecuencias. De este modo, podemos calcular a  $JS_q$  como una función de la posición del cursor. Más concretamente, si llamamos  $n^1$  a la posición del cursor, quedarán determinadas dos subsecuencias, una de longitud  $n^1$  y otra de longitud  $n^2 = L - n^1$ . Para ilustrar los efectos de las posibles elecciones de los pesos, calculamos la q-DJS de dos formas diferentes:

- i)* pesos iguales ( $1/2$ ) para ambas subsecuencias, tanto para la que está a la derecha del cursor como la que está a la izquierda; es decir calculamos  $JS_q^{(\frac{1}{2}, \frac{1}{2})}$
- ii)* pesos  $\pi^{(1)} = \frac{n^1}{L}$  y  $\pi^{(2)} = \frac{n^2}{L}$ ; o sea evaluamos  $JS_q^{(n^1/L, n^2/L)}$

La Figura 3.2, muestra los valores obtenidos para la divergencia  $JS_q$ , promediada sobre las 2000 secuencias para ambas elecciones de los pesos como una función de la posición del cursor:  $\langle JS_q^{(\pi^{(1)}, \pi^{(2)})}(n^1) \rangle$ . Vemos un comportamiento marcadamente diferente para las dos elecciones. En el segundo caso, la divergencia alcanza un máximo global en el punto de fusión entre las dos subsecuencias ( $L_1$ ) mientras que para la primera elección, si bien  $JS_q$  alcanza un máximo en el punto de fusión, también logra máximos en los bordes de la

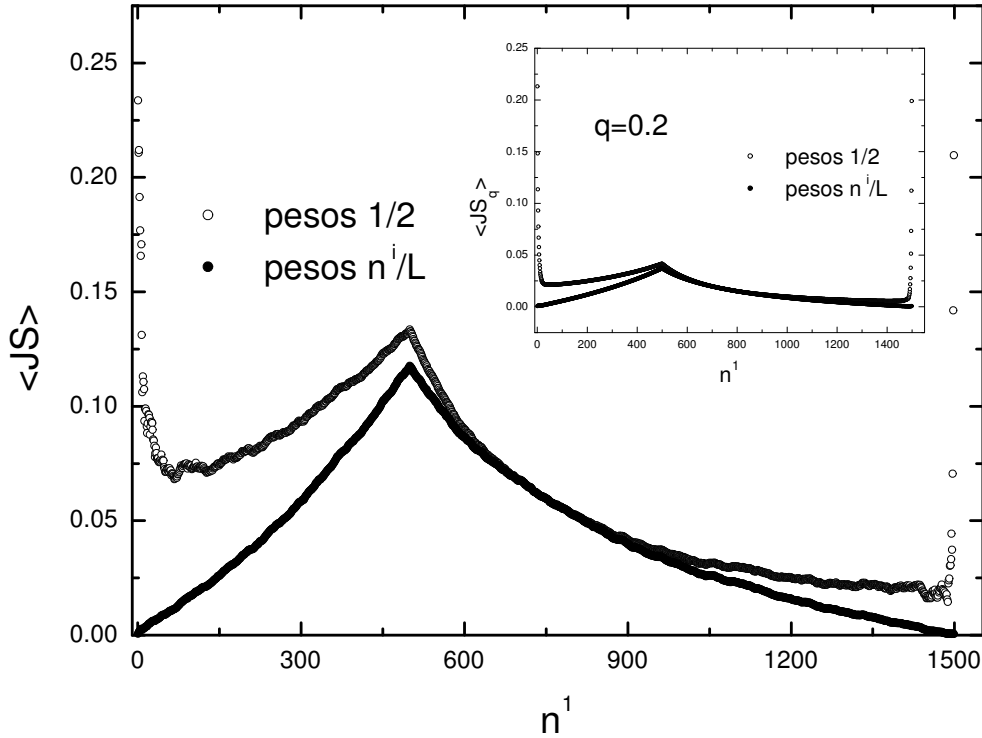


Figura 3.2:  $\langle JS_q \rangle$  en función de la posición del curso móvil para ambas elecciones de los pesos. En este caso se concatenaron 2 secuencias generadas con probabilidades,  $P^{(1)}=(0.2, 0.8)$  y  $P^{(2)}=(0.7, 0.3)$ , y con longitudes  $L_1 = 500$  y  $L_2 = 1000$  respectivamente. El proceso de segmentación se realizó utilizando el parámetro  $q = 1$ . En ambos casos se detecta el punto de fusión de las secuencias. Las curvas corresponden a un promedio sobre 2000 secuencias. En el inserto se muestra el mismo comportamiento para  $q=0.2$

secuencia, esto se debe a efectos de tamaño finito. Claramente este hecho inclina la balanza a favor de la segunda opción, si se trata de disponer de una herramienta que permita la detección de los puntos en donde el contenido composicional de una secuencia cambia.

En la Figura 3.3 se muestra el comportamiento de  $JS_q$  para diferentes valores de  $q$ . En esta figura se pone de manifiesto el crecimiento de la magnitud del máximo de  $JS_q$  en función del parámetro  $q$ . Esto resulta de importancia, pues podrían haber casos en donde la detección de las diferencias composicionales sea dificultosa, y entonces una elección adecuada de  $q$  tenga un efecto amplificador de la magnitud del máximo en la  $JS_q$ . Más allá de esto, podemos pensar que habrá un valor adecuado de  $q$  para cada problema particular

que nos interese estudiar.

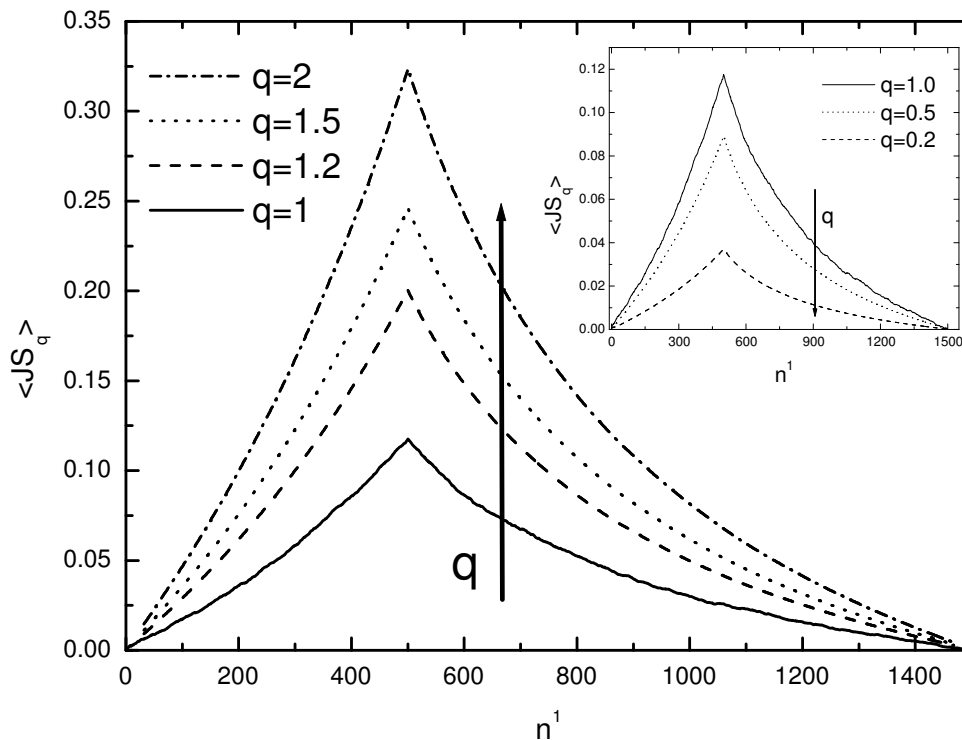


Figura 3.3: La  $\langle JS_q \rangle$  para diferentes valores de  $q$  mayores que 1. Nuevamente se concatenaron dos secuencias con probabilidades  $P^{(1)}=(0.2, 0.8)$  y  $P^{(2)}=(0.7, 0.3)$ , y longitudes  $L_1 = 500$  y  $L_2 = 1000$ . Las curvas corresponden al promedio sobre 2000 secuencias. En el inserto se muestra la  $\langle JS_q \rangle$  para valores de  $q < 1$

El comportamiento arriba indicado, que muestra la potencialidad de la divergencia  $JS_q$  en el proceso de detección de cambios en la distribución de símbolos a lo largo de una secuencia, ha sido utilizado en el desarrollo de un método de segmentación de secuencias simbólicas. El mismo será descrito en detalle en el próximo capítulo. Por ahora sólo mencionaremos que se trata de un esquema iterativo. Como tal requiere de un criterio de detención. Sin este criterio la segmentación terminaría recién con la secuencia dividida en los símbolos individuales. Como se verá, en el algoritmo de segmentación desarrollado en base a la divergencia  $JS_q$ , el proceso se detiene cuando ya no se encuentran diferencias significativas entre las potenciales subsecuencias. Para precisar lo que entendemos por “diferencia significativa”, introduciremos una cantidad llamada *nivel de significancia* [58].

### 3.3.2. Nivel de significancia

Sea nuevamente una secuencia  $\mathcal{S}$  la cual, en principio se puede separar en  $m$  subsecuencias de contenido homogéneo. Un caso límite es el que corresponde a que estas  $m$  subsecuencias hayan sido generadas a partir de la misma distribución de probabilidad. Dado un valor observado de  $JS_q = x$ , queremos evaluar la probabilidad de obtener un valor igual o menor que  $x$ , bajo la hipótesis anterior de que todas las subsecuencias han sido generadas con idéntica distribución de probabilidad. A aquella probabilidad se la llama el *nivel de significancia* del dado valor  $x$ :

$$s_q(x) \equiv \text{Prob}\{JS_q \leq x\}. \quad (3.16)$$

$s(x)$  no tiene una expresión analítica simple; sin embargo podemos conseguir una aproximación usando la expansión de Taylor de  $JS_q$  hasta términos cuadráticos [58].

#### Cálculo aproximado de $s_q(x)$

Consideremos una secuencia simbólica  $\mathcal{S}$  de longitud  $L$  con símbolos tomados del alfabeto  $\mathcal{A}$ . Supongamos que después de un proceso de segmentación, la secuencia  $\mathcal{S}$  se puede separar en  $m$  subsecuencias  $\mathcal{S}^{(j)}$  de longitudes  $L_j$  respectivamente. Tomaremos los pesos  $\pi^{(j)} = L_j/L$ . Vamos a denotar con  $f_i^{(j)}$  la frecuencia de ocurrencia del símbolo  $i$  –ésimo en la subsecuencia  $j$  –ésima, y con  $f_i$  a la frecuencia del símbolo  $i$  –ésimo a lo largo de toda la secuencia  $\mathcal{S}$ . Obviamente,

$$f_i = \sum_{j=1}^m \pi^{(j)} f_i^{(j)}.$$

Podemos escribir la ecuación (3.13) de forma compacta de la siguiente manera,

$$\begin{aligned} JS_q^{(\pi^{(1)}, \dots, \pi^{(m)})}(F_1, \dots, F_m) &= \sum_{j=1}^m \sum_{i=1}^k \pi^{(j)} \left(f_i^{(j)}\right)^q \log_q \left(f_i^{(j)}\right) \\ &\quad - \sum_{j=1}^m \sum_{i=1}^k \pi^{(j)} \left(f_i^{(j)}\right)^q \log_q(f_i), \end{aligned} \quad (3.17)$$

donde  $\log_q$  está dado por la ecuación 3.3. Aquí estamos usando la notación  $F^{(j)} = \{f_i^{(j)}, j = 1, \dots, k\}$ . Luego, la ecuación (3.17) se puede escribir como:

$$JS_q = \sum_{j=1}^m \sum_{i=1}^k \pi^{(j)} \left(f_i^{(j)}\right)^q \left[ \log_q \left(f_i^{(j)}\right) - \log_q(f_i) \right]. \quad (3.18)$$

Usando la siguiente propiedad del  $\log_q$ ,

$$B^{(1-q)} \log_q \left( \frac{A}{B} \right) = \log_q A - \log_q B,$$

tenemos,

$$JS_q = \sum_{j=1}^m \sum_{i=1}^k \pi^{(j)} \left( f_i^{(j)} \right)^q f_i^{(1-q)} \log_q \left( f_i^{(j)} / f_i \right). \quad (3.19)$$

Estudiamos qué sucede cuando  $f_i \sim f_i^{(j)}$ , es decir cuando la probabilidad de encontrar el símbolo  $i$  en la secuencia  $j$  es aproximadamente igual a la probabilidad de encontrar el símbolo  $i$  en toda la secuencia. Para ello tengamos presente el desarrollo:

$$\log_q(x/a) \sim \frac{x-a}{a} + \frac{(x-a)^2}{2a^2} + (1-q) \frac{(x-a)^2}{2a^2}. \quad (3.20)$$

Entonces, podemos escribir la ecuación (3.19) en la forma:

$$JS_q = \sum_{j=1}^m \sum_{i=1}^k \pi^{(j)} \left( f_i^{(j)} \right)^q f_i^{(1-q)} \times \left[ \frac{\left( f_i^{(j)} - f_i \right)}{f_i} + \frac{\left( f_i^{(j)} - f_i \right)^2}{2f_i^2} + (1-q) \frac{\left( f_i^{(j)} - f_i \right)^2}{2f_i^2} \right]. \quad (3.21)$$

En el orden de aproximación en el que estamos trabajando,  $f_i^{(j)} / f_i \sim 1$  y por lo tanto el término de primer orden  $\rightarrow 0$ :

$$JS_q = \sum_{j=1}^m \sum_{i=1}^k \pi^{(j)} \left( \frac{f_i^{(j)}}{f_i} \right)^q \times \left( f_i^{(j)} - f_i \right) + \sum_{j=1}^m \sum_{i=1}^k \pi^{(j)} \left( f_i^{(j)} \right)^q f_i^{(1-q)} \times \left[ \frac{\left( f_i^{(j)} - f_i \right)^2}{2f_i^2} + (1-q) \frac{\left( f_i^{(j)} - f_i \right)^2}{2f_i^2} \right] \quad (3.22)$$

Puesto que  $\left( \frac{f_i^{(j)}}{f_i} \right)^q \sim 1$  y  $\sum_{j=1}^m \sum_{i=1}^k \pi^{(j)} \left( f_i^{(j)} - f_i \right) = 0$ , se tiene

$$JS_q = \frac{\chi^2}{2L} + (1-q) \frac{\chi^2}{2L}, \quad (3.23)$$

donde  $\chi^2 = L \sum_j \sum_i \frac{f_i^{(j)} - f_i}{f_i^2}$

La distribución de probabilidades para la cantidad  $\chi^2$  converge, para valores grandes de  $L$ , a la distribución  $\chi^2$  con  $\nu$  grados de libertad [79]<sup>3</sup>. Luego  $\frac{2L.JS_q}{2-q}$ , converge a la distribución  $\chi^2$ , lo que nos permite estimar  $s_q(x)$ ,

$$s_q(x) \sim F_\nu \left( \frac{2L}{2-q} .x \right) \equiv \frac{\gamma(\nu/2, L.x/(2-q))}{\Gamma(\nu/2)} \quad (3.24)$$

donde  $\Gamma$  y  $\gamma$  son las funciones gamma de primera y segunda especie, respectivamente y  $\nu = (k-1)(m-1)$ .

Para  $q = 1$  se reobtiene el valor obtenido por otros autores para el nivel de significancia:  $s_1(x) \sim \frac{\gamma(\nu/2, N.x)}{\Gamma(\nu/2)}$  [58].

### 3.4. Propiedad de Ramificación

En general las propiedades estadísticas de una secuencia simbólica dependen del alfabeto que se usa para describirla. Existen casos en lo que esto es particularmente relevante [80].

Sea  $\Gamma_j, j = 1, \dots, r$   $r \leq k$  una partición del alfabeto  $\mathcal{A}$ . A cada miembro de esta partición le asignamos un caracter  $\gamma_j$  (en general, tomado de otro listado de símbolos). Vamos a denotar por  $\Gamma = \{\gamma_1, \dots, \gamma_r\}$  a este nuevo alfabeto (alfabeto agrupado). A partir de la secuencia  $\mathcal{S}$ , con caracteres del alfabeto  $\mathcal{A}$ , podemos construir una nueva secuencia con caracteres pertenecientes al alfabeto  $\Gamma$ , sustituyendo cada símbolo  $a_i$  por  $\gamma_j$ , si  $a_i \in \Gamma_j$ . De este modo se puede relacionar las probabilidades de ocurrencia de determinado símbolo del primero de los alfabetos con las probabilidades de ocurrencia del correspondiente símbolo del alfabeto agrupado.

La entropía de Boltzmann-Gibbs tiene una interesante propiedad, conocida como propiedad de ramificación, la cual permite relacionar las entropías evaluadas usando las probabilidades calculadas con los diferentes alfabetos. Esta propiedad dice que:

$$H[\mathcal{S}, \mathcal{A}] = H[\mathcal{S}, \Gamma] + \sum_{j=1}^r \Theta_j H[\mathcal{S}, \Gamma_j], \quad (3.25)$$

donde  $\Theta_j = k_j/L$  y  $k_j$  es el número de símbolos de  $\mathcal{S}$  que pertenecen a  $\Gamma_j$ . Se puede ver que la propiedad de ramificación para la entropía de Tsallis se escribe:

$$H_q[\mathcal{S}, \mathcal{A}] = H_q[\mathcal{S}, \Gamma] + \sum_{j=1}^r \Theta_j^q H_q[\mathcal{S}, \Gamma_j], \quad (3.26)$$

Teniendo presente la relación (2.11), vemos que la propiedad de ramificación (3.25) para

---

<sup>3</sup> $P(\chi^2, \nu)$  es la probabilidad de que el valor observado  $\chi^2$  para un modelo correcto sea menor que un valor  $\chi^2$ ,  $P(\chi^2, \nu) = \frac{\gamma(\nu/2, \chi^2/2)}{\Gamma(\nu/2)}$ .



la entropía de BG, tiene una contrapartida en la divergencia  $JS$ :

$$JS^{(\pi^{(1)}, \pi^{(2)})}[\mathcal{A}] = JS^{(\pi^{(1)}, \pi^{(2)})}[\Gamma] + \sum_j \Theta_j JS^{(\tilde{\pi}^{(1)}, \tilde{\pi}^{(2)})}[\Gamma_j], \quad (3.27)$$

donde se han escrito explícitamente los alfabetos que se usan en cada término. En esta expresión

$$\tilde{\pi}^{(1)} = \pi^{(1)} \theta_j^{(\alpha)} / \Theta_j, \quad \tilde{\pi}^{(2)} = \pi^{(2)} \theta_j^{(\beta)} / \Theta_j,$$

y

$$\theta_j^{(\alpha)} = k_j^{(\alpha)} / N_1, \quad \theta_j^{(\beta)} = k_j^{(\beta)} / N_2$$

con  $k_j^{(\alpha)}$  ( $k_j^{(\beta)}$ ) el número de símbolos pertenecientes a  $\Gamma_j$  en las subsecuencias  $\mathcal{S}^{(1)}$  ( $\mathcal{S}^{(2)}$ ). Esta expresión nos permite evaluar la DJS entre dos subsecuencias, en término de las divergencias parciales entre las mismas dos subsecuencias asociada con el alfabeto agrupado  $\Gamma$  y los alfabetos parciales  $\Gamma_j$ .

Lamentablemente, la deducción de la propiedad de ramificación para la q-DJS no es tan directa. Sin embargo, con un poco de álgebra es posible arribar a ella:

$$\begin{aligned} JS_q^{(\pi^1, \pi^2)}[\mathcal{A}] &= JS_q^{(\pi^{(1)}, \pi^{(2)})}[\Gamma] + \sum_j \Theta_j \frac{[(\theta_j^{(\alpha)})^{q-1} + (\theta_j^{(\beta)})^{q-1}]}{2} JS_q^{(\tilde{\pi}^{(1)}, \tilde{\pi}^{(2)})}[\Gamma_j] + \\ &\frac{(1-q)}{2} \sum_j \Theta_j \tilde{\pi}^1 (\theta_j^{(\alpha)})^{q-1} K_q[\alpha_i^{(j)}, p_i^{(j)}] \left( 2 \log_q \theta_j - \log_q \frac{\theta_j^{(\alpha)}}{\theta_j^{(\beta)}} \right) + \\ &\frac{(1-q)}{2} \sum_j \Theta_j \tilde{\pi}^2 (\theta_j^{(\alpha)})^{q-1} K_q[\beta_i^{(j)}, p_i^{(j)}] \left( 2 \log_q \theta_j - \log_q \frac{\theta_j^{(\beta)}}{\theta_j^{(\alpha)}} \right) \end{aligned} \quad (3.28)$$

donde  $\alpha_i^{(j)}$  y  $\beta_i^{(j)}$  son las probabilidades de ocurrencia del símbolo  $a_i$  en las subsecuencias  $\mathcal{S}^{(1)}$  y  $\mathcal{S}^{(2)}$  respectivamente, considerando el alfabeto  $\Gamma_j$ ;  $p_i^{(j)}$  es la probabilidad de ocurrencia del símbolo  $a_i$  en toda la secuencia  $\mathcal{S}$ , considerando sólo los símbolos que pertenecen a  $\Gamma_j$ .  $K_q$  es la divergencia Kullback no logarítmica dada por la ecuación (3.6). Para  $q = 1$  reobtenemos la identidad (3.27). Los últimos términos en la ecuación (3.28) están relacionados al carácter no extensivo del formalismo. Es importante notar que este término involucra los alfabetos  $\Gamma$  y  $\Gamma_j$ .

### 3.5. El carácter métrico de la q-DJS

En esta sección demostraremos una de las propiedades más importante de la DJS: su carácter métrico. Esta propiedad es de gran relevancia formal y práctica. Por un lado, muy pocas de las distancias entre distribuciones de probabilidad propuestas son efectivamente métricas. Por el otro, el carácter métrico de una distancia tiene gran importancia cuando se desarrollan esquemas iterativos, pues permite enunciar criterios de convergencia del proceso recursivo.

Ya hemos dicho que para  $\pi^{(1)} = \pi^{(2)} = \frac{1}{2}$ ,  $JS_q^{(\pi^{(1)}, \pi^{(2)})}$  es definida positiva y simétrica, y por lo tanto su raíz cuadrada satisface ambas propiedades. Demostraremos ahora que la raíz cuadrada  $D_q \equiv \sqrt{2JS_q}$  cumple además la desigualdad triangular. Esta propiedad para la  $JS$  ha sido probada por otros autores de dos formas distintas [57, 56]. En nuestra demostración para la q-JSD, seguiremos a Endres et al [57].

Usando la siguiente propiedad del  $\log_q$ :

$$\log_q(a/b) = \frac{\log_q(a) - \log_q(b)}{1 + (1 - q) \log_q(b)}, \quad (3.29)$$

es posible reescribir  $2JS_q$  como la suma de términos de la forma:

$$Q_q[x, y] = -x \log_q \left( \frac{x+y}{2x} \right) - y \log_q \left( \frac{x+y}{2y} \right). \quad (3.30)$$

Notemos que  $Q_q[x, y] \geq 0 \quad \forall q > 0$  y  $\forall x, y \geq 0$ . Si  $Q_q$  satisface la desigualdad:

$$\sqrt{Q_q[x, y]} \leq \sqrt{Q_q[x, z]} + \sqrt{Q_q[z, y]} \quad \forall x, y, z \geq 0 \quad (3.31)$$

entonces  $D_q$  satisface la desigualdad triangular. Esta última afirmación se puede demostrar usando la desigualdad de Minkowski. Por lo tanto bastará demostrar la desigualdad (3.31).

Introduzcamos la función

$$F_q(\psi, \zeta, \chi) = \sqrt{Q_q(\psi, \zeta)} + \sqrt{Q_q(\zeta, \chi)}, \quad (3.32)$$

y la pensemos como una función de  $\zeta$  para  $\psi$  y  $\chi$  fijos. En  $\zeta = \psi$  y  $\zeta = \chi$ , la función  $F_q$  toma el valor  $\sqrt{Q_q(\psi, \chi)}$ . Veremos que en estos dos puntos  $F_q$  alcanza su valor mínimo. Si esto es cierto, entonces la desigualdad (3.31) efectivamente se cumple. En la figura 3.4 se muestra un gráfico de la función  $F_q$  para  $q = 2$ .

Estudiemos ahora la derivada  $\frac{\partial F_q}{\partial \zeta}$ . Tras un poco de álgebra se obtiene:

$$2\sqrt{\zeta} \frac{\partial F_q(\psi, \zeta, \chi)}{\partial \zeta} = g_q(x) + g_q(ax), \quad (3.33)$$

con  $x = \psi/\zeta$  y  $a = \chi/\psi > 1$  (suponemos  $\psi < \chi$ ; si  $\psi = \chi = 0$  la desigualdad (3.31) se satisface trivialmente) y

$$g_q(x) = \frac{\frac{1-q}{2} \left(\frac{x+1}{2x}\right)^{-q} \log_q(x) - \log_q\left(\frac{x+1}{2}\right)}{\sqrt{Q_q(x, 1)}}. \quad (3.34)$$

Esta función es discontinua en  $x = 1$ . En particular se puede mostrar (aplicando dos veces la regla de L'Hopital) que

$$\lim_{x \rightarrow 1} g_q^2(x) = q$$

y

$$\lim_{x \rightarrow 1^\pm} g_q(x) = \mp \sqrt{q}.$$

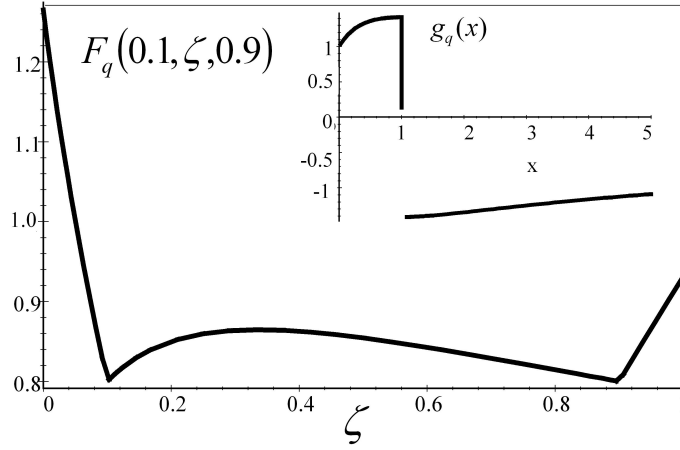


Figura 3.4:  $F_q$  en función de  $\zeta$ , para valores fijos de  $\psi$  y  $\chi$ , y para  $q = 2$

Además, se puede verificar que

$$\lim_{x \rightarrow 0} g_q(x) = \sqrt{-\log_q(1/2)}$$

y

$$\lim_{x \rightarrow \infty} g_q(x) = 0.$$

Entonces,

$$|g_q(x)| \leq \sqrt{q} \quad \forall q \leq q^*$$

y  $\forall x$ . El número  $q^*$  es el mayor valor de  $q$  tal que  $\frac{dg_q(x)}{dx} \geq 0$  (excepto para  $x = 1$ , donde esta derivada no está definida)

Nótese que si  $q \leq q^*$ , entonces  $\sqrt{-\log_q(1/2)} < \sqrt{q}$ .

Consideremos ahora tres regiones:

- i)  $\zeta < \psi$
- ii)  $\psi < \zeta < \chi$
- iii)  $\zeta > \chi$

En cada una de ellas, se cumple:

- i)  $\zeta < \psi$ . Tenemos  $x > 1$  y  $ax > 1$ . Por eso,

$$g_q(x) + g_q(ax) < 0,$$

lo que significa que la derivada  $\frac{\partial F_q}{\partial \zeta} < 0$ .

ii)  $\psi < \zeta < \chi$ . En este caso  $x < 1$  y  $x > 1/a$ . Para  $\zeta \sim \psi$ ,  $x \sim 1^-$  y entonces,

$$g_q(x) + g_q(ax) > 0.$$

Para  $\zeta \sim \chi$ ,  $x \sim 1/a$ , y por esto se cumple,

$$g_q(x) + g_q(ax) < 0.$$

Concluimos que, entre  $\psi$  y  $\chi$ , la función  $F_q$  tiene un máximo local.

iii)  $\zeta > \chi$ . Finalmente en esta región  $x < 1$  y  $ax < 1$ , y en consecuencia

$$g_q(x) + g_q(ax) > 0,$$

esto es, en esta región la derivada de  $F_q$  es nuevamente positiva.

Así podemos concluir que la función  $F_q$  tiene a lo sumo dos mínimos: uno en  $\zeta = \psi$  y el otro en  $\zeta = \chi$ . Es importante tener en cuenta que las desigualdades verificadas por  $g_q(x) + g_q(ax)$ , se satisfacen porque la derivada  $dg_q/dx \geq 0$  (ver inserto de la figura). Por esta razón la desigualdad (3.31) vale solamente si  $q$  pertenece al intervalo  $[1, q^*]$ .

La condición  $dg_q/dx \geq 0$  se puede estudiar numéricamente. Haciendo esto se encuentra que  $q^* = 4,4$ .

Se puede pensar que el requerimiento  $q \in [1, q^*]$  es restrictivo e inusual; sin embargo varios resultados interesantes obtenidos en el marco de la  $q$ ME, se pueden derivar para valores de  $q$  dentro de este intervalo. Además, en el seno del análisis funcional, este tipo de restricciones es bastante usual, ver por ejemplo referencia [81].

Teniendo en cuenta la dependencia de la  $q$ DJS con un parámetro continuo (el parámetro  $q$ ), hemos demostrado el siguiente teorema:

**Teorema 1** Para  $q \in [1, q^*]$ , la familia de funciones

$$D_q : M_+^1(N) \times M_+^1(N) \longrightarrow \mathfrak{R} \quad (3.35)$$

con

$$M_+^1(N) = \{p_i \in \mathfrak{R} : p_i \geq 0 \text{ y } \sum_i p_i = 1\} \quad (3.36)$$

constituye una familia monoparamétrica de métricas para el espacio  $M_+^1(N)$ .

De hecho el carácter métrico de  $D_q$  vale para el caso  $q = 1$ . Topsøe [56] ha demostrado esta propiedad recurriendo a algunos resultados del análisis armónico clásico. Por ejemplo, se puede verificar que la DJS satisface todas las hipótesis del siguiente teorema de Schoenberg [82]:

**Teorema 2 (Schoenberg)** *Sea  $\mathcal{X}$  un conjunto y  $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathfrak{R}$  una aplicación (un núcleo) tal que, para  $(x, y) \in \mathcal{X} \times \mathcal{X}$*

$$\begin{aligned} K(x, y) &= 0 \Leftrightarrow x = y \\ K(x, y) &= K(y, x) \\ K(x, y) &\geq 0 \end{aligned} \tag{3.37}$$

*y tal que, para todo conjunto finito  $(c_i)_{i \leq n}$  de números reales y todos los conjuntos finitos  $(x_i)_{i \leq n}$  de puntos de  $\mathcal{X}$ , la implicación*

$$\sum_{i=1}^n c_i = 0 \Rightarrow \sum_{i,j} c_i c_j K(x_i, x_j) \leq 0 \tag{3.38}$$

*vale, entonces  $(\mathcal{X}, \sqrt{K})$  es un espacio métrico que puede ser inmerso isométricamente como un subespacio de un espacio de Hilbert real.*

Este hecho hace que además de emerger el carácter métrico para  $D \equiv \sqrt{2JS}$ , se pueda establecer un isomorfismo métrico (isometría) entre el espacio  $(M_+^1(N), \sqrt{2JS})$  y un subconjunto de algún espacio de Hilbert real. Como veremos en el capítulo 5, este resultado tiene su relevancia en el marco de la TI cuántica.

## 3.6. Comentarios Finales

La DJS es una excelente herramienta estadística en procesos de segmentación. En este capítulo hemos propuesto una generalización de la DJS en el marco de la qME. Estudiamos sus propiedades más importantes en el contexto del análisis de secuencias simbólicas. Probamos su carácter métrico en el espacio  $M_+^1(N)$ , y definimos una familia monoparamétrica de métricas asociada al parámetro de deformación  $q$ .



# Capítulo 4

## Complejidad de Secuencias Simbólicas

*El estudio de los entes vivos es la mejor manera de apreciar cuán primitiva es aún la Física.*<sup>1</sup>

En este capítulo introducimos una medida de complejidad para secuencias simbólicas. Ella se basa en el proceso de segmentación mencionado en el capítulo anterior y se define como la entropía de la distribución de las longitudes de los dominios relativamente uniformes en los que se descompone la secuencia bajo estudio.

Antes de definir la medida de complejidad propuesta, revisaremos otras cantidades sugeridas como medidas de complejidad. Aprovecharemos esta revisión, para repasar las principales propiedades que una medida de complejidad debe poseer.

### 4.1. Introducción

En los últimos años, el término *complejidad* se ha vuelto muy frecuente en la literatura científica [83]. Esto ha llevado a la introducción de diversas medidas de complejidad en distintos contextos temáticos [84], [85], [86], [87]. Sin embargo, no hay aún un consenso acerca de una definición precisa de complejidad que permita su cuantificación. Quizás una de las mayores dificultades es la falta de un lenguaje común entre las diferentes áreas de la ciencia en las que el concepto se desea introducir.

Por ejemplo, si la complejidad se incrementa o no a lo largo de la evolución, es una de las cuestiones centrales de la biología evolutiva. Esta cuestión ha recibido respuestas marcadamente diferentes entre los especialistas. Hay quienes sugieren que la complejidad se incrementa con la evolución, otros que sostienen que aún no hay suficiente evidencia a favor o en contra de tal incremento, y los que lo rechazan completamente. A pesar de estas opiniones tan dispares, hay un acuerdo sobre la carencia de un significado preciso del

---

<sup>1</sup>A. Einstein.

concepto de complejidad cuando se refiere a un organismo biológico. Esta situación puesta de manifiesto en el ámbito de la biología se repite en otras áreas del conocimiento.

Como hemos mencionado en los capítulos anteriores, la incerteza asociada a un proceso probabilístico se puede caracterizar a través de la entropía de Shannon de la distribución de probabilidad asociada con ese proceso. En los años 50 del siglo pasado, Kolmogorov y Sinai adaptaron la TI de Shannon al estudio de sistemas dinámicos. De allí surgieron medidas tales como la entropía métrica, los exponentes de Lyapunov y la dimensión fractal, los cuales se usan en Física para detectar y cuantificar comportamientos caóticos deterministas. Sin embargo estas medidas de aleatoriedad e impredecibilidad no sirven para capturar la *estructura* que emerge de la inter-relación de las partes de un sistema.

Por otro lado, hay acuerdo en que un sistema completamente ordenado y uno totalmente desordenado no poseen estructura, mientras que un sistema complejo puede involucrar orden y desorden simultáneamente, ubicándose entre los dos extremos. Esto se ve claramente en el ámbito de la Física, donde dos casos límites deben considerarse. Por un lado, un cristal perfecto (sistema totalmente ordenado) y por el otro, el gas ideal (representante de un sistema completamente desordenado). Ambos sistemas son de complejidad nula (o extremadamente baja). En general, una medida de complejidad apropiadamente definida debería alcanzar su máximo en algún nivel intermedio entre el orden de lo completamente regular y el desorden de lo totalmente aleatorio. Esta característica deseable para una medida de complejidad se conoce como la propiedad de *joroba de camello*.

En cierto sentido, introducir una medida de complejidad, es desarrollar un método o esquema para determinar la *estructura* de un sistema. Entre estos métodos podemos distinguir los deterministas y los estadísticos.

Entre las medidas deterministas cabe mencionar la complejidad de Kolmogorov - Chaitin [88] (descrita ya en la Introducción) y la complejidad de Lempel-Ziv [89], ambas definidas para secuencias simbólicas. Estas medidas se basan en un proceso de compresión de la secuencia. De ello resulta que una secuencia aleatoria tiene un valor grande (infinito) de ambas complejidades, y en este sentido se piensan más como medidas de aleatoriedad que de complejidad estructural.

Entre las medidas de complejidad estadística las que mayor atención han recibido son la profundidad termodinámica, introducida por Lloyd y Pagels [90], que mide también la aleatoriedad de un sistema. En la misma categoría están la medida introducida por Crutchfield y Young [91], la cual satisface la propiedad de *joroba de camello*, y la complejidad Lopez-Ruiz, Mancini y Calbet [3], descrita en la Introducción.

Muchas veces, los procesos físicos se pueden considerar como algoritmos que describen la dinámica del sistema. Por ello, muchas de las medidas de complejidad propuestas resultan ser medidas “computacionales”. Ejemplo de esto es la profundidad lógica de Bennet, la cual mide el (menor) tiempo requerido para correr un programa que reproduce la configuración del sistema. Otras medidas del mismo tipo son la de Grassberger [92], y la de Koppel. No obstante lo original de estas ideas, tienen sus limitaciones. Por caso, la complejidad de un organismo biológico no puede cuantificarse caracterizando la dinámica de los procesos relacionados. Las medidas de complejidad biológica se refieren a la forma, a la función o a la secuencia genética que codifica la información del organismo bajo estudio.



Generalmente cuando se consideran seres vivos se tiene interés en la complejidad estructural. Así McShea [93] ha estudiado varias medidas de complejidad estructural basadas en el número de tipos de células, diferentes tipos de pares de miembros, y dimensión fractal de estructuras en los ammonoides. Este autor [94] también introdujo una medida para la complejidad funcional de los organismos. Otras propuestas de medidas de complejidad, realizadas desde una perspectiva biológica, se pueden encontrar en los trabajos de Bell y Mooers [95]. Todos estos intentos de describir la complejidad de un organismo, dejan en claro lo dificultosa que es la tarea de derivar una medida universal para la complejidad estructural o funcional, dada la variedad de formas y funciones que exhiben los seres vivos en la naturaleza.

Una de las maneras de aproximarse a la definición de una medida de complejidad en biología, es considerar la secuencia de nucleótidos que constituyen el genoma de la especie. Esta complejidad, debería ser un reflejo de la complejidad del organismo al cual esa secuencia da origen [83]. Adami y colaboradores [96] introdujeron una medida de complejidad para secuencias genómicas, denominada *complejidad física*, a través de las correlaciones entre los símbolos de la secuencia del genoma y una descripción del entorno en el que esta secuencia es funcional. La complejidad física se refiere a la cantidad de información que almacena una secuencia sobre un entorno particular. Para un genoma, este entorno es el ambiente en el que vive y se replica el organismo que lo alberga. Técnicamente la complejidad física se define como la complejidad Kolmogorov compartida entre la secuencia y la descripción del ambiente en el que la secuencia va a ser interpretada. Ella es un ejemplo de la complejidad efectiva desarrollada en forma independiente por Gell-Mann y Lloyd [84].

## 4.2. Sistemas complejos

Hay consenso en caracterizar a los sistemas complejos como sistemas que no se pueden descomponer ni analizar por partes. Los sistemas complejos están formado por muchos elementos no lineales que interactúan entre sí [97]. Estas interacciones le dan al sistema la capacidad de auto-organizarse [98]. Dado que la complejidad emerge de las interacciones de las unidades individuales, éstas deben estar presentes al momento de definir una medida que cuantifique la complejidad de un sistema. Cuando las diferentes partes de un sistema, por ejemplo, las moléculas de un gas ideal en equilibrio, no interactúan entre sí, su comportamiento global se puede entender en términos de la suma simple de sus partes separadas. Pero, una vez que aparecen interdependencias, esto no es más válido y para cuantificar la complejidad necesitamos una medida que considere esos vínculos [86].

Una medida de complejidad adecuada debería tener en cuenta este hecho cuando se aplica a un sistema complejo, asignando un valor superior al de la suma de las complejidades de sus partes. Es decir, la complejidad del *todo* es mayor o igual que la suma de las complejidades de las partes. A esta propiedad la llamaremos súper-aditividad.

En el estudio de secuencias la propiedad de súper-aditividad para la complejidad (extensiva) se puede escribir de la siguiente forma:

$$(L_1 + L_2) C_S \geq L_1 C_{S_1} + L_2 C_{S_2} \quad (4.1)$$

donde  $C_S$  denota la complejidad de la secuencia completa y  $C_{S_1}$  ( $C_{S_2}$ ) las complejidades de dos subsecuencias arbitrarias  $S_1$  ( $S_2$ ) con longitudes  $L_1$  ( $L_2$ ) respectivamente, y tales que su concatenación da la secuencia completa  $S$  ( $L_1 + L_2 = L$ ).

Concluimos esta sección notando que en el estudio de los sistemas complejos, las secuencias simbólicas tienen un rol importante. Tal como afirma Adami [83], en muchos casos específicos, es posible establecer esquemas de relación entre el sistema bajo estudio y secuencias simbólicas. Ejemplo de ello, lo proveen los esquemas:

- sistema dinámico  $\leftrightarrow$  secuencia de símbolos que produce (1)
- organismo vivo  $\leftrightarrow$  secuencia de bases de su genoma (2)

### 4.3. Una medida de complejidad

El punto de partida de nuestra definición es el proceso de segmentación de la secuencia bajo estudio. En la subsección siguiente haremos una descripción detallada del mismo. Por lo pronto tengamos presente que el proceso de segmentación permite separar a la secuencia en trozos o parches (segmentos) de uniformidad composicional.

Si prestamos atención a las longitudes de esos parches, tras el proceso de segmentación pueden darse dos casos extremos:

- i.* todos los segmentos resultantes tienen la misma longitud (secuencia periódica),
- ii.* la secuencia no ha sido segmentada (secuencia aleatoria).

Estos dos casos, que están en correspondencia con el cristal y el gas ideal, tendrán, de acuerdo a nuestra definición, una complejidad igual a cero. Queda entonces caracterizar a la distribución de largos de los segmentos en los cuales la secuencia ha sido descompuesta, que tomaremos como la de mayor complejidad.

Cuando la probabilidad de medir un valor particular de alguna cantidad varía inversamente como una potencia de ese valor, se dice que la cantidad sigue una ley de potencia. La importancia de las distribuciones que siguen una ley de potencias en Física y áreas relacionadas, ha quedado evidenciada por la ubicuidad de tales leyes en una gran variedad de fenómenos. Este tipo de ley rige tanto la frecuencia del uso de palabras en cualquier lenguaje humano, como el número de cráteres lunares de un determinado tamaño [99]. Más interesante aún, este tipo de comportamiento emerge también al estudiar la dinámica cerebral. En efecto, se sabe que el cerebro crea continuamente redes funcionales complejas correspondientes al tráfico entre regiones. En ese caso se encuentra que la probabilidad de que  $k$  regiones estén temporalmente correlacionadas con una determinada región, satisface

una ley de la forma  $k^{-\mu}$  con  $\mu \approx 2$  [97]. Este ejemplo resulta, a nuestro entender, altamente significativo, pues la dinámica cerebral es un caso paradigmático de lo que se conoce como auto-organización, y sin lugar a dudas, de lo que podemos entender como un sistema complejo. A su vez, la auto-organización se reconoce como el mecanismo que modela a una gran cantidad de sistemas en la Naturaleza.

De acuerdo con lo mencionado previamente, consideramos razonable tomar como una secuencia de alta complejidad, a una que tenga una distribución de longitudes de segmentos de composición relativamente uniforme, dada por una ley de potencia, es decir, la probabilidad  $P(l)$  de encontrar un segmento de composición relativamente homogénea con longitud  $l$ , está dada por:

$$P(l) \sim \frac{1}{l^\mu}. \quad (4.2)$$

Supondremos además que la secuencia más compleja es aquella que tiene máxima interdependencia entre sus segmentos, y para cuantificar esta interdependencia, recurrimos a la función de autocorrelación  $C(l)$ . La interdependencia es máxima cuando esta función es plana. Existe una relación interesante entre el exponente  $\mu$  en la ecuación (4.2), y el comportamiento de la función de autocorrelación [12]. Para una ley de distribución de longitudes dada por (4.2), se puede mostrar que la desviación estándar del contenido de símbolos de la secuencia,  $F(l)$ , se comporta de la forma:

$$F(l) \sim l^\alpha \quad (4.3)$$

y la función autocorrelación sigue una ley de potencia

$$C(l) \sim \frac{1}{l^\gamma} \quad (4.4)$$

con  $\gamma = 2 - 2\alpha$ . Para un exponente  $\mu \leq 2$  corresponde un exponente  $\alpha = 1$  y por lo tanto  $\gamma = 0$ , es decir se tiene una función autocorrelación plana [12]. Entonces para secuencia muy largas, una autocorrelación plana está asociada con una distribución de longitudes de segmentos que obedece una ley de potencia con exponente  $\mu \leq 2$ .

Notemos que cualquier exponente  $\mu \leq 2$  lleva a una función de autocorrelación plana, sin embargo, el exponente  $\mu = 1$  corresponde a una distribución de patrones autosimilar a lo largo de la secuencia [100]. Estas consideraciones sugieren tomar como la secuencia más compleja a una con distribución de longitudes de segmentos de composición relativamente uniforme, dada por la ley (4.2) con  $\mu = 1$ .

Más adelante surgirán otros motivos, de índole práctica, que refuerzan la elección  $\mu = 1$ .

### 4.3.1. Método de segmentación

El método de segmentación de secuencias que describiremos a continuación es consistente con el requisito de comparar el todo con las partes. El mismo hace uso de la divergencia de Jensen-Shannon (2.10) y se ha aplicado exitosamente al estudio de secuencias de ADN [101]. La DJS permite cuantificar la diferencia composicional de los segmentos tentativos

en los cuales la secuencia podría separarse. Es importante destacar que el método de segmentación permite poner de manifiesto la estructura fractal compleja encontrada en las secuencias de ADN, dada por la presencia de correlaciones de corto y largo alcance [102].

Los dominios composicionales están definidos como subsecuencias con una composición de bases diferente a la de las dos subsecuencias adyacentes para un dado nivel de umbral. El nivel de umbral está relacionado con el nivel de detalle con el que se analiza la secuencia y al variarlo se obtienen distintas particiones correspondientes a diferentes niveles de significación estadística. Para valores grandes del umbral se obtienen pocos segmentos pero con una gran diferencia estadística entre ellos. Por el contrario, para umbrales pequeños se obtiene un mayor número de segmentos, pero con una diferencia estadística no muy significativa.

Con el fin de hacer el presente trabajo autocontenido, describiremos brevemente el procedimiento de segmentación. Para más detalles ver referencia [101]. Un cursor se desplaza a lo largo de cada posición de la secuencia completa. La posición del cursor determina dos subsecuencias, la que está a su izquierda y la que está a su derecha. Se calculan las frecuencias relativas de ocurrencia de los distintos símbolos para cada una de las subsecuencias y con ellas, el valor de la divergencia en cada paso. La posición del cursor que corresponde al máximo se elige como punto de corte, tras verificar que el valor máximo alcanzado por la divergencia supere el valor de umbral elegido. Este proceso se repite con las dos subsecuencias resultantes hasta que ninguno de los valores de la divergencia superen el umbral. De esta manera se define un criterio para detener la iteración.

Si se segmenta una secuencia con una estructura simple, se pueden encontrar (excluyendo las fluctuaciones aleatorias) dominios homogéneos. Sin embargo, cuando la secuencia que se analiza presenta correlaciones de largo alcance, relajando el valor del umbral se encuentran nuevos dominios dentro de dominios que antes se habían tomado como homogéneos.

Para secuencias de ADN, la presencia de estos dominios dentro de dominios, indica una heterogeneidad composicional compleja, la cual es consistente con una jerarquía biológica [102].

### 4.3.2. Definición de la complejidad

Consideremos una secuencia simbólica  $\mathcal{S}$  de longitud  $L$ . Vamos a suponer que por un proceso de segmentación como el descrito en la sección anterior podemos segmentar la secuencia en  $N_s$  dominios de diferente contenido composicional. Sean  $l_i, i = 1 \dots N_s$ , las longitudes de los segmentos obtenidos después de la segmentación, de modo que:

$$\sum_{i=1}^{N_s} l_i = L \quad (4.5)$$

En general estas longitudes no son todas distintas entre sí, es decir, podemos encontrar varios segmentos de la misma longitud.

Consideremos ahora una partición  $\mathcal{A} = \{A_j\}_{j=1}^\nu$  del intervalo  $[1, L]$ :

$$1 = A_1 < A_2 < \dots < A_{\nu-1} < A_\nu = L. \quad (4.6)$$

Denotamos por  $\tilde{N}_j$  el número de segmentos cuya longitud pertenece al intervalo  $[A_{j-1}, A_j)$ . Debe cumplirse la condición  $\sum_j \tilde{N}_j = N_s$ . Finalmente vamos a denotar por  $f_j$  la frecuencia de ocurrencia de los segmentos con longitudes en el intervalo  $[A_{j-1}, A_j)$ :

$$f_j = \frac{\tilde{N}_j}{N_s}; \quad \sum_j f_j = 1 \quad (4.7)$$

A partir de estas frecuencias  $F = \{f_j\}$ , podemos evaluar la entropía de la distribución de longitudes de los segmentos.

$$H_S[F, \mathcal{A}, D_u] = H_S[F] = - \sum_{j=2}^\nu f_j \log f_j \quad (4.8)$$

Para una secuencia  $\mathcal{S}$  dada, esta cantidad depende de la partición  $\mathcal{A}$  y del valor del nivel de significancia usado en la segmentación de la secuencia  $D_u$ , i.e. depende del detalle con el que la secuencia ha sido analizada. Es por esto que incluimos explícitamente la partición  $\mathcal{A}$  y el valor de  $D_u$  como uno de los argumentos de  $H_S$ .

Sin embargo hay dos casos en los que la entropía (4.8) no depende de la elección de una partición particular:

1. una secuencia *idealizada* periódica y
2. una secuencia *idealizada* aleatoria.

Aquí con *idealizada* queremos decir que su carácter regular o aleatorio se detecta para cualquier valor del nivel de detalle que se use en el proceso de segmentación. En el primer caso existe sólo un valor (el período) para la longitud de los segmentos. Tendremos  $f_J = 1$  para algún valor  $2 \leq J \leq \nu$  y  $f_j = 0$  para todo otro  $j$ . Entonces para una secuencia periódica  $H_S = 0$  para cualquier partición del intervalo  $[1, L]$ . Análogamente, debido al hecho de que una secuencia aleatoria no se segmenta para cualquier nivel de detalle significativo (por la propia definición de significativo), sólo una de las  $f_j$  es diferente de cero:  $f_\nu = 1$ . Tenemos también en este caso  $H_S = 0$ . Estos dos casos extremos son los correspondientes al cristal perfecto y al gas ideal en el contexto físico. Retomando la forma de los esquemas de la página 50, podemos escribir:

- cristal perfecto  $\leftrightarrow$  secuencia periódica
- gas ideal  $\leftrightarrow$  secuencia aleatoria

La entropía  $H_S$  tiene información sobre la segmentación de la secuencia, y se anula para orden y desorden completos, es decir satisface la propiedad de la joroba de camello. Estas características sugieren estudiarla como una medida de complejidad. Sin embargo, para caracterizar realmente la secuencia bajo estudio, la medida de complejidad debe ser independiente de cualquier parámetro arbitrario. Para esto debemos fijar una partición  $\mathcal{A}$ , de modo que una vez que hayamos hecho esa elección, la cantidad  $H_S$  no incluya explícitamente en sus argumentos a  $\mathcal{A}$ .

Al buscar caracterizar lo que entenderemos por la secuencia más compleja, haremos las siguientes consideraciones: pensemos que tras el proceso de segmentación, a un dado nivel de detalle, la secuencia  $\mathcal{S}$  se descompone en  $N_s$  segmentos de contenido composicional uniforme, y supongamos que podemos identificar una ley de potencia para la distribución de longitudes de esos segmentos:

$$N_l = \frac{N_s}{Z(\mu, \lambda^*)} l^{-\mu} \quad (4.9)$$

donde  $Z(\mu, \lambda^*) = \sum_{l=1}^{\lambda^*} l^{-\mu}$ ,  $\lambda^*$  es una longitud de corte. El valor de corte  $\lambda^*$  tiene que ver con el tamaño finito de la secuencia  $\mathcal{S}$ . Su valor se puede obtener de la condición:

$$N_s \frac{Z(\mu - 1, \lambda^*)}{Z(\mu, \lambda^*)} = L \quad (4.10)$$

De la ley de distribución (4.9), y para una partición dada  $\mathcal{A}$ , podemos evaluar las frecuencias de los segmentos

$$f_j = \frac{1}{N_s} \sum_{l \in [A_j, A_{j+1}-1]} N_l.$$

Hemos argumentado ya a favor de pensar que la secuencia más compleja será una que tiene una distribución de largos de la forma (4.9) con  $\mu = 1$ . Nos queda entonces buscar la partición  $\mathcal{A}$  que hace que la entropía (4.8) alcance un valor máximo. Debido a una propiedad fundamental de la entropía, ésta alcanza su valor máximo para una partición  $\mathcal{A}$  en la que todas las frecuencias  $f_j$  son iguales. Es decir, para que la entropía sea máxima el número de segmentos en cada intervalo  $[A_{j-1}, A_j)$  debe ser el mismo. Puesto que la secuencia es finita y existe una longitud de corte, existirá también un valor  $j^*$  tal que  $f_j = 0$  para  $j > j^*$ . Entonces el máximo de la entropía ocurre para el valor  $j^*$  más grande consistente con la condición de uniformidad para las  $f_j$ . Resulta, en este caso, un valor de entropía igual a  $\log j^*$ . El máximo será alcanzado para la secuencia *más estirada* posible, de forma de alcanzar el mayor valor  $j^*$ . Es por esto que debemos encontrar la partición del intervalo  $[1, L]$  para la cual el número de segmentos en cada uno de los intervalos consecutivos sea el mismo e igual a uno. Para más detalles ver Apéndice B. Estos requerimientos se pueden

expresar a través del siguiente conjunto de ecuaciones

$$\begin{aligned}
 1 + \frac{1}{2^\mu} + \dots + \frac{1}{(A_2 - 1)^\mu} &= \frac{1}{A_2^\mu} + \dots + \frac{1}{(A_3 - 1)^\mu} \\
 \frac{1}{A_3^\mu} + \dots + \frac{1}{(A_4 - 1)^\mu} &= \frac{1}{A_4^\mu} + \dots + \frac{1}{(A_5 - 1)^\mu} \\
 &\vdots \\
 \frac{1}{A_{j^*-2}^\mu} + \dots + \frac{1}{(A_{j^*-1} - 1)^\mu} &= \frac{1}{A_{j^*-1}^\mu} + \dots + \frac{1}{(\lambda^*)^\mu}
 \end{aligned} \tag{4.11}$$

con  $\mu = 1$ .

Como estamos buscando el máximo valor para  $j^*$  es obvio, a partir del conjunto de ecuaciones previas, que  $A_2$  debe ser igual a 2. Esto se debe a la condición de tener un segmento en cada uno de los intervalos. Usando el conjunto de ecuaciones (4.11) se pueden obtener el resto de los  $A_j$ , de tal forma de determinar la partición buscada.

Tras este razonamiento, llegamos a nuestra definición de complejidad de una secuencia  $S$  de longitud  $L$ :

$$C_S = H_S[F_L], \tag{4.12}$$

donde  $H_S[F_L]$  es la entropía de la distribución de segmentos calculada en la partición del intervalo  $[1, L]$  dada por (4.11) con  $\mu = 1$ .

La evaluación de la complejidad (4.12) para una secuencia arbitraria  $S$  de longitud  $L$  requiere:

1. Calcular la partición  $\mathcal{A}$  correspondiente a la longitud  $L$  de acuerdo a (4.11) para  $\mu = 1$ .
2. Segmentar la secuencia para un cierto valor del nivel de significancia  $D_u$ , y determinar las frecuencias dadas por (4.7) para la partición  $\mathcal{A}$ .
3. Finalmente evaluar la entropía  $H_S$  dada por la ecuación (4.8).

Como ya ha sido dicho, hay también motivos prácticos para tomar el exponente  $\mu = 1$ . En efecto, para valores mayores de  $\mu$  compatibles con la condición de autocorrelación plana ( $\mu \leq 2$ ), la entropía  $H_S[F_L]$ , evaluada siguiendo los pasos recién indicados, toma valores extremadamente bajos (en la mayoría de los casos cero). Para secuencias reales, las longitudes de los segmentos que se obtienen después del proceso de segmentación, quedan todas dentro del mismo intervalo de la partición si se elige  $\mu > 1$ , ya que las particiones resultan cada vez vez más *estiradas* a medida que se toman valores de  $\mu$  más grandes.

## 4.4. Aplicaciones y Resultados

Un aspecto deseable para una medida de complejidad es que dependa del nivel de detalle con el que se analiza el sistema en estudio [103]. Llamativamente, muy pocas de las medidas

de complejidad introducidas para la caracterización de secuencias simbólicas, lo satisface. En particular, la medida introducida en la sección anterior lo cumple. En este caso, el nivel de detalle está cuantificado a través del nivel de umbral usado en la segmentación,  $D_u$ . En los ejemplos presentados a continuación, se pondrá de manifiesto esa dependencia a través de gráficos conocidos como *perfiles de complejidad*.

### Aplicación a secuencias genómicas

Aplicaremos ahora la medida de complejidad  $C_S$  a secuencias reales de ADN con diferentes estructuras en lo que al tipo de correlación entre sus componentes (nucleótidos) se refiere. En todos los ejemplos usamos para la segmentación el alfabeto *normal* o de cuatro símbolos  $\{A, C, G, T\}$ . Las evaluaciones de la medida de complejidad propuesta, nos permiten, por un lado, estudiar sus principales propiedades, tales como la dependencia con el nivel de detalle y la propiedad de aditividad, y por el otro podemos estudiarla como una herramienta adecuada para revelar ciertas características estructurales en el ADN, por ejemplo el contenido de intrones y exones, y su relación con aspectos evolutivos de los genomas.

En la figura 4.1 se muestran los valores de la complejidad  $C_S$  en función del nivel de umbral,  $D_u$  (perfil de complejidad) para la secuencia HUMTCRADCV y la ECO110k. La secuencia HUMTCRADCV es una secuencia de ADN humano que presenta correlaciones de largo alcance [104], mientras que la ECO110k es una secuencia de ADN bacterial sin este tipo de correlaciones. Se puede observar que la complejidad de la secuencia humana es mayor que la bacterial, al menos en el rango de valores de umbral entre 20 y 50. Un aspecto destacable de la dependencia de  $C_S$  con el nivel de umbral es que llega a un aplanamiento pero no se anula, lo cual es marcadamente distinto a lo que ocurre con otras medidas de complejidad propuestas [102]. En la misma figura se grafica la complejidad  $C_S$  para una secuencia aleatoria (generada por computadora) con la misma composición de bases que la ECO110k. En este caso se observa que la complejidad es nula a partir de cierto valor de umbral, lo que nos lleva a elegir  $D_u$ , tal que  $20 \leq D_u \leq 50$  como rango de interés.

### Super-aditividad

Otro aspecto investigado de la medida de complejidad  $C_S$ , está relacionado a la propiedad de super-aditividad, (4.1). En la figura 4.2 se muestran los perfiles de complejidad para las secuencias completas ECO110k, y la Beta-globulina humana HUMHBB y para la suma pesada de dos subsecuencias arbitrarias de esas dos secuencias. Claramente la ecuación (4.1) se verifica. También es posible comprobar que la complejidad  $C_S$  de cualquier auto-concatenación, es decir, la concatenación de una secuencia consigo misma, es exactamente igual a la de la secuencia original siempre que el punto de unión resulte ser un corte en la segmentación. De no darse esta coincidencia, lo cual es bastante poco probable, ocurre un aumento o una disminución de la complejidad de la auto-concatenación, insignificante para secuencias largas.



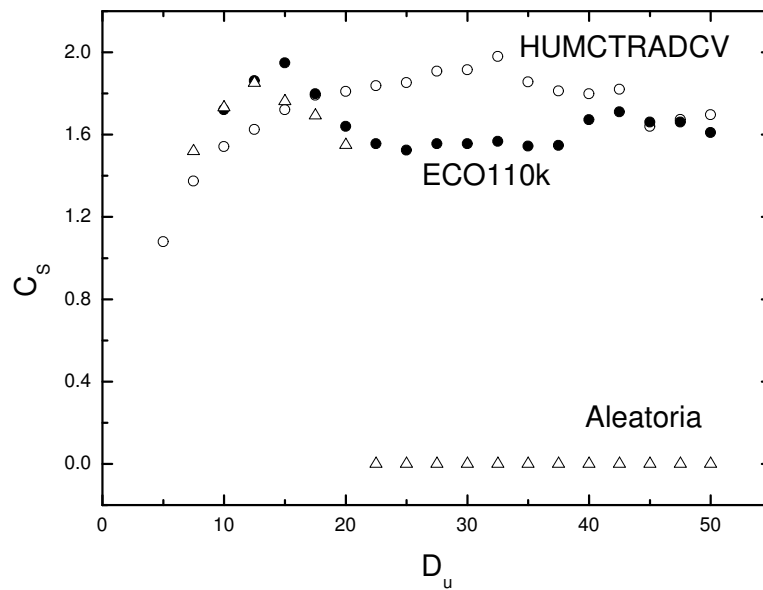


Figura 4.1:  $C_S$  en función del nivel de umbral  $D_u$  para dos secuencias genómicas y una secuencia aleatoria generada artificialmente con idéntica composición que la ECO110k. Los círculos llenos corresponden a la secuencia ECO110k (111408 pares de bases, bp), los círculos vacíos corresponden a las secuencia de ADN humano HUMCTRADCV (98308 bp), y los triángulos corresponden a la secuencia aleatoria. El proceso de segmentación se realizó utilizando el alfabeto normal (alfabeto de 4 símbolos) para describir la secuencia.

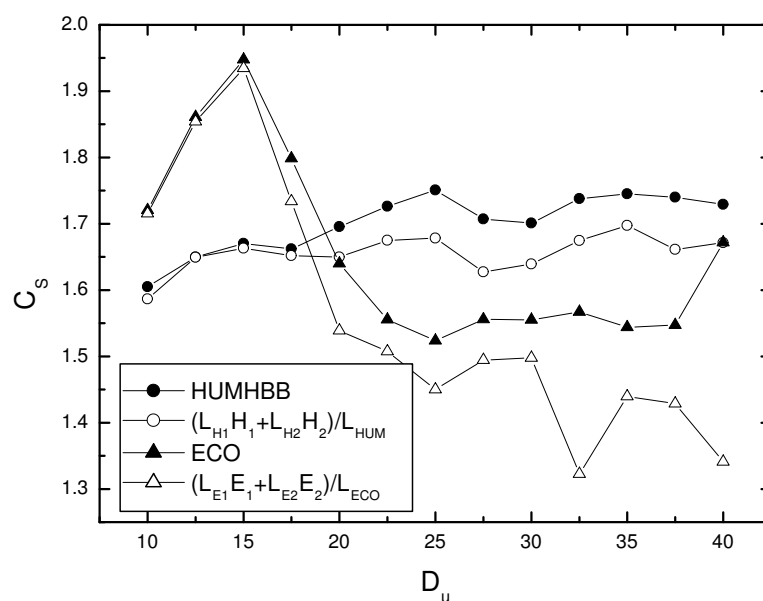


Figura 4.2: Perfiles de complejidad para las secuencias ECO110k ( $L_{ECO} = 111408$  bp) y HUMHBB ( $L_{HUM} = 73308$  bp). Los símbolos llenos corresponden a la complejidad para las secuencias completas, y los símbolos vacíos corresponden a la suma pesada de las complejidades para dos subsecuencias arbitrarias de cada una de las secuencias. Las subsecuencias fueron elegidas de forma tal que su concatenación fuera igual a la secuencia completa ( $L_{E_1} = 57120$  bp y  $L_{E_2} = 54288$  bp;  $L_{H_1} = 42720$  bp y  $L_{H_2} = 30588$ bp).

### Correlaciones

Es sabido que sólo una porción pequeña del genoma de los organismos más altos en la escala biológica, codifica información para la secuencia de aminoácidos de las proteínas [105]. El rol de los intrones (regiones continuas no codificantes en el ADN) permanece aún sin conocerse. El estudio de las propiedades estadísticas de las regiones no codificantes muestra la existencia de correlaciones de largo alcance que indican la presencia de un orden estructural subyacente en los intrones y los segmentos intergénicos. Este orden estructural es revelado por los perfiles de complejidad que se muestran en la figura 4.3, donde hemos graficado la complejidad para las regiones codificantes y no codificantes del cromosoma humano 22.

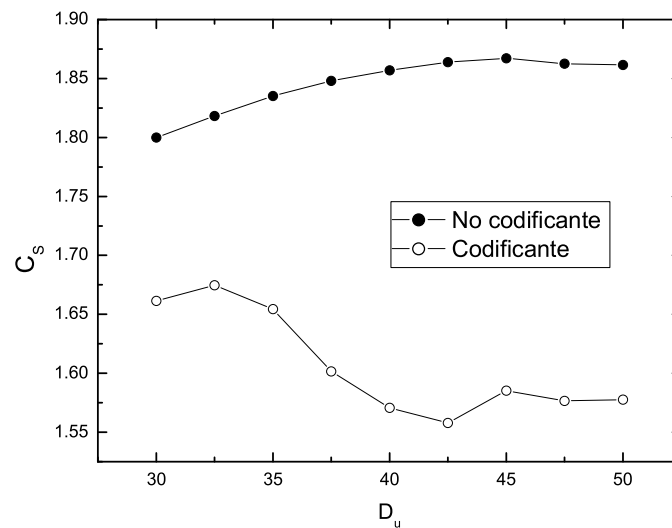


Figura 4.3:  $C_S$  para las regiones codificada y no codificada del cromosoma humano 22.

## Complejidad y evolución

*La evolución es un proceso que ocurre de cualquier modo; es prueba y error en la licuadora de la selección natural, y sobreviven no los más fuertes sino los mejor adaptados. Adaptación -repitió Guzman-, y no fuerza.<sup>2</sup>*

Las secuencias genómicas son una valiosa fuente de información sobre la historia evolutiva de las especies [106]. En particular es posible relacionar algunas características estadísticas observadas a lo largo de las secuencias genómicas a las influencias de una variedad de procesos incluyendo la evolución [107]. En este contexto evaluamos la complejidad  $C_S$ , para secuencias homólogas de ADN de diferentes especies; en particular para la cadena de genes pesados de myosina. En general se puede observar que existe un acuerdo entre la complejidad biológica de las especies y los valores de  $C_S$ . Es importante notar que existe una relación entre el porcentaje de intrones y las correlaciones de largo rango en la secuencia. Este rasgo claramente puesto de manifiesto por  $C_S$  se puede ver en la figura 4.4.

## 4.5. Comentarios Finales

En este capítulo hemos introducido una medida de complejidad basada en la segmentación entrópica de secuencias simbólicas que satisface las nociones intuitivas de complejidad. Además de anularse para secuencias periódicas y aleatorias, la complejidad es máxima para aquellas secuencias que presentan distribuciones de longitudes de segmentos que obedecen una ley de potencia y autocorrelación plana. Estas distribuciones están presentes ubicuamente en la Naturaleza, y hay evidencias de que están relacionadas con procesos con capacidad de auto-organización.

La medida de complejidad introducida es super-aditiva y depende del nivel de detalle con el que se analiza la secuencia. Una propiedad destacable de esta complejidad es que es fácilmente computable. Hasta ahora nos hemos restringido a su aplicación a secuencias genómicas. Un objetivo de esta línea de investigación es estudiar la aplicación de  $C_S$  a sistemas dinámicos. Esto requerirá la definición de un mapa entre la evolución del sistema dinámico bajo estudio y una secuencia simbólica [5].

---

<sup>2</sup>Paola Kaufmann. El Lago.

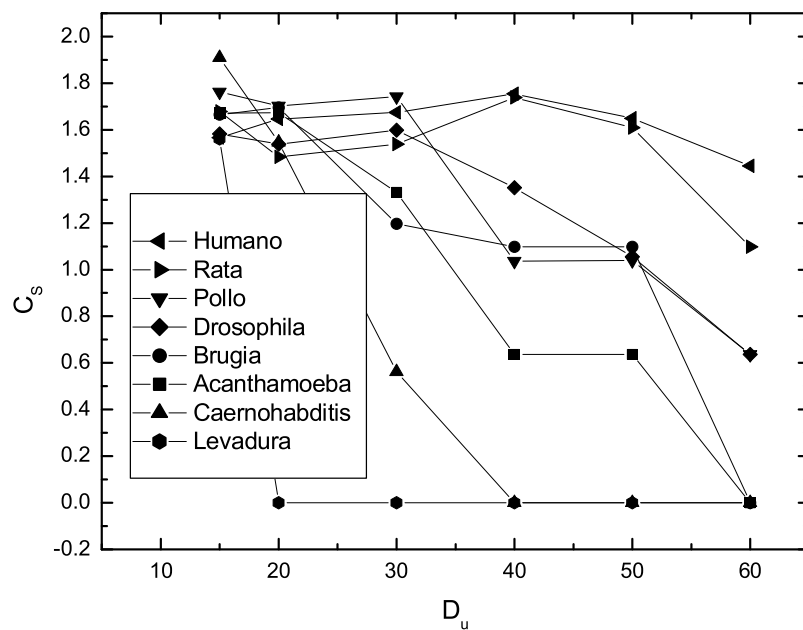


Figura 4.4: Perfiles de complejidad para la cadena de genes pesados de miosina para diferentes especies (longitud total, porcentaje de intrones): Humano (28438bp, 74%), Rata (25759bp, 77%), Pollo (31111bp, 74%), Mosca del vinagre (22663bp, 66%), Brugia (11766bp, 32%), Acanthamoeba (5894bp, 10%), Caenorhabditis (10780bp, 14%), Levadura (6108bp, 0%)



# Capítulo 5

## Distancias y Filogénesis

*Me fascinaba la idea de mi propia impenetrabilidad, de que nadie tuviese acceso al sonido de mi mundo interior. Me maravillaba la cantidad de información, opiniones, percepciones, ideas y proyectos que podía albergar sin que nadie sospechara cuánto se movía tras la estable fachada de mi rostro.*<sup>1</sup>

Desde cierto punto de vista, uno de los rasgos fundamentales que distingue a un ser inanimado de uno vivo, es que este último es portador de información y tiene la capacidad de transmitir esa información a su descendencia, es decir, es portador de un *programa genético* [108]. Esta información está contenida en las secuencias de ADN. El ADN consiste de una doble hélice formada por dos hebras de polinucleótidos arrolladas en sentido antiparalelo, y que se replica (prácticamente) de manera fiel de una generación a otra. La pequeña proporción de cambios (mutaciones) que puedan producirse constituye la base de la evolución, permitiendo que el programa genético evolucione y dé lugar a la enorme variedad de formas de vida que hoy conocemos.

Cada uno de los caracteres elementales de los seres vivos corresponde a un gen, es decir, a un segmento de la secuencia de ADN que especifica una proteína, la cual tiene un rol determinado en el metabolismo o en la estructura celular de esos seres. Esto da lugar a los rasgos característicos del individuo (fenotipo). Otra parte de la cadena de ADN se dedica a regular el funcionamiento de los genes.

Las relaciones filogenéticas (filogenia) son, en general, los patrones de la historia común compartida entre replicadores biológicos. Ejemplos de tales replicadores son las especies y los genes. A un nivel molecular, la filogenética es el estudio de las relaciones evolutivas entre las especies y los genes. En tales estudios hay una mezcla de biología molecular y técnicas estadísticas.

Existen varias razones por las que los datos moleculares, particularmente los contenidos en las secuencias de ADN, son más versátiles para los estudios filogenéticos que los datos morfológicos y fisiológicos. En primer lugar, las secuencias genéticas son más fáciles de

---

<sup>1</sup>Gioconda Belli. El pergamino de la seducción.

modelar que el análisis comparativo de, por ejemplo, huesos y órganos. En segundo lugar, los datos moleculares son directamente generalizables a otras taxonomías; por ejemplo, comparar una planta con una bacteria, haciendo uso de su morfología, puede ser una tarea difícil, mientras que el uso de los datos de las secuencias genéticas, permite hacer tal comparación.

Para sacar provecho de estas ventajas, se han desarrollado diversas técnicas de análisis estadísticos de los datos moleculares. Como se indicó con cierto detalle en la Introducción, muchos de estos desarrollos involucran la definición de distancias entre secuencias simbólicas.

En el presente capítulo nos proponemos encarar el problema del análisis filogenético a la luz de algunos de los desarrollos presentados previamente. Nuestro tratamiento combina, el método de segmentación expuesto en el capítulo anterior, con un modo de medir la separación entre dos secuencias simbólicas. Este último está basado en la distancia introducida por M. Li, (1.14) [26].

## 5.1. Ideas para el análisis filogenético

Una pregunta que surge al realizar el análisis de secuencias de ADN, en el marco de los estudios filogenéticos, es dónde está almacenada la información genética relevante. Una posibilidad es que esta información esté en cada nucleótido, incluida su posición dentro de la secuencia. Otra posible respuesta, y teniendo en cuenta el proceso de segmentación descrito en el capítulo anterior, es que la información esté fundamentalmente dada por la distribución de largos de los segmentos con homogeneidad composicional. De hecho, el que la información esté en ambos aspectos arriba mencionados, es una tercera posibilidad.

Parte de lo que resta del presente capítulo tiene por objetivo esclarecer lo concerniente a esta pregunta. Por lo pronto mencionaremos algunas ideas que permiten explorar la respuesta correcta.

### 5.1.1. Manipulación de las secuencias

Como primer paso en el análisis comparativo de secuencias, vamos a proceder, de acuerdo al método descrito en el capítulo anterior, a la segmentación de la secuencia. Con el fin de explorar la relevancia del contenido de información en la posición de cada nucleótido, dentro de cada uno de los segmentos, uno puede reordenar los símbolos dentro de cada segmento. El reordenamiento de los símbolos destruye la información contenida en la secuencia de nucleótidos y sólo conserva la información de los segmentos.

Se puede optar por uno de los siguientes reordenamientos:

1. Los símbolos dentro de cada segmento se ordenan colocando primero todas las *A*, luego todas las *C*, después las *G* y por último las *T* (sin que el orden de las letras importe, pero una vez elegido se conserva para todas las secuencias que estudiamos).



De esta forma, un segmento

$$ACGTGGTCA,$$

se reescribe como

$$AACCGGGTT.$$

2. Damos posiciones aleatorias a los símbolos dentro de cada segmento, tal que el mismo segmento,

$$ACGTGGTCA$$

toma la forma,

$$GTACGACGT.$$

En el proceso de segmentación, valores pequeños del nivel de umbral llevan a un mayor número de segmentos y es por esto que la información que se destruye en el reordenamiento de símbolos en cada segmento decrece a medida que el nivel de umbral escogido disminuye.

Una vez que las secuencias, correspondientes a cada una de las especies que se están estudiando, han sido segmentadas y sus símbolos reordenados, se puede calcular la distancia entre las secuencias resultantes. Para ello usaremos la medida (1.14). Esto permite calcular una matriz distancia, siendo cada elemento de esta matriz representante de la “distancia” entre las especies.

La estructura de los árboles que se obtienen reordenando los símbolos en las secuencias sólo coincide con la estructura de los existentes en la literatura para valores de umbral muy pequeño. Esto nos lleva a concluir que la posición (el orden) de los símbolos en la secuencia, contiene información relevante para la filogenia.

Es conocido por los especialistas, que el ADN más relevante para los estudios filogenéticos de los vertebrados, es el ADN mitocondrial. Por ello, y tratándose de hacer este capítulo autocontenido, haremos una breve descripción de sus principales características.

### 5.1.2. El ADN mitocondrial

Las mitocondrias son las fábricas de energía de la célula. Son las encargadas de suministrar la mayor parte de la energía que necesita la célula para llevar a cabo sus actividades. Se trata de unos organelos que están en el citoplasma de toda célula eucariota. Curiosamente estos organelos tienen su propio código genético, parecido al del núcleo de la célula. Hay quienes sostienen que la actual célula de los seres pluricelulares es una simbiosis de varias células anteriores: la mitocondria sería una célula autónoma, anterior, que encontró ventajas en vivir dentro de una célula mayor.

Las secuencias de ADN mitocondrial son en general las que más se usan como fuente de información filogenética entre los vertebrados. El ADN mitocondrial se hereda solo de la madre, lo cual permite trazar líneas genéticas directas. Además no se recombina, lo que indica que los cambios se deben a mutaciones a lo largo de muchas generaciones. El proceso de recombinación en el ADN nuclear (con la excepción del cromosoma Y) mezcla secciones de ADN de la madre y del padre, creando así una historia genética mezclada e ilegible.

El ADN mitocondrial es parecido al ADN de las bacterias; es una cadena cerrada, sin extremos por lo cual se puede replicar ilimitadamente. Esto le da carácter de *ente inmortal* y es diferente del ADN cromosómico que se encuentra en el núcleo de las células. Las secuencias de ADN mitocondrial son, en general, más cortas que las de ADN cromosómico, y la mayor parte de estas secuencias codifica, es decir, hay poca cantidad de ADN basura. La figura 5.1 muestra el genoma mitocondrial humano. En general el genoma mitocondrial de todos los mamíferos tiene esta forma.

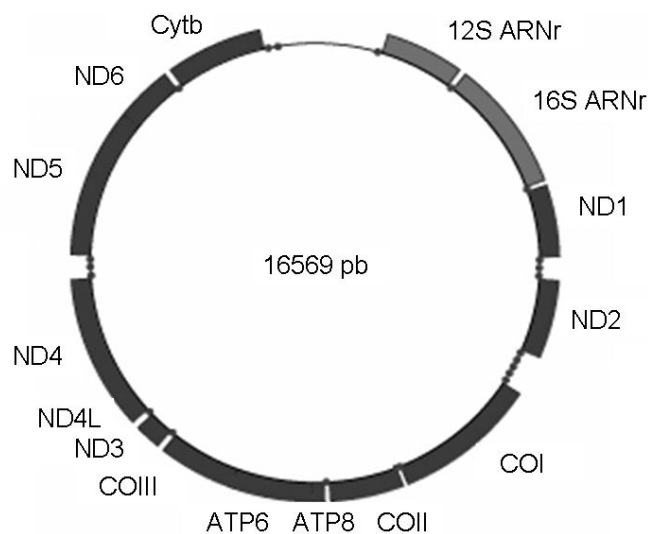


Figura 5.1: ADN mitocondrial (ADNmt) del humano. (Imagen tomada de Brown. Genomes2)

Este genoma es pequeño y compacto, con escasos espacios huecos, tanto que los genes ATP6 y ATP8 se solapan. Abreviaciones: ATP6, ATP8, genes para la subunidad 6 y 8 de ATPasa; COI, COII, COIII: genes para las subunidades I, II y III de la citocromo c oxidasa; Cytb: genes para la apocytochromo b; ND1-ND6: genes para la subunidad 1-6 de la NADH hidrogenasa. ARN ribosómico y ARN de transferencia son dos tipos de ARN no-codificante.

## 5.2. Contrucción de los árboles

... una de cada diez mil especies deja algún registro fósil, que pueda identificarla, que alcanza para armar un relato, una suerte de apuesta. Eso significa que la mayoría de las especies de todos los tiempos se han ido sin dejar más rastro de su presencia sobre la tierra que el de un cangrejo sobre la arena.<sup>2</sup>

### 5.2.1. Esquema de cálculo

A partir de la medida de distancia (1.14), y manipulando las secuencias mitocondriales completas de un grupo de mamíferos estudiaremos en qué parte de la secuencia se encuentra la información relevante para la construcción de relaciones filogenéticas.

Utilizaremos la distancia  $d_K$  para contruir la matriz distancia, cuyos elementos son las distancias entre todos los pares de las secuencia bajo estudio, y a partir de esta matriz, utilizando el paquete PHYLIP [109], construiremos el árbol filogenético correspondiente.

PHYLIP, el *paquete de inferencia filogenética*, consiste de 35 programas, que permiten inferir filogenia (árboles evolutivos). Esta inferencia se puede hacer por parsimonia, compatibilidad, métodos de matrices distancia y máxima verosimilitud. Con PHYLIP también se puede calcular el consenso y la distancia entre árboles y calcular matrices de distancia de distintas maneras. El paquete trabaja con diferentes tipos de datos. Estos pueden ser secuencias de nucleótidos, secuencias de proteínas, frecuencias de genes, fragmentos o partes de una secuencia, distancias y caracteres discretos o continuos en general.

De los programas del paquete, utilizamos el programa NEIGHBOR que construye árboles por agrupamiento sucesivo de linajes; y el programa TREEDIST que calcula la distancia entre árboles. Ésta se calcula considerando todas las ramificaciones que pueden existir entre los dos árboles que se están comparando.

Como la complejidad Kolmogorov, en general, no se puede calcular, usamos el programa GenCompress [110] para aproximar las cantidades  $K(\mathcal{S}_1)$ ,  $K(\mathcal{S}_1|\mathcal{S}_2)$  y  $K(\mathcal{S}_1\mathcal{S}_2)$ . Estas se necesitan para evaluar la distancia  $d_K$  entre cada una de las secuencias.

### 5.2.2. Motivación

La relación filogenética entre primates, ferungulados<sup>3</sup> y roedores ha sido, hasta hace poco tiempo, motivo de debate. Notablemente se ha observado que distintos grupos de proteínas, llevan a la construcción de diferentes árboles, relativizando en consecuencia los resultados emergentes del análisis filogenético. El origen de estas incongruencias ha sido estudiado en detalle por Y. Cao *et al.* en 1998. En este trabajo estos investigadores compararon el árbol que se obtiene de las 12 proteínas codificadas por la hebra H del ADNmt, concatenadas,

---

<sup>2</sup>Paola Kaufmann. El Lago.

<sup>3</sup>Ferungulados: artiodáctilos + cetáceos + perisodáctilos + carnívoros:

y el que se obtiene del gen individual NDI que codifica para la NADH hidrogenasa. En el trabajo se estudiaron también árboles generados por otros genes individuales.

Las conclusiones más importantes de ese trabajo, se pueden resumir del siguiente modo. Dado que los primates, roedores y ferungulados pueden estar relacionados de diferentes maneras, se pueden generar tres tipos de árboles, como se muestra en la figura 5.2. Mientras que las 12 proteínas concatenadas soportan el árbol 1, los estudios realizados con el gen NDI, soportan el árbol 2, y el resultado no se modifica agregando especies al análisis. A su vez, Cao *et al.* secuenciaron los genes de especies adicionales y las agregaron al estudio filogenético para de esta forma disminuir el ruido estocástico que podría haber sido responsable de las diferencias. Además de la secuenciación de genes de especies adicionales, se realizaron estudios evolutivos más sofisticados. Sin embargo ninguno de estos procesos cambió la conclusión previa y el mayor número de especies contrario a lo esperado, reforzó los resultados contradictorios obtenidos para el gen NDI.

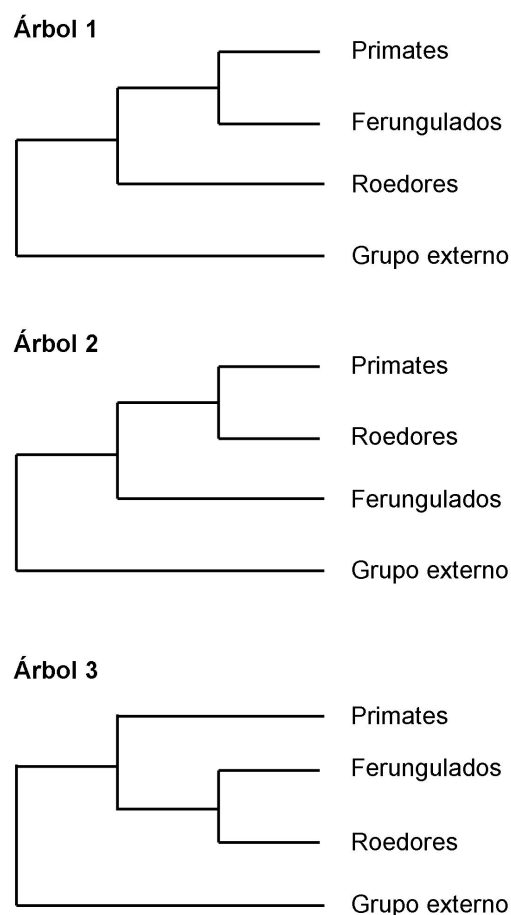


Figura 5.2: Los tres árboles posibles entre primates, ferungulados y roedores.

Este grupo de mamíferos también fue estudiado por Li *et al.* utilizando los genomas mitocondriales completos [26]. Los estudios realizados utilizando los genomas completos o las 12 proteínas codificadas por la secuencias mitocondriales, permitieron establecer la ramificación temprana de los roedores, conduciendo a la conclusión de que el análisis del gen ND1 lleva a un árbol filogénico incorrecto. En nuestro trabajo comparamos los árboles obtenidos con los árboles generados por M. Li [26] y que concuerdan con los últimos resultados obtenidos por Y. Cao y colaboradores [111].

### 5.2.3. Materiales y métodos

#### Análisis de 20 secuencia mitocondriales

Las especies con las que trabajamos fueron las siguientes: rata (*Ratus norvegivus*), ratón (*Mus musculus*), foca gris (*Halichoerus grypus*), foca (*Phoca vitulina*), gato (*Felis catus*), rinoceronte blanco (*Caratherium simum*), caballo (*Equus caballus*), ballena (*Balaenoptera physalus*), ballena azul (*Balaenoptera musculus*), vaca (*Bos taurus*), guibon (*Hylobates lar*), gorila (*Gorilla gorilla*), humano (*Homo sapiens*), chimpancé (*Pan troglodytes*), chimpancé pigmeo (*Pan paniscus*), orangután de Bornea (*Pongo pygmaeus*), orangután de Sumatra (*Pongo pigmaeus abelii*), comadreja (*Didelphis virginiana*), canguro (*Macropus robustus*) y ornitorrinco (*Ornithorhynchus anatinus*). Las tres últimas las utilizamos como grupo externo. Todas la secuencias se bajaron del banco de genes.

#### Algoritmo de compresión

Gencompress [110] es un algoritmo de compresión sin pérdida, diseñado especialmente para secuencias genéticas, que se basa en la búsqueda de repeticiones aproximadas. El algoritmo se basa en la idea de que la secuencias de ADN, secuencias complejas, es decir no aleatorias, son compresibles y toma en cuenta que 2 bits son suficientes para almacenar cada una de las 4 bases del alfabeto ( $\{A, C, G, T\}$ ). Es por esto que este programa especial lleva a tasas de compresión más altas que los programas de compresión de textos estándares, que expanden los archivos de las secuencias, asignando más de 2 bits a cada base.

#### Construcción de árboles

Para construir los árboles utilizamos varios de los algoritmos que con este fin incluye el paquete PHYLIP [109]. Los resultados que se presentan a continuación corresponden al programa *Neighbor Joining* (NJ) [112] en el paquete mencionado. Elegimos este algoritmo con el fin de hacer una comparación consistente con los resultados presentados por otros autores [26, 111].

#### Genes o intergenes

Con el fin de estudiar qué parte de la secuencia codifica la información relevante desde el punto de vista filogenético, vamos a pensar, en un modelo simplificado de una secuencia de

ADN constituida por genes (sólo consideramos los que codifican para proteínas) e intergenes (segmentos que están entre los genes). Recordemos que los genes son segmentos dentro de la secuencia que tienen alguna función, y que los genes que codifican proteínas son aquellos que tienen la información respecto a los rasgos característicos de un individuo. Otros genes codifican para ADN ribosómico, o de transferencia. Las proteínas tienen un rol determinado en el metabolismo o en la estructura celular, dando lugar así a algún rasgo del aspecto del individuo, lo que ya mencionamos como el fenotipo.

A diferencia de otros trabajos [26, 111], trabajaremos con las secuencias genómicas completas, pero destruyendo la información contenida en los genes y luego en los intergenes, dejando cada una de las partes complementarias de la secuencia (intergenes, genes) según el caso sin alterar. Para destruir la información, en los genes o intergenes, asignaremos posiciones aleatorias a los símbolos que pertenezcan a estos segmentos. Dando a los símbolos posiciones aleatorias, destruimos los codones, y por tanto, los aminoácidos que los codones codifican para dar origen a las proteínas, según corresponda.

Con estas secuencias, completas pero con las modificaciones recién mencionadas, calculamos la matriz distancia para las 20 secuencias en estudio. A partir de esta matriz, construimos el árbol correspondiente a cada caso.

Las matrices distancias obtenidas en cada caso difieren de la matriz que se obtiene usando los genomas completos. Generamos los árboles correspondientes y vemos que el árbol que resulta dejando los genes intactos, tiene la misma estructura que el árbol que se obtiene usando los genomas completos. Esto no ocurre con el árbol que se obtiene dejando intactas las regiones correspondientes a intergenes. A continuación se muestran los árboles que se obtienen con los genomas mitocondriales completos y con los genes o intergenes desordenados, para las 20 especies con las que hemos trabajado.

Finalmente calculamos las distancias entre los diferentes árboles. La matriz obtenida se muestra en la figura 5.6. Se ve que, si bien la distancia entre el árbol original y el árbol que se obtiene de los genomas con los genes es menor, hay una diferencia entre estos árboles.

### 5.3. Comentarios Finales

En este capítulo hemos aplicado una de las distancias genéticas comentadas en la Introducción combinada con el método de segmentación basado en la DJS. Hemos mostrado cómo se trabaja con las medidas de distancia genética y con la construcción de filogénesis utilizando software disponible. De los resultados del trabajo podemos concluir que el orden de los símbolos dentro de la secuencia es relevante para estudios de filogénesis, en particular la información que contienen los genes que codifican para proteínas.

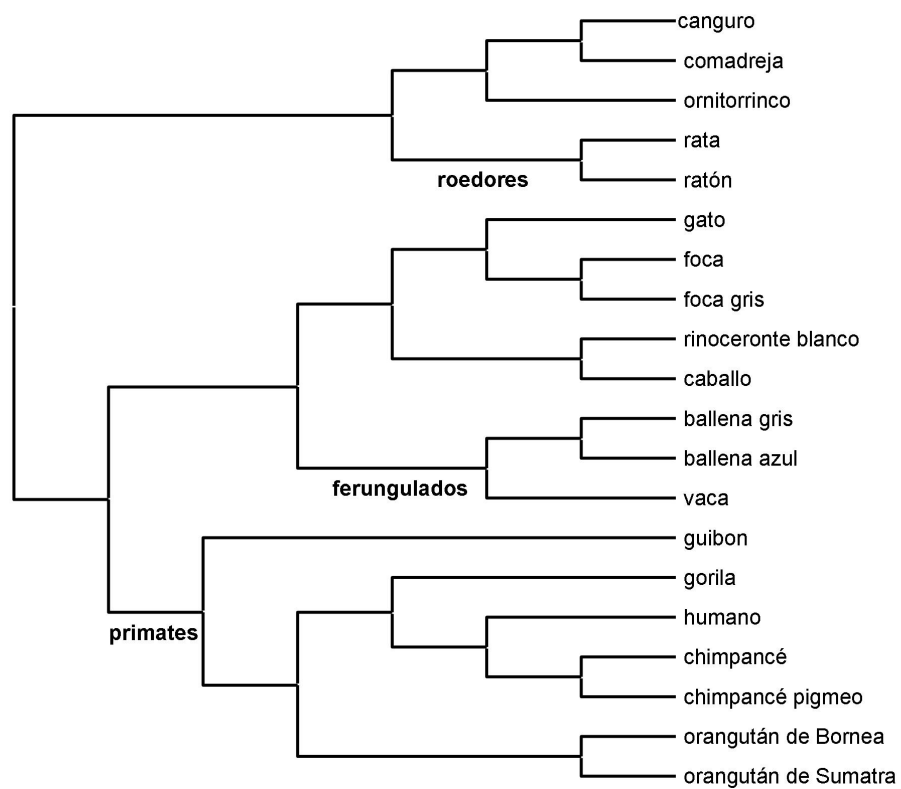


Figura 5.3: Árbol con los genomas mitocondriales completos.

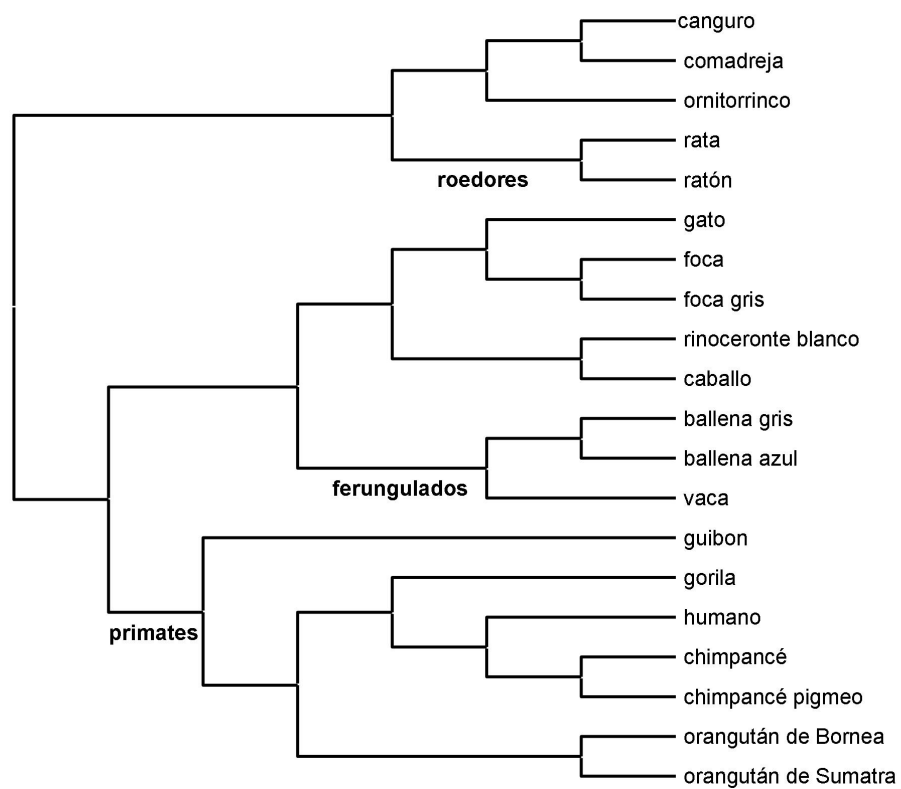


Figura 5.4: Árbol con los genes sin modificar y los intergenes desordenados.



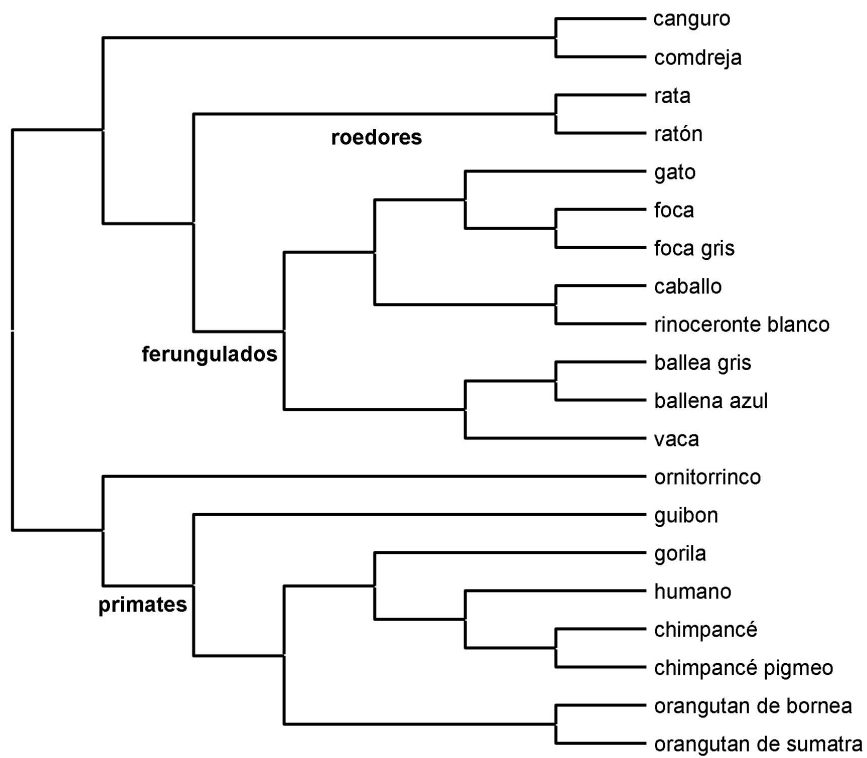


Figura 5.5: Árbol con los símbolos correspondientes a los genes desordenados, se puede observar que la información filogenética ha sido destruida.

árboles	completo	gen	intergen
completo	0	0.5358	0.6558
gen	0.5358	0	0.3445
intergen	0.6558	0.3445	0

Figura 5.6: Matriz distancia entre los diferentes árboles.



## Capítulo 6

# La Divergencia de Jensen-Shannon Cuántica

*– Dije cerdito –contestó Alicia–. ¡Y a ver si dejas de andar apareciendo y desapareciendo tan de golpe! ¡Me da mareo!*

*–De acuerdo –dijo el Gato.*

*Y esta vez desapareció despacito, con mucha suavidad, empezando por la punta de la cola y terminando por la sonrisa, que permaneció un rato allí, cuando el resto del Gato ya había desaparecido.*

*–¡Vaya! –se dijo Alicia–. He visto muchísimas veces un gato sin sonrisa, ¡pero una sonrisa sin gato! ¡Es la cosa más rara que he visto en toda mi vida!<sup>1</sup>*

Como se reseñó en la Introducción, el concepto de distancia entre estados juega un papel fundamental en distintos contextos de la teoría cuántica. Dada la naturaleza estadística de esta teoría física, la distancia entre estados termina teniendo una contrapartida como distancia entre distribuciones de probabilidad. Esta afirmación se formalizará más adelante. Por ahora, sólo queremos contextualizar el problema de definir distancias entre estados cuánticos, con el resto del presente trabajo de tesis.

Aún cuando, y con diferentes motivaciones, se han propuesto varias medidas de distancias entre estados cuánticos, muchas de estas cantidades tienen diferentes problemas; algunos de tipo formal y otros de tipo operacional.

En este capítulo introduciremos una nueva distancia entre estados cuánticos, la cual tiene notables ventajas formales y de cálculo, frente a otras anteriormente propuestas. Nuestra definición se hará en dos partes: la primera estará dedicada a la definición de la distancia entre estados puros, y en la segunda lo haremos para estados mixtos.

Antes de estudiar a la DJS cuántica, repasaremos algunos aspectos formales básicos

---

<sup>1</sup>Lewis Carroll. Alicia en el País de las Maravillas.

de la Mecánica Cuántica, en particular aquellos asociados con el proceso de medición. También revisaremos algunas de las medidas de distancias que se utilizan comúnmente en Física Cuántica y que pueden relacionarse con nuestra propuesta.

## 6.1. Estados y operaciones en Física Cuántica

El estado de un sistema físico colecta la información disponible sobre el sistema y que puede obtenerse a través de mediciones.

Matemáticamente, un estado cuántico es un elemento de un espacio de Hilbert complejo separable,  $\mathcal{H}$ . El producto escalar de  $\mathcal{H}$ ,  $\langle \Phi | \Psi \rangle$   $\Phi, \Psi \in \mathcal{H}$ , induce la norma  $\|\Psi\|^2 = \langle \Psi | \Psi \rangle$ .<sup>2</sup>

Si se tiene máxima información sobre el estado de un sistema cuántico, en el sentido de que el sistema ha sido preparado fijando los valores de un sistema completo de observables, se dice que el sistema está en un *estado puro*; en tal caso está representado por un vector de estado normalizado  $|\Psi\rangle$  en el espacio de Hilbert,  $\|\Psi\| = 1$ . Este vector de estado está unívocamente determinado excepto por una fase, y el estado en sí mismo corresponde a un rayo unidad en  $\mathcal{H}$ , es decir al conjunto  $\{e^{i\lambda}|\Psi\rangle \mid \lambda \in \mathfrak{R}\}$ .

Un sistema cuántico muy elemental es uno que tiene dos estados posibles. En este caso el espacio de Hilbert asociado es uno de dimensión dos. Los elementos de la base del espacio de vectores se pueden identificar como  $|0\rangle$  y  $|1\rangle$ , y cualquier vector (estado) en este espacio de Hilbert - espacio complejo ( $C^2$ )- se puede expresar en la forma

$$|\Psi\rangle = \alpha|0\rangle + \beta|1\rangle, \quad (6.1)$$

donde  $\alpha$  y  $\beta$  son números complejos que satisfacen la condición  $|\alpha|^2 + |\beta|^2 = 1$ .

Si se quiere incluir la posibilidad de información parcial sobre el estado de un sistema, los rayos unidad no son suficientes para describirlo. El concepto de *estado mezcla* incorpora ignorancia sobre un estado cuántico. Desde el punto de vista del formalismo, un estado mezcla será representado por un operador  $\rho$  definido sobre  $\mathcal{H}$  que satisface:

1.  $\rho$  es autoadjunto,  $\rho = \rho^\dagger$ .
2.  $\rho$  es un operador definido semipositivo,  $\rho \geq 0$ .
3.  $\rho$  tiene traza unidad;  $Tr\rho = 1$ .

La representación de un estado por medio de un operador con las anteriores propiedades, incluye tanto a los estados mezclas como a los estados puros; en este último caso el operador  $\rho$  es idempotente:  $\rho^2 = \rho$ .

Cualquier estado mezcla admite una representación de la forma

---

<sup>2</sup>A lo largo del presente capítulo, usaremos indistintamente  $\Psi$  o la notación de brackets  $|\Psi\rangle$ , para representar a un elemento del espacio  $\mathcal{H}$ .

$$\rho = \sum_{i=1}^n p_i |\phi_i\rangle\langle\phi_i| \quad (6.2)$$

con  $\{p_i\}_{i=1}^n$  números reales no negativos tales que  $\sum_i p_i = 1$  y las proyecciones  $|\phi_i\rangle\langle\phi_i|$ ,  $i = 1, \dots, n$ , corresponden a un conjunto completo (normalizado) de autoestados. En general, un estado tiene un número infinito de tales representaciones, y tanto los pesos ( $p_i$ ), como las proyecciones, pueden ser diferentes en las distintas descomposiciones.

La ecuación de Schrödinger gobierna la dinámica de los estados de un sistema cuántico aislado. La evolución temporal corresponde al *mapa dinámico*

$$\rho \mapsto \sigma = U\rho U^\dagger, \quad (6.3)$$

donde  $U : \mathcal{H} \rightarrow \mathcal{H}$  es un operador unitario, en general, dependiente del tiempo.

## Mediciones

Muchos de los resultados más sorprendentes de la Mecánica Cuántica están relacionados con las propiedades particulares de los procesos de medición [113]. Desde un punto de vista formal, una medición en Mecánica Cuántica se describe por medio de una colección de operadores,  $\{M_m\}$ , que actúan sobre el espacio de Hilbert  $\mathcal{H}$ ; el índice  $m$  etiqueta los posibles resultados de la medición. Si el sistema está en el estado  $|\Psi\rangle$  inmediatamente antes de la medición, la probabilidad de que el resultado de la medición sea  $m$  está dada por

$$p_m = \langle\Psi|M_m^\dagger M_m|\Psi\rangle,$$

y el estado del sistema después de la medición es

$$\frac{M_m|\Psi\rangle}{\sqrt{\langle\Psi|M_m^\dagger M_m|\Psi\rangle}}$$

Un caso especial de los procesos de medición lo dan las *mediciones proyectivas*. Una medición proyectiva se describe por un observable,  $\mathcal{M}$ , es decir un operador Hermitiano definido sobre el espacio de los estados del sistema:

$$\mathcal{M} = \sum_m m P_m,$$

donde  $P_m$  es el proyector sobre el autoespacio de  $\mathcal{M}$  con autovalor  $m$ . Si el estado del sistema es  $\Psi$ , la probabilidad de medir  $m$  está dada por

$$p_m = \langle\Psi|P_m|\Psi\rangle.$$

Supongamos que  $\{M_m\}$  es una familia completa de proyectores ortogonales,

$$\begin{aligned} M_m M_{m'} &= \delta_{m,m'} M_m \\ \sum_m M_m M_m^\dagger &= I, \end{aligned}$$

y que describen una medición. Si el sistema está en el estado  $\Psi$ , la probabilidad del resultado  $m$ , está dada por  $p_m = \langle \Psi | M_m^\dagger M_m | \Psi \rangle$ . Si introducimos los operadores

$$E_m = M_m^\dagger M_m,$$

se tiene que  $\sum_m E_m = I$  y  $p_m = \langle \Psi | E_m | \Psi \rangle$ . Entonces, el conjunto de operadores  $E_m$  es suficiente para determinar las probabilidades de los diferentes resultados de la medición. El conjunto completo  $E_m$  se conoce como un POVM (Positive Operator-Value Measure).

Más adelante volveremos a estos aspectos formales del proceso de medición y lo relacionaremos con sus importantes implicancias en el marco de la información y la computación cuántica [114].

Un tema íntimamente relacionado con el de la medición, es el de la distinguibilidad entre estados, es decir, con la manera en que se puede discernir entre dos estados vecinos  $|\Psi^{(1)}\rangle$  y  $|\Psi^{(2)}\rangle$ , por medio de un dispositivo de medición determinado. Es de notar que el problema de distinguibilidad entre estados de un sistema físico tiene su relevancia en el contexto clásico; pero la naturaleza del problema en MC es altamente no trivial [28].

En las secciones siguientes introduciremos una medida de distinguibilidad entre estados cuánticos a partir de la divergencia de Jensen-Shannon. En primer lugar trataremos el problema para estados puros y luego para estados mezclas.

## 6.2. Parte I: Estados puros

### 6.2.1. Introducción

Consideremos los autoestados  $|\phi_1\rangle, \dots, |\phi_N\rangle$ , de un operador hermitiano  $\mathcal{A}$  asociado con un instrumento de medición; por simplicidad supondremos que no hay degeneración. Como sabemos, tras el proceso de medición se obtendrá alguno de los  $N$  resultados posibles,  $a_i$ . Si el sistema ha sido preparado en el estado normalizado  $|\Psi^{(1)}\rangle$ , la probabilidad de ocurrencia del valor  $a_i$  está dada por  $p_i^{(1)} = |\langle \phi_i | \Psi^{(1)} \rangle|^2$ . Si en cambio el sistema ha sido preparado en el estado  $|\Psi^{(2)}\rangle$ , la probabilidad de obtener el resultado  $a_i$  será:  $p_i^{(2)} = |\langle \phi_i | \Psi^{(2)} \rangle|^2$ . Como los  $|\phi_i\rangle$  constituyen una base completa,

$$\sum_i |\phi_i\rangle \langle \phi_i| = I, \quad (6.4)$$

y por lo tanto

$$\sum_i p_i^{(1)} = \sum_i p_i^{(2)} = 1. \quad (6.5)$$

Como antes denotemos al conjunto de las distribuciones de probabilidad discretas  $M_+^1(N)$

3.

---

<sup>3</sup>La generalización de estos desarrollos al caso continuo es directa.

$$M_+^1(N) = \{(p_1, \dots, p_N); 0 \leq p_i \leq 1; \sum_i p_i = 1\} \quad (6.6)$$

Sea  $\mathcal{S}$  el conjunto de estados normalizados en el espacio de Hilbert  $\mathcal{H}^{n+1}$ ,  $n + 1 = N$ . A cada estado  $|\Psi\rangle$  en  $\mathcal{S}$  se le asigna un elemento  $\{p_i\}$  de  $M_+^1(N)$  a través de la aplicación  $\mathcal{F}_{\mathcal{A}}$

$$\begin{aligned} \mathcal{F}_{\mathcal{A}} : \mathcal{S} \subset \mathcal{H}^{n+1} &\rightarrow M_+^1(N) \\ |\Psi\rangle &\rightarrow \{p_i\} \end{aligned} \quad (6.7)$$

donde  $p_i = |\langle \phi_i | \Psi \rangle|^2$ . Obviamente, la aplicación  $\mathcal{F}_{\mathcal{A}}$  es consistente con las normalizaciones (6.5) y (6.6).

Vamos a denotar con  $d_M(P^{(1)}, P^{(2)})$  a una distancia definida sobre el espacio de las distribuciones de probabilidad  $M_+^1(N)$ , es decir una aplicación que va de  $M_+^1(N) \times M_+^1(N)$  en  $\mathfrak{R}$  y que satisface las propiedades (1.2). A través de la aplicación  $\mathcal{F}_{\mathcal{A}}$  se puede asociar a  $d_M$  una distancia definida en el espacio  $\mathcal{H}^{n+1}$ ,

$$d_{\mathcal{H}}^{\mathcal{A}}(|\Psi^{(1)}\rangle, |\Psi^{(2)}\rangle).$$

Notemos que la distancia  $d_{\mathcal{H}}^{\mathcal{A}}(|\Psi^{(1)}\rangle, |\Psi^{(2)}\rangle)$  definida sobre  $\mathcal{H}^{n+1}$  depende del dispositivo de medición  $\mathcal{A}$ .

Nuestro objetivo es encontrar una distancia representativa de  $d_M(P^{(1)}, P^{(2)})$  en el espacio de Hilbert que sea independiente de la base  $|\phi_k\rangle$  y para esto buscaremos el máximo de esta distancia  $d_{\mathcal{H}}^{\mathcal{A}}$  sobre todos los instrumentos de medición  $\mathcal{A}$ . Encontrando este máximo, hallaremos la mejor discriminación posible entre dos estados <sup>4</sup>.

A continuación aplicaremos este esquema de correspondencia a distintas distancias introducidas por otros autores.

### 6.2.2. Distancia Wootters

En 1981, en un trabajo de gran importancia, W. Wootters [34] usó la idea de distinguibilidad entre estados, para definir una medida de distancia entre preparaciones cuánticas, llamada distancia estadística. Es importante notar que la noción de distinguibilidad introducida por Wootters es aplicable en general, más allá de la MC propiamente dicha.

Siguiendo a Wootters, diremos que dos distribuciones de probabilidad son indistinguibles en un dado número de experimentos, si la diferencia entre ellas es más pequeña que el tamaño de una fluctuación estadística típica. De modo más preciso, diremos que las distribuciones de probabilidad  $P^{(1)} = (p_1^{(1)}, p_2^{(1)}, \dots, p_N^{(1)})$  y  $P^{(2)} = (p_1^{(2)}, p_2^{(2)}, \dots, p_N^{(2)})$  son indistinguibles tras  $L$  pruebas ( $L \rightarrow \infty$ ) si y sólo si la condición

$$\frac{\sqrt{L}}{2} \left\{ \sum_{i=1}^N \frac{(\delta p_i)^2}{p_i} \right\}^{1/2} < 1 \quad (6.8)$$

<sup>4</sup>Notación: Denotaremos con  $d_M$  a una distancia definida sobre  $M_+^1(N)$ ; con  $d_{\mathcal{H}}^{\mathcal{A}}$  a la correspondiente distancia sobre  $\mathcal{H}^{n+1}$  que se obtiene por la correspondencia  $\mathcal{F}_{\mathcal{A}}$ ; y con  $D_{\mathcal{H}}$  al máximo de  $d_{\mathcal{H}}^{\mathcal{A}}$  sobre todos los aparatos  $\mathcal{A}$ .

con  $\delta p_i = p_i^{(1)} - p_i^{(2)}$ ,

Obviamente, diremos que las distribuciones  $P^{(1)}$  y  $P^{(2)}$  son distinguibles en  $L$  experimentos si se cumple la desigualdad:

$$\frac{\sqrt{L}}{2} \left\{ \sum_{i=1}^N \frac{(\delta p_i)^2}{p_i} \right\}^{1/2} \geq 1 \quad (6.9)$$

Este criterio de distinguibilidad involucra una distancia entre dos distribuciones de probabilidad, la cual en el límite  $L \rightarrow \infty$ , toma la forma:

$$d_M^W(P^{(1)}, P^{(2)}) = \lim_{L \rightarrow \infty} \frac{1}{\sqrt{L}} \times Y_L, \quad (6.10)$$

donde  $Y_L$  es el número de distribuciones de probabilidad intermedias, cada una de las cuales es distinguible (en  $L$  experimentos) de sus vecinas. Tras un poco de álgebra (6.10) se puede reescribir como [34]

$$d_M^W(P^{(1)}, P^{(2)}) = \arccos \left[ \sum_i \left( p_i^{(1)} \right)^{1/2} \left( p_i^{(2)} \right)^{1/2} \right]. \quad (6.11)$$

En el contexto cuántico, la distancia de Wootters corresponde al número de estados distinguibles entre dos estados dados, cuando se analizan con el mismo dispositivo de medición. Así, la distancia estadística está completamente determinada por el tamaño de las fluctuaciones que ocurren en las mediciones que se realizan para distinguir un estado de otro. Si tenemos en cuenta la aplicación (6.7), a la distancia (6.11) le corresponde una distancia entre estados cuánticos, digamos entre los estados  $\Psi^{(1)}$  y  $\Psi^{(2)}$ :

$$d_{\mathcal{H}}^{W, \mathcal{A}}(|\Psi^{(1)}\rangle, |\Psi^{(2)}\rangle) = \arccos \left( \sum_i |\langle \phi_i | \Psi^{(1)} \rangle| |\langle \phi_i | \Psi^{(2)} \rangle| \right). \quad (6.12)$$

Naturalmente esta distancia depende del aparato de medición,  $\mathcal{A}$ , que se utiliza para distinguir un estado del otro. Surge entonces, la siguiente pregunta: Cuál es la mejor discriminación que podemos hacer entre dos estados? Podemos responder a ella, buscando el máximo de la distancia (6.12) sobre todos los aparatos de medición:

$$\max_{\mathcal{A}} d_{\mathcal{H}}^{W, \mathcal{A}}(|\Psi^{(1)}\rangle, |\Psi^{(2)}\rangle) \quad (6.13)$$

Fácilmente podemos calcular este máximo, teniendo presente que la desigualdad

$$\sum_i |\langle \phi_i | \Psi^{(1)} \rangle| |\langle \phi_i | \Psi^{(2)} \rangle| \geq |\langle \Psi^{(1)} | \Psi^{(2)} \rangle|, \quad (6.14)$$

vale para todo conjunto ortonormal  $\{|\phi_i\rangle\}$ . Suponiendo las descomposiciones

$$\begin{aligned} |\Psi^{(1)}\rangle &= \sum_k a_k |\phi_k\rangle \\ |\Psi^{(2)}\rangle &= \sum_k b_k |\phi_k\rangle, \end{aligned} \quad (6.15)$$



se tiene

$$\begin{aligned} |\langle \Psi^{(1)} | \Psi^{(2)} \rangle| &= \left| \sum_k a_k b_k^* \right| \leq \sum_k |a_k b_k^*| \\ &\leq \sum_k |\langle \phi_k | \Psi^{(1)} \rangle| |\langle \phi_k | \Psi^{(2)} \rangle|. \end{aligned} \quad (6.16)$$

De la desigualdad (6.16), junto a la naturaleza decreciente de la función arc cos en el intervalo  $[0, 1]$ , resulta que la distancia

$$D_{\mathcal{H}}^W(|\Psi^{(1)}\rangle, |\Psi^{(2)}\rangle) = \arccos(|\langle \Psi^{(1)} | \Psi^{(2)} \rangle|), \quad (6.17)$$

maximiza  $d_{\mathcal{H}}^{W,A}$ . Este es uno de los resultados más importante del trabajo de Wootters del año 1981: geoméricamente, la distancia (6.17) representa el ángulo entre los dos estados (rayos)  $|\Psi^{(1)}\rangle$  y  $|\Psi^{(2)}\rangle$ .

Para concluir este repaso de la distancia de Wootters, mencionemos que ella ha sido extendida a estados mezclas en un trabajo de Braunstein y Caves [36]. Volveremos a este punto en la segunda parte de este capítulo.

### 6.2.3. Distancia Hellinger

Otra medida de distancia definida sobre el espacio de las distribuciones de probabilidad es la dada por la expresión:

$$(d_M^H)^2(P^{(1)}, P^{(2)}) = \frac{1}{2} \sum_i \left| \sqrt{p_i^{(1)}} - \sqrt{p_i^{(2)}} \right|^2, \quad (6.18)$$

Utilizando el mismo esquema que el aplicado para la distancia Wootters, se llega a que la contrapartida de la distancia  $d_M^H$  en el espacio de Hilbert, es:

$$(d_{\mathcal{H}}^{H,A})^2 = \frac{1}{2} \sum_i \{ |\langle \phi_i | \Psi^{(1)} \rangle| - |\langle \phi_i | \Psi^{(2)} \rangle| \}^2, \quad (6.19)$$

o equivalentemente:

$$1 - \sum_i |\langle \phi_i | \Psi^{(1)} \rangle| |\langle \phi_i | \Psi^{(2)} \rangle|. \quad (6.20)$$

Nuevamente, para encontrar el máximo de  $d_{\mathcal{H}}^{H,A}$  podemos usar la desigualdad (6.16), llegando de este modo a:

$$(D_{\mathcal{H}}^H)^2(|\Psi^{(1)}\rangle, |\Psi^{(2)}\rangle) = 1 - |\langle \Psi^{(1)} | \Psi^{(2)} \rangle|, \quad (6.21)$$

Esta distancia se conoce como la distancia Hellinger y geoméricamente se puede interpretar como el ángulo medio entre los vectores  $|\Psi^{(1)}\rangle$  y  $|\Psi^{(2)}\rangle$  en el espacio de Hilbert [115].

### 6.2.4. Distancia Bhattacharyya

Otra distancia, utilizada tanto sobre el conjunto de distribuciones de probabilidad, como en el espacio de Hilbert de la MC, es la que se origina de los coeficientes Bhattacharyya. Para dos distribuciones de probabilidad  $P^{(1)}$  y  $P^{(2)}$ , estos coeficientes se definen como [116]:

$$B(P^{(1)}, P^{(2)}) = \sum_i \sqrt{p_i^{(1)}} \sqrt{p_i^{(2)}} \quad (6.22)$$

A partir de estos coeficientes se puede definir una distancia entre las distribuciones de probabilidad:

$$d_M^B(P^{(1)}, P^{(2)}) = -\ln(B(P^{(1)}, P^{(2)})). \quad (6.23)$$

La distancia asociada en el espacio de Hilbert es

$$d_{\mathcal{H}}^{B,A} = -\ln \sum_i |\langle \phi_i | \Psi^{(1)} \rangle| |\langle \phi_i | \Psi^{(2)} \rangle|. \quad (6.24)$$

Notemos que la distancia Wootters también se puede expresar en término de los coeficientes  $B(P^{(1)}, P^{(2)})$  como

$$d_M^W(P^{(1)}, P^{(2)}) = \arccos(B(P^{(1)}, P^{(2)})).$$

Es importante mencionar que ni la distancia Wootters ni la distancia Bhattacharyya son métricas puesto que no verifican la desigualdad triangular.

La distancia asociada a (6.23) en el espacio de Hilbert es

$$d_{\mathcal{H}}^{B,A} = -\ln \sum_i |\langle \phi_i | \Psi^{(1)} \rangle| |\langle \phi_i | \Psi^{(2)} \rangle|. \quad (6.25)$$

Notando que la función  $-\ln(x)$  es decreciente y en base a la desigualdad (6.14), llegamos a que

$$D_{\mathcal{H}}^B(|\Psi^{(1)}\rangle, |\Psi^{(2)}\rangle) = -\ln |\langle \Psi^{(1)} | \Psi^{(2)} \rangle|, \quad (6.26)$$

es el máximo de la distancia Bhattacharyya

En estos ejemplos, centramos nuestra atención en en la búsqueda de los máximos para cada una de las distancias, logrando escribir a cada una de ellas, como una función de la métrica (Riemanniana) natural en el espacio de Hilbert: el ángulo entre los rayos correspondientes a los estados. Es de gran importancia observar que esta cantidad es invariante bajo la acción del operador unitario evolución temporal,  $U$ .

### 6.2.5. Métrica Fubini-Study

En los ejemplos anteriores hemos construido una distancia sobre el espacio de Hilbert, a partir de una distancia definida sobre el conjunto  $M_+^1(N)$ . Al estudiar la métrica de Fubini-Study, procederemos de manera inversa.

Recordemos que el espacio de Hilbert  $\mathcal{H}^{n+1}$  es isomorfo al espacio proyectivo complejo  $n$ -dimensional,  $\mathcal{P}^n$

$$\mathcal{P}^n = (C^{n+1} - \{0\}) / \sim. \quad (6.27)$$

La relación de equivalencia está dada por,

$$|\Psi\rangle \sim |\Phi\rangle \Leftrightarrow \exists \lambda \in C - 0 \quad |\Psi\rangle = \lambda |\Phi\rangle. \quad (6.28)$$

En el espacio  $\mathcal{P}^n$  la métrica Fubini-Study  $\theta_{FS}$  entre dos estados  $\Psi$  y  $\tilde{\Psi}$  se define de acuerdo a la relación:

$$\cos^2\left(\frac{\theta_{FS}}{2}\right) \equiv \frac{\langle \Psi | \tilde{\Psi} \rangle \langle \tilde{\Psi} | \Psi \rangle}{\langle \Psi | \Psi \rangle \langle \tilde{\Psi} | \tilde{\Psi} \rangle}, \quad (6.29)$$

Para  $|\Psi\rangle \sim |\tilde{\Psi}\rangle$ , se tiene  $\theta_{FS} = 0$ . La máxima separación entre los estados se alcanza para  $\theta_{FS} = \pi$ .

Sea  $\mathcal{S} \subset \mathcal{P}^n$  el conjunto de estados normalizados en  $\mathcal{P}^n$ , y,  $|\Psi\rangle$  y  $|\Psi\rangle + |d\Psi\rangle$  dos estados muy próximos en  $\mathcal{S}$ . La condición de normalización implica:

$$2\text{Re}\langle \Psi | d\Psi \rangle = -\langle d\Psi | d\Psi \rangle. \quad (6.30)$$

Reemplazando  $|\tilde{\Psi}\rangle = |\Psi\rangle + |d\Psi\rangle$  en la ecuación (6.29), podemos calcular la distancia Fubini-Study entre dos estados infinitesimalmente cercanos. En este caso, a partir de la definición (6.29)

$$\cos^2\left(\frac{\delta\theta_{FS}}{2}\right) \simeq \left[1 - \frac{1}{2!} \left(\frac{\delta\theta_{FS}}{2}\right)^2 + \dots\right]^2 \simeq 1 - \frac{(\delta\theta_{FS}^2)}{4}, \quad (6.31)$$

con lo cual

$$\delta\theta_{FS}^2 = 4(\langle d\Psi | d\Psi \rangle - |\langle \Psi | d\Psi \rangle|^2). \quad (6.32)$$

Si

$$|d\Psi_{\perp}\rangle \equiv |d\Psi\rangle - |\Psi\rangle \langle \Psi | d\Psi \rangle,$$

es la proyección ortogonal sobre  $|\Psi\rangle$  de  $|d\Psi\rangle$ , la métrica Fubini-Study toma la forma [36]

$$\delta\theta_{FS}^2 = 4\langle d\Psi_{\perp} | d\Psi_{\perp} \rangle. \quad (6.33)$$

Una aproximación alternativa a la métrica Fubini-Study se puede encontrar en la referencia [117].

Supongamos que podemos hacer los siguientes desarrollos para  $|\Psi\rangle$  y  $|\tilde{\Psi}\rangle = |\Psi\rangle + |d\Psi\rangle$ :

$$\begin{aligned} |\Psi\rangle &= \sum_i \sqrt{p_i} |\phi_i\rangle \\ |\tilde{\Psi}\rangle &= \sum_i \sqrt{p_i + dp_i} |\phi_i\rangle, \end{aligned} \quad (6.34)$$

La forma más general de estos desarrollos debería incluir sendas fases, pero ellas pueden eliminarse a través de una transformación de base adecuada (ver referencia [36]). Entonces, la distancia Fubini-Study entre estos estados, hasta segundo orden en  $dp_i$ , se escribe

$$\delta\theta_{FS}^2(|\Psi\rangle, |\tilde{\Psi}\rangle) = \frac{1}{4} \sum_i \frac{dp_i^2}{p_i}. \quad (6.35)$$

Dada la condición de normalización de los estados  $\Psi$  y  $\tilde{\Psi}$ , la expresión (6.35) puede pensarse como la distancia, correspondiente a la métrica Fubini-Study, entre las distribuciones  $\{p_i\}$  y  $\{p_i + dp_i\}$  sobre el espacio  $M_+^1(N)$ .

### 6.2.6. Un nuevo criterio de distinguibilidad basado en la DJS

*Alicia se quedó un rato contemplando pensativa la seta, en un intento de descubrir cuáles serían sus dos lados, y, como era perfectamente redonda, el problema no resultaba nada fácil. Así pues, extendió los brazos todo lo que pudo alrededor de la seta y arrancó con cada mano un pedacito.*

*–Y ahora –se dijo–, ¿cuál será cuál? <sup>5</sup>*

En el presente contexto, denotaremos a la divergencia de Jensen-Shannon, entre las distribuciones  $P^{(1)}$  y  $P^{(2)}$ , por:

$$d_M^{JS}(P^{(1)}, P^{(2)}) = H\left(\frac{P^{(1)} + P^{(2)}}{2}\right) - \frac{1}{2}H(P^{(1)}) - \frac{1}{2}H(P^{(2)}), \quad (6.36)$$

donde  $H(P)$  denota la entropía de Shannon.

Nos proponemos encontrar la contrapartida cuántica de esta expresión. Para ello, consideremos nuevamente los estados  $|\Psi\rangle$  y  $|\tilde{\Psi}\rangle$ , dados por las ecuaciones (6.34). Utilizando la correspondencia entre estados y distribuciones de probabilidad (6.7), podemos evaluar la DJS entre estos estados, es decir estamos evaluando la distancia asociada en el espacio de Hilbert  $d_{\mathcal{H}}^{JS,A}(|\Psi\rangle, |\tilde{\Psi}\rangle)$ . Expandiendo la DJS en  $dp_i$ , fácilmente se encuentra que la primera contribución no nula corresponde a términos de orden cuadrático:

$$d_{\mathcal{H}}^{JS,A}(|\Psi\rangle, |\tilde{\Psi}\rangle) = \frac{1}{8} \sum_i \frac{dp_i^2}{p_i}, \quad (6.37)$$

<sup>5</sup>Lewis Carroll. Alicia en le País de las Maravillas.

es decir, a este orden, la DJS coincide (salvo por un factor 1/2) con la métrica Fubini-Study (6.35). Al mismo orden, existe una relación similar entre la DJS y las distancias de Wootters y la de Batacharyya, como puede demostrarse por simple inspección:

$$\begin{aligned} d_{\mathcal{H}}^{JS,\mathcal{A}} &= \frac{1}{2} \left( d_{\mathcal{H}}^{W,\mathcal{A}} \right)^2, \\ d_{\mathcal{H}}^{JS,\mathcal{A}} &= \frac{1}{2} \left( d_{\mathcal{H}}^{B,\mathcal{A}} \right)^2. \end{aligned} \quad (6.38)$$

Las ecuaciones (6.38) han sido establecidas hasta segundo orden en  $dp_i$ . Cuando estas relaciones son investigadas a un orden superior, emergen algunos resultados interesantes. Nos restringimos a un sistema con dos estados, siendo inmediata la extensión al caso más general. Sean  $P^{(1)} = (p, q)$  y  $P^{(2)} = (p + dp, q - dp)$  con  $p + q = 1$  dos distribuciones de probabilidad vecinas y evaluamos la DJS hasta orden  $dp^4$ . Tenemos

$$\begin{aligned} d_M^{JS} = & - \frac{1}{8} \frac{1}{(p-1)p} dp^2 + \frac{1}{16} \frac{2p-1}{p^2(p-1)^2} dp^3 \\ & - \frac{7}{192} \frac{3p^2 - 3p + 1}{p^3(p-1)^3} dp^4. \end{aligned} \quad (6.39)$$

Por otro lado, la distancia Wootters evaluada hasta el mismo orden en  $dp$ , es:

$$\begin{aligned} \frac{1}{2} (d_M^W)^2 = & - \frac{1}{8} \frac{1}{(p-1)p} dp^2 + \frac{1}{16} \frac{2p-1}{p^2(p-1)^2} dp^3 \\ & - \frac{1}{384} \frac{44p^2 - 44p + 15}{p^3(p-1)^3} dp^4. \end{aligned} \quad (6.40)$$

Sorprendentemente se puede ver la coincidencia entre ambos desarrollos, (6.39) y (6.40), hasta orden  $dp^3$ , siendo la diferencia en el cuarto orden muy pequeña (sólo un factor  $\frac{1}{192}$ ). En otras palabras, la relación

$$d_M^{JS} = \frac{1}{2} (d_M^W)^2 \quad (6.41)$$

se puede establecer hasta tercer orden en  $dp$ .

La Figura 6.1 muestra un gráfico de las distancias  $d_M^{JS}$  y  $\frac{(d_M^W)^2}{2}$  para las distribuciones binarias  $P^{(1)} = (t, 1-t)$  y  $P^{(2)} = (x, 1-x)$ . Las distancias han sido graficadas en función de  $b$  y se ha tomado  $t = 0.5$ .

Algunas observaciones sobre la relación (6.41) merecen ser explicitadas. En primer lugar no debe perderse de vista que es una relación válida hasta el tercer orden en  $dp$ , es decir, no es una identidad válida a todo orden. A su vez, puede reescribirse trivialmente en la forma  $\sqrt{2d_M^{JS}} = d_M^W$ . Como ya ha sido demostrado, precisamente el término de la izquierda de esta igualdad satisface las propiedades de una métrica. Por último, la relación (6.41) insinúa una nueva formulación del criterio de distinguibilidad de Wootters (6.9), en término de la DJS:

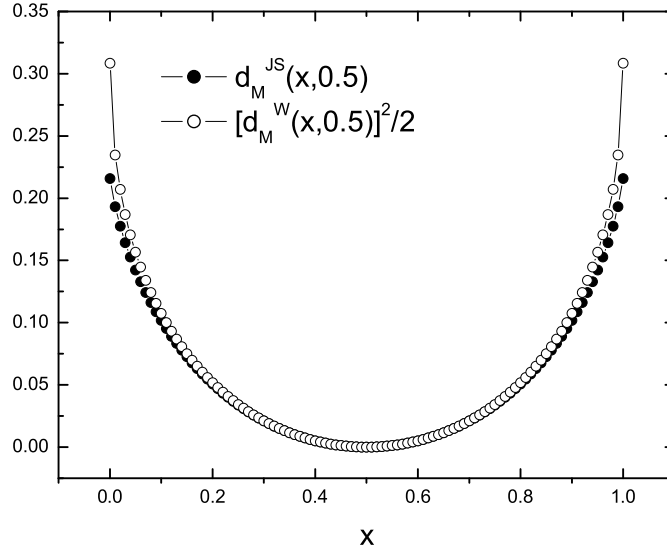


Figura 6.1:  $d_M^{JS}$  y  $\frac{(d_M^W)^2}{2}$  para  $P^{(1)} \sim P^{(2)}$ . Se eligen  $P^{(1)} = (t, 1-t)$  y  $P^{(2)} = (x, 1-x)$  y se evalúan las correspondientes distancias en función de  $x$  para  $t = 0.5$

*Dos distribuciones de probabilidad  $P^{(1)}$  y  $P^{(2)}$  son distinguibles después de  $L$  experimentos ( $L \rightarrow \infty$ ) si y sólo si*

$$[d_M^{JS}(P^{(1)}, P^{(2)})]^{1/2} > \frac{1}{\sqrt{2L}} \quad (6.42)$$

Es claro que la desigualdad (6.8) es equivalente a la desigualdad (6.42) para dos distribuciones suficientemente “cercanas”.

Entendemos que esta reformulación del criterio de distinguibilidad tiene una notable ventaja frente a la formulación original de Wootters, pues queda expresado en término de una verdadera métrica para el espacio de las distribuciones de probabilidad  $M_+^1(N)$ . Además, dado que la DJS tiene su origen en TI, nos permite establecer un puente entre esta teoría y el problema de distinguibilidad de distribuciones de probabilidad.

Queda pendiente todavía encontrar la distancia en el espacio de Hilbert asociada con la DJS. Dada la correspondencia (6.7), y el conjunto completo  $\{|\phi_i\rangle\}$  asociado con el aparato  $\mathcal{A}$ , podemos evaluar la DJS entre dos estados,  $d_{\mathcal{H}}^{JS, \mathcal{A}}(|\Psi\rangle, |\Phi\rangle)$ . En este caso el cálculo del máximo de  $d_{\mathcal{H}}^{JS, \mathcal{A}}$  no es tan directo como para las otras distancias estudiadas. Sin embargo podemos dar respuesta a este punto a través del siguiente razonamiento:

Consideremos el espacio de Hilbert de dimensión 2, y sean  $|\Psi^{(1)}\rangle$  y  $|\Psi^{(2)}\rangle$  dos estados arbitrarios normalizados (nuevamente la extensión a un número mayor de dimensiones es directa). Tenemos  $|\langle\Psi^{(1)}|\Psi^{(2)}\rangle| = \cos\varphi$  para  $\varphi \in [0, \pi/2]$ ; así  $\varphi$  es la distancia Wootters

entre  $|\Psi^{(1)}\rangle$  y  $|\Psi^{(2)}\rangle$ .

Tomemos  $\{|\phi_i\rangle\}_{i=1}^2$  una base ortonormal para  $\mathcal{H}^2$ . Cualquier otra base ortonormal  $\{|\tilde{\phi}_i\rangle\}_{i=1}^2$  se puede relacionar con  $\{|\phi_i\rangle\}$  a través de una rotación

$$\begin{aligned} |\tilde{\phi}_1(\theta)\rangle &= \frac{e^{i\theta}}{\sqrt{2}}|\phi_1\rangle + \frac{e^{-i\theta}}{\sqrt{2}}|\phi_2\rangle \\ |\tilde{\phi}_2(\theta)\rangle &= -\frac{e^{i\theta}}{\sqrt{2}}|\phi_1\rangle + \frac{e^{-i\theta}}{\sqrt{2}}|\phi_2\rangle \end{aligned} \quad (6.43)$$

con  $\theta \in [0, 2\pi]$ . Consideremos

$$p_i^{(j)} \equiv |\langle \phi_i | \Psi^{(j)} \rangle|^2$$

y

$$\tilde{p}_i^{(j)}(\theta) \equiv |\langle \tilde{\phi}_i(\theta) | \Psi^{(j)} \rangle|^2.$$

La correspondencia (6.7) permite escribir  $\langle \phi_i | \Psi^{(j)} \rangle = \sqrt{p_i^{(j)}} e^{i\alpha_i^{(j)}}$  donde  $\alpha_i^{(j)}$  son fases reales arbitrarias. Con un poco de álgebra llegamos a:

$$\tilde{p}_1^{(1)}(\theta) = \frac{p_1^{(1)} + p_2^{(1)}}{2} + \sqrt{p_1^{(1)} p_2^{(1)}} \cos(2\theta + \alpha_2^{(1)} - \alpha_1^{(1)}), \quad (6.44)$$

y

$$\tilde{p}_1^{(2)}(\theta) = \frac{p_1^{(2)} + p_2^{(2)}}{2} + \sqrt{p_1^{(2)} p_2^{(2)}} \cos(2\theta + \alpha_1^{(2)} - \alpha_2^{(2)}). \quad (6.45)$$

Además,  $\tilde{p}_2^{(1)} = 1 - \tilde{p}_1^{(1)}$  y  $\tilde{p}_2^{(2)} = 1 - \tilde{p}_1^{(2)}$ . Sin pérdida de la generalidad podemos tomar  $|\phi_1\rangle = |\Psi^{(1)}\rangle$ , entonces  $p_1^{(1)} = 1$ ,  $\alpha_1^{(1)} = 0$ ,  $p_2^{(1)} = 0$ ,  $\sqrt{p_1^{(2)}} = \cos \varphi$  y  $\sqrt{p_2^{(2)}} = \sin \varphi$ . Con estos ingredientes podemos evaluar

$$\sqrt{2d_{\mathcal{H}}^{JS, \tilde{\phi}}(\tilde{P}^{(2)}, \tilde{P}^{(1)})}$$

como una función de  $\theta$ .

La figura 6.2, muestra un gráfico de esta función para dos valores de  $\varphi$ . Por otro lado la figura 6.3 muestra la misma función, pero con un barrido continuo del ángulo  $\varphi$ . Esta figura pone de manifiesto que la función  $\sqrt{2d_{\mathcal{H}}^{JS, \tilde{\phi}}(\tilde{P}^{(2)}, \tilde{P}^{(1)})}$  está acotada superiormente por el ángulo  $\varphi$ , es decir está acotada por la distancia Wootters. De este modo podemos establecer las siguientes desigualdades:

$$D_{\mathcal{H}}^W(|\Psi^{(1)}\rangle, |\Psi^{(2)}\rangle) \geq d_{\mathcal{H}}^{W, \mathcal{A}}(|\Psi^{(1)}\rangle, |\Psi^{(2)}\rangle) \geq \sqrt{2d_{\mathcal{H}}^{JS, \mathcal{A}}(|\Psi^{(1)}\rangle, |\Psi^{(2)}\rangle)} \quad (6.46)$$

Esto nos permite concluir que  $D_{\mathcal{H}}^W(|\Psi^{(1)}\rangle, |\Psi^{(2)}\rangle)$  “representa” a  $\sqrt{2d_{\mathcal{H}}^{JS, \mathcal{A}}}$  en el espacio de Hilbert.

Dos estados que son distinguibles bajo el criterio de Jensen-Shannon, son obviamente distinguibles bajo el criterio establecido por Wootters.

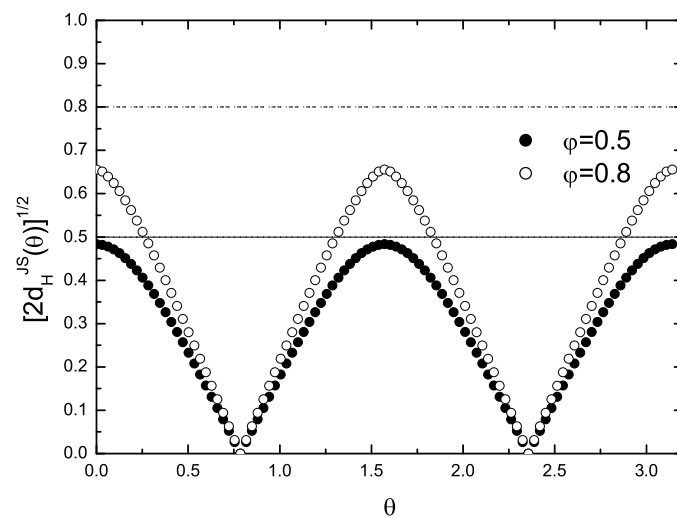


Figura 6.2:  $\sqrt{2d_{\mathcal{H}}^{JS, \tilde{\phi}}(\tilde{p}^{(2)}, \tilde{p}^{(1)})}$  en función de  $\theta$  para  $\phi = 0,5$  y  $\phi = 0,8$

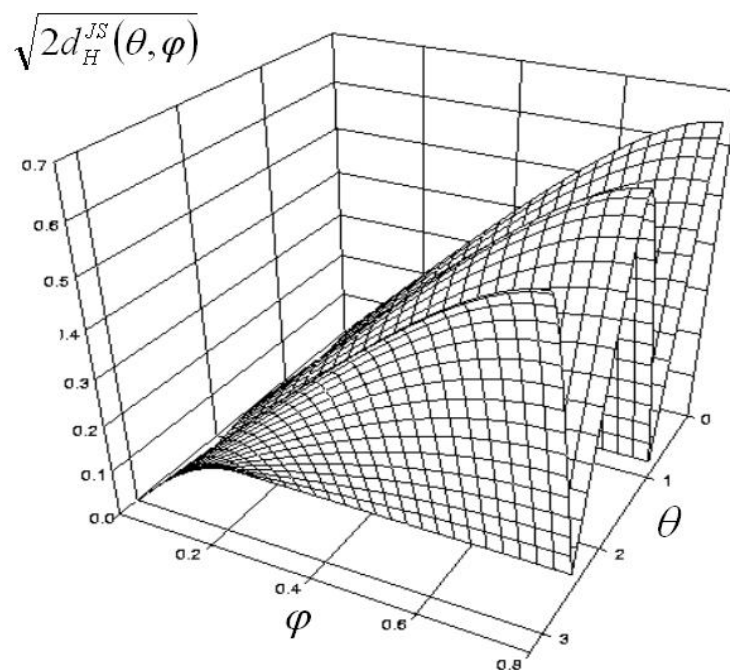


Figura 6.3: Gráfico 3D de  $\sqrt{2d_{\mathcal{H}}^{JS, \tilde{\phi}}(\tilde{p}^{(2)}, \tilde{p}^{(1)})}$  en función de  $\theta$  y  $\phi$ . Se puede apreciar claramente que esta función está acotada por el plano  $z = \phi$ .



Varias consecuencias importantes pueden extraerse de los resultados recién expuestos. En primer lugar la DJS puede tomarse como una medida unificada de distancia entre distribuciones de probabilidad y por la correspondencia ya explicitada, entre estados cuánticos puros. Además nos permite leer el teorema de Schoenberg, enunciado en el capítulo 3, a la luz de la relación (6.41) y de las desigualdades (6.46). Por último nos permiten establecer un puente entre la TI y el problema de distinguibilidad entre estados cuánticos.

En esta primera parte, focalizamos nuestra atención al estudio de estados puros. A continuación veremos cómo es posible extender estos resultados para estados mixtos.

### 6.3. Parte II: Estados mezcla

Para un sistema descrito por una matriz u operador densidad  $\rho$ , la entropía de von Neumann está definida por

$$H_N(\rho) = -Tr(\rho \log \rho) \quad (6.47)$$

con  $Tr$  el operador traza. Esta cantidad es el análogo cuántico de la entropía de Shannon, y es una medida del grado de mezcla de un estado.  $H_N$  se anula para un estado puro y alcanza su máximo valor para un estado máximamente mezclado, representado por una matriz densidad es proporcional a la matriz identidad;  $\rho = \frac{1}{\dim[\mathcal{H}]}I$ .

Entre las propiedades más importantes de  $H_N$ , destacamos:

- Positividad:  $H_N(\rho) \geq 0$  y  $H_N(\rho) = 0 \Leftrightarrow \rho$  es un estado puro.
- Invariancia frente a una evolución unitaria: si  $\sigma = U\rho U^\dagger$ , para un operador unitario  $U$ , entonces  $H_N(\sigma) = H_N(\rho)$ .
- Concavidad:  $H_N(\lambda\rho_1 + (1 - \lambda)\rho_2) \geq \lambda H_N(\rho_1) + (1 - \lambda)H_N(\rho_2) \quad \forall \lambda \in [0, 1]$
- Aditividad:  $H_N(\rho_1 \otimes \rho_2) = H_N(\rho_1) + H_N(\rho_2)$
- Relación con la entropía de Shannon: Si  $\rho = \sum_i p_i \rho_i$  es una mezcla de estados cuánticos  $\rho_i$  con  $p_i$  un conjunto de números reales positivos, tales que  $\sum_i p_i = 1$ , entonces [28]:

$$H_N \left( \sum_i p_i \rho_i \right) \leq \sum_i p_i H_N(\rho_i) + H(\{p_i\}) \quad (6.48)$$

con  $H$  la entropía de Shannon.

Una de las distancias más estudiadas en el contexto de los estados mezclas, se deriva de la entropía de Von Neumann. Esta medida se conoce como la entropía relativa y permite introducir la noción de distinguibilidad entre estados mezclas. A continuación veremos sus principales propiedades.

### 6.3.1. La entropía relativa

La entropía relativa de un operador  $\rho$  con respecto a un operador  $\sigma$ , ambos en  $\mathcal{S}(\mathcal{H})$ , es decir el conjunto de operadores hermitianos definidos positivos y de traza unidad, está dada por <sup>6</sup>:

$$S(\rho\|\sigma) = Tr[\rho(\log \rho - \log \sigma)] \quad (6.49)$$

donde  $\log$  es la función logaritmo en base 2.  $S(\rho\|\sigma)$  da una medida de cuán diferente es  $\rho$  con respecto a  $\sigma$ , en el sentido de distinguibilidad estadística, y satisface las siguientes propiedades:

- Es no negativa y se anula si y sólo si  $\rho = \sigma$ .
- Es no simétrica.
- Es no acotada.

La entropía relativa está bien definida sólo cuando el soporte de  $\sigma$  es igual o más grande que el soporte de  $\rho$ . Si esta condición no se cumple, la entropía relativa diverge  $(+\infty)$  [118]. Recordemos que el soporte de un operador es el subespacio desarrollado por los autovectores correspondientes a autovalores no nulos. Este requerimiento es muy restrictivo, y se viola en situaciones físicamente relevantes; por ejemplo cuando  $\sigma$  es un estado de referencia puro <sup>7</sup> [119].

A continuación enumeraremos las principales propiedades de la entropía relativa que la hacen una distancia atractiva en el ámbito de la MC. La demostración de cada de ellas puede encontrarse en los excelentes trabajos de Wehrl y Vedral [120] y [41].

1.  $S$  es invariante ante transformaciones unitarias:

$$S(U\rho U^\dagger\|U\sigma U^\dagger) = S(\rho\|\sigma) \quad (6.50)$$

para cualquier operador unitario  $U$ . Esta propiedad es bastante natural, pues una transformación unitaria representa una rotación en el espacio de Hilbert y la distancia entre dos estados debe ser invariante frente a una rotación.

2. Teorema H generalizado:

$$S(\Phi\rho\|\Phi\sigma) \leq S(\rho\|\sigma), \quad (6.51)$$

para cualquier mapa completo positivo  $\Phi$  que preserve la traza, dado por

$$\Phi\sigma = \sum_i V_i\sigma V_i^\dagger, \quad \sum_i V_i^\dagger V_i = 1, \quad (6.52)$$

<sup>6</sup>En el resto del presente capítulo, usaremos la notación  $D(\|)$  para indicar una distancia entre operadores.

<sup>7</sup>Esto tiene una correspondencia directa con el requerimiento de que las distribuciones de probabilidad deben ser absolutamente continuas para que la divergencia Kullback-Leibler esté bien definida.

Este resultado es muy significativo, pues la forma más general en la que un sistema cuántico abierto evoluciona, está representada matemáticamente por un mapa del tipo dado por la ecuación (6.52). Técnicamente, este tipo de mapa se conoce como un mapa  $CP$ . Pensada la entropía relativa como una medida de distinguibilidad entre estados mezclas, la ecuación (6.51) significa que una evolución no unitaria produce una disminución de la distinguibilidad entre los estados (esto justifica llamar a (6.51) el teorema H). Naturalmente, una evolución unitaria es un caso particular de un mapa  $CP$ .

Otro ejemplo de este tipo de mapa, es el de uno con una representación de la forma  $\sum_i P_i \sigma P_i$ , donde  $P_i$  es un conjunto completo de proyectores ortogonales ( $P_i^\dagger = P_i$  and  $P_i^2 = P_i$ ). En este caso se tiene

$$S(P_i \rho P_i \| P_i \sigma P_i) \leq S(\rho \| \sigma) \quad (6.53)$$

3.  $S$  es conjuntamente convexa

$$S\left(\sum_i \lambda_i \rho^{(i)} \| \sum_i \lambda_i \sigma^{(i)}\right) \leq \sum_i \lambda_i S(\rho^{(i)} \| \sigma^{(i)}), \quad (6.54)$$

donde los  $\lambda_i$  son números reales positivos tales que  $\sum_i \lambda_i = 1$ .

4. Sean  $\rho^{AB}$  y  $\sigma^{AB}$  dos matrices densidad de un sistema compuesto  $AB$ , representado por el espacio de Hilbert  $\mathcal{H}_A \otimes \mathcal{H}_B$ . Entonces  $S$  satisface

$$\begin{aligned} S(\rho^A \| \sigma^A) &\leq S(\rho^{AB} \| \sigma^{AB}) \\ S(\rho^B \| \sigma^B) &\leq S(\rho^{AB} \| \sigma^{AB}) \end{aligned} \quad (6.55)$$



Figura 6.4: Los objetos se hacen menos distinguibles cuando se dispone de información parcial sobre ellos.

donde  $\rho^A = Tr_B \rho^{AB}$  y  $\rho^B = Tr_A \rho^{AB}$  con  $Tr_A$  y  $Tr_B$  las trazas parciales sobre los respectivos espacios. Esta propiedad de la entropía relativa, tiene una interesante interpretación. En efecto, tomar la traza sobre una parte de un sistema lleva a una pérdida de información sobre esa parte, y por lo tanto su distinguibilidad es menor. Esta afirmación se encuentra representada esquemáticamente en la figura 6.4.

5.  $S$  es adititiva: Nuevamente, si  $\rho^A \otimes \rho^B$  y  $\sigma^A \otimes \sigma^B$  representan matrices densidad de un sistema compuesto  $AB$ , entonces se cumple:

$$S(\rho^A \otimes \rho^B \| \sigma^A \otimes \sigma^B) = S(\rho^A \| \sigma^A) + S(\rho^B \| \sigma^B), \quad (6.56)$$

con  $\rho^A, \sigma^A \in S(\mathcal{H}_A)$  y  $\rho^B, \sigma^B \in S(\mathcal{H}_B)$ .

6.  $S$  verifica la identidad de Donald [121]. Supongamos que los operadores densidad  $\rho_i$  ocurren con probabilidad  $p_i$ , lo que resulta en un estado promedio  $\rho = \sum_i p_i \rho_i$  y consideremos un operador densidad arbitrario  $\sigma$ . Luego,

$$\sum_i p_i S(\rho_i \| \sigma) = \sum_i p_i S(\rho_i \| \rho) + S(\rho \| \sigma). \quad (6.57)$$

### 6.3.2. Una alternativa a la entropía relativa: La DJS Cuántica

La extensión natural de la expresión (6.36) de la divergencia de Jensen-Shannon, a los operadores densidad es

$$JS(\rho \| \sigma) = H_N \left( \frac{\rho + \sigma}{2} \right) - \frac{1}{2} H_N(\rho) - \frac{1}{2} H_N(\sigma) \quad (6.58)$$

la cual puede reescribirse en términos de la entropía relativa en la forma

$$JS(\rho \| \sigma) = \frac{1}{2} \left[ S \left( \rho \| \frac{\rho + \sigma}{2} \right) + S \left( \sigma \| \frac{\rho + \sigma}{2} \right) \right] \quad (6.59)$$

Llamaremos a esta cantidad la divergencia de Jensen-Shannon cuántica (DJSC).

Si  $\rho$  y  $\sigma$  son operadores densidad con conjuntos completos de autovectores  $\{|r_i\rangle\}$  y  $\{|s_i\rangle\}$  respectivamente, tal que

$$\rho = \sum_i r_i |r_i\rangle \langle r_i|,$$

y

$$\sigma = \sum_j s_j |s_j\rangle \langle s_j|,$$

la DJSC entre  $\rho$  y  $\sigma$  se expresa como

$$JS(\rho \| \sigma) = \frac{1}{2} \left[ \sum_{k,i} |\langle t_k | r_i \rangle|^2 r_i \log \left( \frac{2r_i}{\lambda_k} \right) + \sum_{k,j} |\langle t_k | s_j \rangle|^2 s_j \log \left( \frac{2s_j}{\lambda_k} \right) \right] \quad (6.60)$$

con

$$\lambda_k = \sum_i r_i |\langle t_k | r_i \rangle|^2 + \sum_j s_j |\langle t_k | s_j \rangle|^2$$

y  $\{|t_k\rangle\}$  un conjunto completo de autovectores normalizados de  $\rho + \sigma$ .

Guiados por las propiedades de la entropía relativa, enunciaremos las principales propiedades de la DJSC:

- Positividad:  $JS(\rho||\sigma) \geq 0$  y nula si y sólo si  $\rho = \sigma$ .
- Simetría:  $JS(\rho||\sigma) = JS(\sigma||\rho)$
- Siempre está bien definida: La restricción impuesta sobre los soportes de  $\rho$  y  $\sigma$  puede ser relajada, y la DJSC está bien definida aún cuando  $\sigma$  sea un estado puro.
- Si  $\rho$  y  $\sigma$  conmutan, y por lo tanto diagonalizan en la misma base de autovectores  $|i\rangle$

$$\rho = \sum_i r_i |i\rangle\langle i| \quad \sigma = \sum_i s_i |i\rangle\langle i|$$

se cumple,

$$JS(\rho||\sigma) = JS(\{r_i\}, \{s_i\}) \quad (6.61)$$

- Acotada:

$$0 \leq JS(\rho||\sigma) \leq 1. \quad (6.62)$$

Esto se puede derivar directamente de la definición de la DJSC, de la relación (6.48) y notando que  $H(1/2, 1/2) = 1$ . La igualdad se cumple si y sólo si los estados  $\rho_i$  tienen soportes en subespacios ortogonales.

La DJSC hereda de la entropía relativa las propiedades 1-4 listadas anteriormente. Todas ellas pueden demostrarse directamente de la representación (6.59). A modo de ejemplo, veamos la prueba de la propiedad 2:

$$\begin{aligned} JS(\Phi\rho||\Phi\sigma) &= \frac{1}{2} \left[ S\left(\Phi\rho||\frac{\Phi\rho + \Phi\sigma}{2}\right) + S\left(\Phi\sigma||\frac{\Phi\rho + \Phi\sigma}{2}\right) \right] \\ &= \frac{1}{2} \left[ S\left(\Phi\rho||\frac{\Phi(\rho + \sigma)}{2}\right) + S\left(\Phi\sigma||\frac{\Phi(\rho + \sigma)}{2}\right) \right] \\ &\leq \frac{1}{2} \left[ S\left(\rho||\frac{\rho + \sigma}{2}\right) + S\left(\sigma||\frac{\rho + \sigma}{2}\right) \right] \\ &= JS(\rho||\sigma) \end{aligned} \quad (6.63)$$

La DJSC no verifica la propiedad de aditividad. Esto se debe a que, en general,  $\rho^A \otimes \rho^B + \sigma^A \otimes \sigma^B$  no se puede expresar como un producto directo de matrices densidad. Por lo tanto, que la entropía relativa satisfaga la propiedad de aditividad, no implica que la DJSC la cumpla. Sin embargo, se puede probar que la DJSC, satisface una “aditividad restringida”:

$$JS(\rho^A \otimes \rho^B || \sigma^A \otimes \rho^B) = JS(\rho^A || \sigma^A) \quad (6.64)$$

donde  $\rho^A, \sigma^A \in S(\mathcal{H}_A)$  y  $\rho^B \in S(\mathcal{H}_B)$ .

Un punto importante que debe resaltarse es que la aditividad restringida es suficiente para probar que la DJSC satisface la propiedad 4 de la entropía relativa. Si bien, como

hemos dicho recientemente, esta propiedad para la DJSC se puede derivar de la homóloga satisfecha por la entropía relativa, es esclarecedor probarla directamente. Para ello seguimos los pasos de Nielsen y Chuang [28].

Sea  $\rho^{AB}$  un operador densidad del sistema compuesto  $AB$ . Se puede demostrar que existen transformaciones unitarias  $U_j$ , definidos sobre el espacio de Hilbert asociado con la parte  $B$ , y números  $\lambda_j$ , ( $\sum_j \lambda_j = 1$ ) tales que,

$$\rho^A \otimes \frac{I}{d} = \sum_j \lambda_j U_j \rho^{AB} U_j^\dagger,$$

con  $d$  la dimensión del espacio de Hilbert  $\mathcal{H}_B$  cuyo operador identidad es  $I$ . Luego, usando las propiedades de convexidad, de invariancia bajo una evolución unitaria y de la aditividad restringida verificadas por la DJSC, tenemos:

$$\begin{aligned} JS(\rho^A \|\sigma^A) &= JS\left(\rho^A \otimes \frac{I}{d} \|\sigma^A \otimes \frac{I}{d}\right) \\ &\leq \sum_j \lambda_j JS(U_j \rho^{AB} U_j^\dagger \|\ U_j \sigma^{AB} U_j^\dagger) \\ &\leq \sum_j \lambda_j JS(\rho^{AB} \|\sigma^{AB}) = JS(\rho^{AB} \|\sigma^{AB}) \end{aligned} \quad (6.65)$$

El comportamiento no creciente de la DJSC sobre un mapa  $CP$  tiene la siguiente consecuencia: sea  $\{E_m\}$  un conjunto POVM [28]. Dados dos operadores  $\rho$  y  $\sigma$  que pertenecen a  $\mathcal{S}(\mathcal{H})$ , podemos introducir las siguientes distribuciones de probabilidad

$$p_m = Tr(\rho E_m) \quad q_m = Tr(\sigma E_m)$$

Siguiendo a Vedral [41] definimos

$$JS_1(\rho \|\sigma) = \sup_{\{E_m\}} JS(\{p_m\}, \{q_m\})$$

donde el supremo se toma sobre todos los POVMs. Dado que  $S(\rho \|\sigma) \geq S_1(\rho \|\sigma)$  con  $S_1(\rho \|\sigma) = \sup_{\{E_m\}} S(\{p_m\}, \{q_m\})$ , podemos derivar la siguiente desigualdad:

$$JS(\rho \|\sigma) \geq JS_1(\{p_m\}, \{q_m\}) \quad (6.66)$$

La aplicación  $\rho \rightarrow Tr(\rho E_m)$  es un mapa  $CP$ .

Para concluir con el listado de las propiedades DJSC, notemos que a partir de la identidad de Donald, podemos obtener una representación útil para la DJSC:

$$2JS(\rho \|\sigma) = S(\rho \|\tau) + S(\sigma \|\tau) - 2S\left(\frac{\rho + \sigma}{2} \|\tau\right), \quad (6.67)$$

donde  $\tau$  es un operador densidad arbitrario. Esta identidad nos permite evaluar la DJSC entre dos operadores densidad en términos de un tercer operador arbitrario.

### 6.3.3. La DJSC y su relación con otras medidas de distancia

En la primera parte de este capítulo, hemos mostrado que la DJS es una buena aproximación a la distancia Wootters (hasta el tercer orden de su desarrollo). Es natural, entonces, preguntarse por una relación similar entre la DJSC y la extensión propuesta por Braunstein y Caves de la distancia Wootters. Para investigar este punto, consideremos dos operadores densidad próximos,  $\rho$  y  $\rho + d\rho$ . Supongamos que  $\rho$  diagonaliza en la base  $|i\rangle$ :

$$\rho = \sum_j p_j |j\rangle\langle j|,$$

Al evaluar la DJSC entre  $\rho$  y  $\rho + d\rho$  hasta segundo orden en  $\varepsilon$ , se llega a:

$$JS(\rho||\rho + d\rho) = \frac{\varepsilon^2}{8} \left[ \sum_j \frac{\omega_{jj}}{p_j} + \sum_{j,k} (p_j - p_k)(\log p_j - \log p_k) |a_{kj}|^2 \right] \quad (6.68)$$

donde hemos supuesto que  $d\rho \equiv \varepsilon\omega$ , con  $\varepsilon \ll 1$ . Las cantidades  $\omega_{jk}$  son los elementos de la matriz  $\omega$  en la base en la que diagonaliza  $\rho$  y donde hemos supuesto que los autovectores de  $\rho + d\rho$  se escriben, (a primer orden en  $\varepsilon$ ) en la forma:

$$|j^{(1)}\rangle = \sum_k (\delta_{kj} + \varepsilon a_{jk}) |k\rangle.$$

Expandiendo

$$\log p_j - \log p_k = \frac{2(p_j - p_k)}{p_j + p_k} - \frac{4(p_j - p_k)^3}{(p_j + p_k)^3} + \dots$$

y tras un poco de álgebra, la expresión (6.68) se puede reescribir como

$$JS(\rho||\rho + d\rho) = \frac{\varepsilon^2}{8} \left[ \sum_j \frac{\omega_{jj}}{p_j} + 2 \sum_{j,k} \frac{(p_j - p_k)^2}{p_k + p_j} |a_{kj}|^2 \right] + \vartheta \quad (6.69)$$

donde  $\vartheta$  es un término que involucra sumas de potencias de  $(p_j - p_k)$  de orden par y no menor que cuatro. Los dos primeros términos coinciden con la métrica introducida por Braunstein y Caves como una generalización para estados mezcla de la distancia Wootters,  $ds_{OD}$  [36]:

$$JS(\rho||\rho + d\rho) \sim \frac{1}{8} ds_{OD}^2$$

Como estos autores notan, la métrica  $ds_{OD}$  está relacionada a su vez con la métrica de Bures,  $\Xi$ :

$$\Xi(\rho||\sigma) = \sqrt{2} [(1 - \text{Tr}(\rho^{1/2} \sigma \rho^{1/2})^{1/2})^{1/2}],$$

y establecen la siguiente relación entre ellas

$$4\Xi^2(\rho||\rho + d\rho) \sim ds_{DO}^2.$$

Entonces podemos concluir que para dos estados vecinos

$$\Xi(\rho||\rho + d\rho) \sim \sqrt{2JS(\rho||\rho + d\rho)}. \quad (6.70)$$

### 6.3.4. Ejemplo de evaluación de la DJSC

Para finalizar esta sección y a modo de ejemplo de evaluación de la DJSC, calcularemos la DJSC entre un operador densidad  $\rho_W$  correspondiente a un estado de Werner para un sistema de dos partículas de spin  $1/2$  [122], y un estado puro  $\sigma$ . Al estado de Werner lo escribimos como

$$\rho_W = F|\Psi^-\rangle\langle\Psi^-| + \frac{1-F}{3} (|\Psi^+\rangle\langle\Psi^+| + |\Phi^+\rangle\langle\Phi^+| + |\Phi^-\rangle\langle\Phi^-|)$$

con

$$|\Psi^\pm\rangle = \frac{1}{\sqrt{2}} (|\uparrow\downarrow\rangle \pm |\downarrow\uparrow\rangle)$$

y

$$|\Phi^\pm\rangle = \frac{1}{\sqrt{2}} (|\uparrow\uparrow\rangle \pm |\downarrow\downarrow\rangle).$$

Como estado puro de referencia elegimos a  $\sigma = |\Psi^-\rangle\langle\Psi^-|$ . En la expresión de arriba,  $F$  es la pureza del estado  $\rho_W$  con respecto al estado de referencia  $\sigma$ . En términos de la pureza  $F$ , la DJSC entre  $\rho_W$  y  $\sigma$ , resulta:

$$JS(\rho_W\|\sigma) = \frac{1}{2} \left[ F \log F - (1+F) \log \left( \frac{1+F}{2} \right) \right] \quad (6.71)$$

El hecho de poder evaluar la divergencia  $JS(\rho\|\sigma)$  aún cuando uno de sus argumentos corresponde a un estado puro, es una clara ventaja de la DJSC con respecto a la entropía relativa, que, como fue ya dicho, en este caso diverge.

### 6.3.5. La DJSC y la TI cuántica

Presentaremos a continuación tres aplicaciones de la DJSC en el marco de la Teoría de Información Cuántica. En primer lugar haremos una interesante interpretación de la DJSC, como el límite superior de la información accesible en un proceso de comunicación cuántica. Luego, mostramos que la distancia propuesta es útil para generar una medida de entrelazamiento, como así también que nos permite proponer una alternativa al concepto de fidelidad, comentado en la Introducción.

La TI clásica estudia el problema de enviar información, como por ejemplo letras de un alfabeto o una secuencia de símbolos, a través de canales de comunicación que operan de acuerdo a las leyes de la Física Clásica. El concepto fundamental de información y computación clásica es el *bit* y tiene dos estados, el 0 y el 1.

Por otro lado, la TI cuántica está motivada por el estudio de canales cuánticos de comunicación, y la posibilidad de poder transmitir información de una forma más eficiente y confiable, que la que posibilitan los canales clásicos. A la fecha, las posibilidades tecnológicas que este campo puede brindar, son todavía difíciles de imaginar [28].

La información y la computación cuántica se contruyen sobre un concepto análogo al del bit: el *bit cuántico* o *qubit*. El *qubit* puede estar en los estados  $|0\rangle$ ,  $|1\rangle$  o en una superposición



de ambos. Al respecto, tengamos presente la forma general del estado de un sistema de dos niveles, expresión (6.1). Es decir, el *qubit* existe en un *continuo* de estados entre  $|0\rangle$  y  $|1\rangle$  hasta que se lo observa, y por eso puede “representar” una cantidad infinita de información. De los postulados fundamentales de la MC, sabemos que el proceso de medición cambia el estado de un *qubit* colapsándolo, de la superposición de  $|0\rangle$  y  $|1\rangle$ , a un estado específico consistente con el resultado de la medición; tras la medición se tiene finalmente un *bit* de información.

Basten estas consideraciones para poner de manifiesto el carácter altamente no trivial del almacenamiento y la transmisión de la información cuántica.

La TI provee de magnitudes que permiten cuantificar las características y capacidades de los procesos de comunicación por canales cuánticos. Por ejemplo, la información accesible en un proceso de comunicación, es una de tales cantidades. Eencialmente mide cuánta es la información del mensaje enviado en relación al mensaje recibido. Aunque parezca extraño, en el contexto cuántico hay una cota superior (y también una inferior) para esta cantidad, la cual es conocida como la cota de Holevo [123]. A continuación haremos una interpretación de la DJSC a la luz de esta cota.

### 6.3.6. Accesibilidad de la información cuántica

La cota Holevo juega un papel muy importante en una gran cantidad de aplicaciones de teoría de la información cuántica [28]. Antes de enunciarla, repasaremos el concepto de información mútua.

Sean  $X$  e  $Y$  dos variables aleatorias. Formalicemos la pregunta que apareció un poco más arriba: ¿cuánto es el contenido de información de la variable  $X$  en relación al contenido de información de la variable  $Y$ ? Recordemos también que la información es una magnitud de significado preciso y su cuantificador es la entropía de Shannon. Denotemos por  $p(x, y)$  a la probabilidad del par  $(X, Y)$ . Para medir la incerteza total acerca del par  $(X, Y)$  podemos introducir la *entropía conjunta* de  $X$  e  $Y$ :

$$H(X, Y) \equiv - \sum_{x,y} p(x, y) \log p(x, y). \quad (6.72)$$

Al conocer el valor de  $X$  tenemos  $H(X)$  bits de información sobre el par  $(X, Y)$ . Análogamente, si conocemos el valor de  $Y$  tenemos  $H(Y)$  de información sobre el par  $(X, Y)$ . Al sumar estas últimas cantidades, estamos duplicando la información que tienen en común  $X$  e  $Y$ ; mientras que la información que no tienen en común se ha contado exactamente una vez. Si restamos de esta suma la información conjunta de  $(X, Y)$ , dada por  $H(X, Y)$ , nos quedamos entonces con la información en común que tienen  $X$  e  $Y$ :

$$IM(X : Y) \equiv H(X) + H(Y) - H(X, Y). \quad (6.73)$$

Esta cantidad se conoce como la *información mútua* de  $X$  e  $Y$ . Estamos ahora en condiciones de enunciar el siguiente

**Teorema 3** (*Cota Holevo*): Supongamos que se prepara el estado  $\rho_i$  con  $i = 1, \dots, n$  con probabilidades  $p_1 \dots p_n$  y que luego se realiza una medición descrita por un conjunto de elementos POVMs,  $\{E_m\}$ , sobre el estado preparado  $\rho_i$ . Supongamos que como resultado de esa medición se obtiene el valor  $j$ . La cota Holevo establece que para cualquiera de tales mediciones se cumple:

$$IM(i : j) \leq \chi = H_N(\rho) - \sum_i p_i H_N(\rho_i) \quad (6.74)$$

donde  $\rho = \sum_x p_x \rho_x$ .

La cantidad  $\chi$  está relacionada con la DJSC. En efecto: si combinamos la extensión propuesta de la DJS como una medida de distancia entre un número arbitrario de distribuciones de probabilidad con diferentes pesos, con nuestra definición de la DJSC, vemos que la cantidad  $\chi$  es precisamente la DJSC, con pesos  $p_i$  entre los operadores densidad  $\rho_i$ :

$$\chi \equiv JS^{(p_1, \dots, p_n)}(\rho_1 \| \dots \| \rho_n). \quad (6.75)$$

Esto da, a nuestro entender, una interesante interpretación física de la DJSC.

### 6.3.7. Entrelazamiento

El entrelazamiento cuántico (quantum entanglement) es un recurso físico asociado con correlaciones no clásicas, de características peculiares, que se dan entre sistemas cuánticos separados [124]. El entrelazamiento es básico en la implementación de procesos de información cuántica.

Un estado de un sistema bi-partito  $AB$  se denomina separable, o sólo clásicamente correlacionado, si se puede escribir como una combinación convexa de estados producto. En términos más formales, se dice que  $\rho \in S(\mathcal{H}_A \otimes \mathcal{H}_B)$  es separable si se puede representar de la forma:

$$\rho = \sum_{i=1}^N p_i \rho_i^A \otimes \rho_i^B, \quad (6.76)$$

donde  $0 \leq p_1, \dots, p_N \leq 1$  y  $\sum p_i = 1$ . Los estados  $\rho_i^A$  se toman del espacio de estados  $S(\mathcal{H}_A)$ , y  $\rho_i^B$  son elementos de  $S(\mathcal{H}_B)$ . Diremos que un estado es entrelazado si no es separable, es decir si no admite una representación de la forma (6.76). Los estados no entrelazados son los más generales que dos operadores imaginarios, pueden crear por operaciones locales y comunicaciones clásicas.

El conjunto de todos los estados no entrelazados es convexo, es decir, una combinación convexa de dos estados no entrelazados es no entrelazado.

Los estados (6.76) son los únicos estados sin entrelazamiento. Todo otro estado tendrá algún grado de entrelazamiento. Para poder indicar con precisión ese grado de entrelazamiento, se han propuesto diferentes medidas del grado de entrelazamiento.

Denotemos por  $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$  el espacio de Hilbert de un sistema cuántico compuesto por dos subsistemas  $A$  y  $B$ . De acuerdo a Vedral, cualquier medida de entrelazamiento  $\mathcal{E}$ , debe satisfacer las siguientes condiciones (necesarias) [125]:

- i.*  $\mathcal{E}(\rho) = 0 \Leftrightarrow \rho$  es separable.
- ii.*  $\mathcal{E}(\rho) = \mathcal{E}(U_A \otimes U_B \rho U_A^\dagger \otimes U_B^\dagger)$  para dos operadores unitarios arbitrarios  $U_A$  y  $U_B$  que actúan sobre los respectivos espacios de Hilbert.
- iii.*  $\mathcal{E}(\rho)$  es convexa respecto a sus argumentos:

$$\mathcal{E}\left(\sum_j \lambda_j \rho_j\right) \leq \sum_j \lambda_j \mathcal{E}(\rho_j)$$

- iv.*  $\mathcal{E}(\rho)$  decrece bajo operaciones cuánticas genéricas, es decir, si  $\{V_i\}$  son operadores tales que  $\sum V_i V_i^\dagger = 1$ ,  $\rho_j = \frac{V_j \rho V_j^\dagger}{\lambda_j}$  y  $\lambda_j = \text{Tr}(V_j \rho V_j^\dagger)$ , entonces

$$\mathcal{E}(\rho) \geq \sum_j \lambda_j \mathcal{E}(\rho_j)$$

Vedral y colaboradores [125] mostraron que una medida de entrelazamiento definida como

$$\mathcal{E}(\rho) = \inf_{\sigma} D(\rho \parallel \sigma) \tag{6.77}$$

satisface las propiedades *i-iv*, si la distancia  $D : \mathcal{S} \otimes \mathcal{S} \rightarrow \mathfrak{R}$ , verifica las siguientes dos propiedades:

- (a)  $D(\rho \parallel \sigma) \geq 0$  y  $D(\rho \parallel \rho) = 0$  para cualquier  $\rho, \sigma \in \mathcal{S}$
- (b)  $D(\Phi \rho \parallel \Phi \sigma) \leq D(\rho \parallel \sigma)$  para cualquier mapa  $CP$ ,  $\Phi$ .

En (6.77) el ínfimo se toma sobre el conjunto de todos los estados separables, que llamaremos  $\mathcal{D}$ .

Estos requerimientos admiten una interesante interpretación gráfica; ver figura 6.5: Sean  $\mathcal{T}$  el conjunto de las matrices densidad de sistemas cuánticos bi-partitos,  $\mathcal{D}$  el conjunto de todos los estados separables y  $\mathcal{E} = \mathcal{T} - \mathcal{D}$  el conjunto de los estados entrelazados. Se busca el operador  $\sigma^*$  en  $\mathcal{D}$  más cercano a  $\rho$  bajo la medida de distancia  $D$ .  $\sigma^*$  aproxima las correlaciones clásicas de  $\rho$  tanto como es posible. Es por esto que  $\mathcal{E}(\rho)$  mide las correlaciones cuánticas. Esto sugiere dos contribuciones diferentes a las correlaciones de  $\rho$ :

- las correlaciones cuánticas,  $\mathcal{E}(\rho)$ , y
- las correlaciones clásicas,  $D(\sigma^* \parallel \sigma_A^* \otimes \sigma_B^*)$ ,

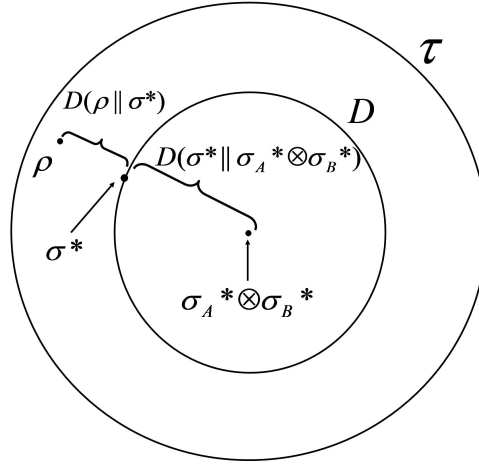


Figura 6.5: Interpretación gráfica de las condiciones sobre la distancia  $D$ ; ver texto para detalles

donde  $\sigma^*$  es el estado no entrelazado que minimiza  $D$  y,  $\sigma_A^*$  y  $\sigma_B^*$  son sus partes reducidas.

Notablemente, la DJSC  $JS(\rho||\sigma)$  satisface los requisitos (a) y (b) impuestos sobre una distancia que permita definir una medida de entrelazamiento adecuado. Por ello, proponemos introducir una medida de entrelazamiento de la forma:

$$\mathcal{E}_{JS}(\rho) = \inf_{\sigma \in \mathcal{D}} JS(\rho||\sigma) \quad (6.78)$$

donde nuevamente, el ínfimo se toma sobre todos los estados separables.

Queda por investigar las ventajas que esta medida del grado de entrelazamiento podría tener frente a otras propuestas, por ejemplo, frente a la homóloga definida a partir de la entropía relativa [125].

### 6.3.8. Fidelidad

La noción de *fidelidad* fue comentada en la Introducción, en el contexto de las diferentes distancias sobre el conjunto de las distribuciones de probabilidad. Naturalmente, se la puede extender como una cantidad definida sobre el conjunto de los operadores densidad [126, 127]. Si  $\rho$  y  $\sigma$  son dos operadores densidad, entonces la fidelidad entre ellos está dada por:

$$F(\rho||\sigma) = Tr \sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}. \quad (6.79)$$

Sus principales propiedades son:

1. La fidelidad es simétrica, invariante bajo transformaciones unitarias y acotada entre 0 y 1.

2. Si  $\rho$  y  $\sigma$  comuntan, es decir son diagonales en la misma base,

$$\rho = \sum_i r_i |i\rangle\langle i|; \quad \sigma = \sum_i s_i |i\rangle\langle i|,$$

para alguna base ortonormal  $|i\rangle$ , entonces:

$$\begin{aligned} F(\rho||\sigma) &= \text{Tr} \sqrt{\sum_i r_i s_i |i\rangle\langle i|} \\ &= \text{Tr} \left( \sum_i \sqrt{r_i s_i} |i\rangle\langle i| \right) \\ &= \sum_i \sqrt{r_i s_i} \\ &= F(r_i, s_i), \end{aligned} \tag{6.80}$$

es decir en este caso la fidelidad se reduce a la fidelidad clásica (1.8) entre los autovalores  $r_i$  y  $s_i$  de  $\rho$  y  $\sigma$ , respectivamente.

3. Para un estado puro  $\sigma = |\Psi\rangle\langle\Psi|$  y un estado arbitrario  $\rho$ , la fidelidad resulta:

$$F(|\Psi\rangle\langle\Psi||\rho) = \sqrt{\langle\Psi|\rho|\Psi\rangle}, \tag{6.81}$$

es decir, es igual a la raíz cuadrada del solapamiento entre  $|\Psi\rangle$  y  $\rho$ .

4. La fidelidad está relacionada a la métrica Bures a través de la expresión,

$$\Xi^2(\rho||\sigma) = 2(1 - F(\rho||\sigma)). \tag{6.82}$$

Prestemos un poco de atención a esta última identidad. La relación existente entre la DJSC y la métrica de Bures, expresión (6.70), sugiere definir una fidelidad alternativa utilizando a la DJSC:

$$F_{JS}(\rho||\sigma) \equiv 1 - JS(\rho||\sigma) \tag{6.83}$$

De la propiedad (6.62),  $F_{JS}$  debe estar acotada entre 0 y 1. En particular, para  $\rho = \sigma$ ,  $F_{JS} = 1$  y  $F_{JS}(\rho||\sigma) = 0$  si y sólo si  $\rho$  y  $\sigma$  tienen soportes sobre bases ortogonales.

Podemos ir un poco más allá con la propuesta (6.83). En efecto, un simple cálculo permite mostrar que la fidelidad entre dos estados  $\rho$  y  $\sigma$  se puede evaluar en términos de las *purificaciones* de estos estados como

$$F(\rho||\sigma) = \max_{|\Phi\rangle} |\langle\Psi|\Phi\rangle| \tag{6.84}$$

donde  $|\Psi\rangle$  es cualquier purificación fija de  $\rho$ , y la maximización es sobre todas las purificaciones de  $\sigma$  [28]. Por otro lado, recordemos que la distancia Wootters entre dos estados

puros  $|\Psi\rangle$  y  $|\Phi\rangle$ , está dada por la ecuación (6.17). La fidelidad se puede expresar de la siguiente forma en términos de esta distancia:

$$\begin{aligned} F(\rho||\sigma) &= \max_{|\Phi\rangle} \cos(D_{\mathcal{H}}^W(|\Psi\rangle|||\Phi\rangle)) \\ &= \max_{|\Phi\rangle} \left[ 1 - \frac{(D_{\mathcal{H}}^W(|\Psi\rangle|||\Phi\rangle))^2}{2} + \dots \right] \end{aligned} \quad (6.85)$$

Teniendo en cuenta la relación entre la distancia Wootters y la DJS [128], dada por la ecuación (6.41) podemos aproximar

$$F(\rho||\sigma) \sim \max_{|\Phi\rangle} (1 - JS(|\Psi\rangle|||\Phi\rangle)) \quad (6.86)$$

la cual pensamos se puede tomar como punto de partida para investigar a (6.83) como una medida de fidelidad.

## 6.4. Comentarios finales

La noción de distinguibilidad de estados cuánticos es de fundamental importancia en computación e información cuánticas. Cuanto mejor podamos distinguir dos estados, más información tendremos sobre ellos. En este capítulo hemos introducido la DJS como una medida de distancia entre estados cuánticos y hemos demostrado que esta medida tiene notables ventajas formales frente a otras medidas anteriormente propuestas. La DJS está siempre bien definida, es acotada y tiene importantes propiedades geométricas. En la primera parte del capítulo nos limitamos al estudio de estados puros y formulamos un criterio de distinguibilidad. En la segunda parte generalizamos la DJS para estados mezcla, estudiamos su relación con otras medidas de distancia y estudiamos varias aplicaciones en el contexto de la TI cuántica.

# A modo de conclusión...

## ¿Qué queda por hacer?

En esta tesis hemos estudiado una distancia entre distribuciones de probabilidad originada en la TI, conocida como la divergencia de Jensen-Shannon (DJS). Hemos realizado un estudio exhaustivo de la misma dentro de tres marcos teóricos fundamentales para la Física actual. Estudiamos sus propiedades y aplicaciones al análisis de secuencias simbólicas en el contexto de la MEBG, luego generalizamos la medida en el marco de la qME y finalmente extendimos su dominio de aplicabilidad al formalismo de la Mecánica Cuántica, tanto para el caso de estados puros como para el de estados mezclas.

En el contexto del análisis de secuencias, queda estudiar la aplicación de  $C_S$  a sistemas dinámicos. Para esto será necesario la definición de un mapa entre la dinámica del sistema y una secuencia de símbolos. En este mismo contexto nos parece interesante investigar la complejidad de secuencias mitocondriales y estudiar posibles relaciones entre las secuencias mitocondriales y cromosómicas de diferentes especies en lo que respecta por ejemplo a evolución.

Se han realizado avances en la demostración de que la DJS es una métrica en el espacio de Hilbert, hemos demostrado que verifica la desigualdad triangular al menos para casos particulares de operadores densidad en un espacio de Hilbert de dimensión  $2 \times 2$ . Sería interesante continuar con esta línea de investigación, debido a la relevancia de tener medidas que sean métricas.

Dentro del formalismo cuántico, estudiamos una generalización de la DJS cuántica a la qME. Esta medida se puede escribir de la siguiente forma para dos operadores densidad,

$$JS_q[\rho|\sigma] = \frac{1}{2}S_q[\rho|(\rho + \sigma)/2] + \frac{1}{2}S_q[\sigma|(\rho + \sigma)/2] \quad (6.87)$$

donde  $S_q$  es la q-entropía relativa [119]. Hemos verificado que 6.87 satisface las siguientes propiedades:

1.  $JS_q$  es definida positiva.
2. Es conjuntamente convexa para  $q \in (0, 1)$ .
3. Satisface la aditividad restringida:

$$JS_q[\rho_1 \otimes \rho_2|\sigma_1 \otimes \rho_2] = JS_q[\rho_1|\sigma_1].$$

4. Decrece monótonamente por mediciones proyectivas.
5. Es invariante frente a transformaciones unitarias para  $q \in (0, 1)$ .
6. Decrece para mapas-CP.

Esta medida podría resultar de utilidad para caracterizar el entrelazamiento cuántico. Se han realizado algunas aplicaciones para sistemas bipartitos de espín  $1/2$ . Quizás es posible determinar algún criterio de separabilidad dependiente del parámetro de deformación  $q$ .

Otro asunto que debería investigarse es el significado de la relación existente entre la cantidad de Holevo,  $\chi$ , y la DJS cuántica.



# Bibliografía

- [1] B. Lesche. Instabilities of Rényi entropies. *J. Stat. Phys.* **27**, pages 419–422, 1982.
- [2] C. Tsallis y E. Brigatti. Nonextensive statistical mechanics: A brief introduction. *Continuum Mech. Thermodyn.* **16**, pages 223–235, 2004.
- [3] R. López-Ruiz, H. Mancini y X. Calbet. A statistical measure of complexity. *Phys. Lett. A* **209**, pages 321–326, 1995.
- [4] M.T. Martin, A. Plastino y O. Rosso. Statistical complexity and disequilibrium. *Phys. Lett. A* **311**, pages 126–132, 2003.
- [5] P.W. Lamberti, M.T. Martin, A. Plastino y O. Rosso. Intensive entropic non-triviality measure. *Physica A* **334**, pages 119–131, 2004.
- [6] R.V. Hamming. Error detecting and error correcting codes. *Bell Sys. Tech. Journal* **29**, pages 147–160, 1950.
- [7] V.I. Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics* **10**, pages 707–710, 1966.
- [8] C.E. Shannon. A Mathematical Theory of Communication. *Bell. Syst. Tech. J.* **27**, pages 379–423, 1948.
- [9] E.T. Jaynes. Information Theory and Statistical Mechanics. *Phys. Rev.* **106**, pages 620–630, 1957.
- [10] E.T. Jaynes. Information Theory and Statistical Mechanics II. *Phys. Rev.* **108**, pages 171–190, 1957.
- [11] R. A. Fisher. Theory of statistical estimation. *Proc. Cambridge Phil. Society*, **22**, pages 700–725, 1925.
- [12] H.E. Stanley, S.V. Buldyrev, A.L. Goldberger, Z.D. Goldberger, S. Havlin, R.N. Mantegna, S.M. Ossadnik, C-K Peng y M. Simons. Statistical mechanics in biology: how ubiquitous are long-range correlations? *Physica A* **205**, pages 214–253, 1994.
- [13] H. Zhou y G. Slater. A metric to search for relevant words. *Physica A* **329**, pages 309–327, 2003.

- [14] F.S. Collins, M. Morgan y A. Patrinos. The Human Genome Project: Lessons from Large-Scale Biology. *Science* **300**, pages 286 – 290 y otros artículos en Special Section: Building on the DNA Revolution, *Science* **300**, 2003.
- [15] J. Arnold y Ñ. Hilton. Genome sequencing: Revelations from a Bread Mould. *Nature* **422**, pages 821 y otros artículos en esta edición especial, DNA 50th anniversary: Double helix at 50, *Nature* **422**, 2003.
- [16] Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, pages 69–87, 2005.
- [17] J.L. Boore y W.M. Brown. Big tree from little genomes: mitochondrial gene order as a phylogenetic tool. *Curr. Opin. Genet. Dev.*, **8**, pages 668–674, 1998.
- [18] S.T. Fitz-Gibbon y C.H. House. Whole genome-based phylogenetic analysis of free-living microorganisms. *Nucleic Acids Res.*, **27**, pages 4218–4222, 1999.
- [19] B.E. Blaisdel. A measure of similarity of sets of sequences not requiring sequence alignment. *Proc. Natl. Acad. Sci. USA*, **83**, pages 5155–5159, 1986.
- [20] T. Wu, J. Burke y D. Davison. A measure of DNA sequence dissimilarity based on Mahalanobis distance between frequencies of words. *Biometrics*, **53**, pages 1431–1439, 1997.
- [21] T. Wu, Y. Hsieh y L. Li. Statistical measures of DNA sequence dissimilarity under Markov chain models of base composition. *Biometrics*, **57**, pages 441–443, 2001.
- [22] A.C. Yang, A.L. Goldberger y C.K. Peng. Genomic classification using a new information-based similarity index: Application to the sars coronavirus. *J. Comput. Biol.* **12(8)**, pages 1103–1116, 2005.
- [23] A.C. Yang, S. Hseu, H. Yien, A. Goldberger y C. Peng. Linguistic analysis of the human heartbeat using frequency and rank order statistics. *Phys. Rev. Lett.* **90**, page 108103, 2003.
- [24] A.C. Yang, C.K. Peng, H.W. Yien y A.L. Goldberger. Information categorization approach to literary authorship disputes. *Physica A* **329**, pages 473–483, 2003.
- [25] M. Li y P. Vitanyi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer-Verlag, New York, 2nd ed., 1997.
- [26] M. Li, J.H. Badger, X. Chen, S. Kwong, P. Kearney y H. Zhang. An information-based sequence distance and its applications to whole mitochondrial genome phylogeny. *Bioinformatics* **17**, pages 149–154, 2001.
- [27] C. Bennet, M. Li y B. Ma. Linking chain letter. *Scientific American*, **76**, page 81, 2003.

- [28] M.A. Nielsen y I.L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, 2000.
- [29] V. Vedral. The role of relative entropy in quantum information theory. *Rev. Mod. Phys.* **74**, pages 197–234, 2002.
- [30] M. Ravicule, M. Casas y A. Plastino. Information and metrics in Hilbert space. *Phys. Rev. A*, **55**, page 1695, 1997.
- [31] L. Knöll y A. Orłowski. Distance between density operators: Applications to the Jaynes-Cummings model. *Phys. Rev. Lett.* **51**, pages 1622–1630, 1995.
- [32] A. Wünsche. *Appl. Phys. B* **60**, page 119, 1995.
- [33] M. Hillery. Nonclassical distance in quantum optics. *Phys. Rev. A* **35**, pages 725–732, 1987.
- [34] W.K. Wootters. Statistical distance and hilbert space. *Phys. Rev. D* **72**, pages 357–362, 1981.
- [35] D.N. Page. Geometrical description of berry’s phase. *Phys. Rev. A* **36**, pages 3479–3481, 1987.
- [36] S.L. Braunstein y C.M. Caves. Statistical distance and the geometry of quantum states. *Phys. Rev. E* **72**, pages 3439–3443, 1994.
- [37] D. Bures. An extension of Kakutani’s theorem on infinite product measures to the tensor product of semifinite  $w^*$ -algebras. *Trans. Amer. Math. Soc.* **135**, page 199212, 1969.
- [38] A. Uhlmann. The “transition probability” in the state space of a  $*$ -algebra. *Rep. Math. Phys.* **9**, pages 273–279, 1976.
- [39] M. Hübner. *Phys. Lett. A* **163**, page 239, 1992.
- [40] R. Jozsa. Fidelity for mixed quantum states. *J. Mod. Opt.* **41**, pages 2315–2323, 1994.
- [41] V. Vedral, M.B. Plenio, M.A. Rippin, P.L. Knight. Quantifying entanglement. *Phys. Rev. Lett.* **78**, pages 2275–2279, 1997.
- [42] K. Życzkowski y W. Stomczyński. The monge distance between quantum states. *J.Phys. A: Math. Gen.* **31**, pages 9095–9104, 1998.
- [43] K. Husimi. Some formal properties of the density matrix. *Proc. Phys. Math. Soc. Japan* **22**, page 264, 1940.

- [44] Jr. W.T. Grandy. Information Theory in Physics. *Am. J. Phys.* **65** (6), pages 466–476, 1997.
- [45] H. Nyquist. Certain Factors Affecting Telegraph Speed. *Bell. Syst. Tech. J.* **3**, pages 324–346, 1924.
- [46] R.V.L. Hartley. Transmission of Information. *Bell. Syst. Tech. J.* **7**, pages 535–563, 1928.
- [47] N. Wiener. *Cybernetics*. Mit Cambridge MA, 1948.
- [48] D.V. Gokhale y S. Kullback. *Information in Contingency Tables*. Marcel Dekker, 1978.
- [49] C.K. Chow y C.N. Liu. Approximation discrete probability distributions with dependence trees. *IEE Trans. Inform. Theory* **IT-14**, pages 462–467, 1968.
- [50] T. Kailath. The divergence and Bhattacharyya distance measure in signals selection. *IEEE Trans. Commun. Technol.* **COM-15**, pages 52–60, 1967.
- [51] M. Ben Bassat. f-entropies, probability of error, and feature selection. *Inform. Contr.* **39**, pages 131–142, 1966.
- [52] T.M. Cover y J.A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc, New York, 1991.
- [53] S. Kullback. *Information Theory and Statistics*. New York: Dover Publications, 1968.
- [54] C. Rao. Differential Geometry in Statistical Inference. *IMS-Lectures Notes* **10**, page 217, 1987.
- [55] J. Lin. Divergence Measures Based on the Shannon Entropy. *IEEE* **37** No. 1, pages 145–151, 1991.
- [56] F. Topsøe. Inequalities for the Jensen-Shannon Divergence. <http://www.math.ku.dk/topsoe/>.
- [57] J. Schindelin D. Endres. A new metric for probability distributions. *IEEE Trans. Inform. Theory* **49** (7), pages 1858–1860, 2003.
- [58] I. Grosse, P. Bernaola Galván, P. Carpena, R. Román-Roldán, J.L. Oliver y H.E. Stanley. Analysis of symbolic sequences using the Jensen-Shannon divergence. *Phys. Rev. E* **65** (041), page 905, 2000.
- [59] C. Tsallis. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* **52**, page 479, 1988.

- [60] C. Tsallis, G. Bemski y R.S. Mendes. Is re-association in folded proteins a case of nonextensivity? *Phys. Lett. A* **257**, page 93, 1999.
- [61] C. Tsallis, J.C. Anjos y E.P. Borges. Fluxes of cosmic rays: A delicately balanced stationary state. *Phys. Lett. A* **310**, page 372, 2003.
- [62] C. Beck. Dynamical foundations of nonextensive statistical mechanics. *Phys. Rev. Lett.* **87**, page 180601, 2001.
- [63] C. Beck, G.S. Lewis y H.L. Swinney. Measuring non-extensivity parameters in a turbulent Couette-Taylor flow,. *Phys. Rev. E* **63**, page 035303, 2001.
- [64] N. Arimitsu y T. Arimitsu. Analysis of velocity derivatives in turbulence based on generalized statistics. *Europhys. Lett.* **60**, page 60, 2002.
- [65] L. Borland. Closed form option pricing formulas based on a non-Gaussian stock price model with statistical feedback. *Phys. Rev. Lett.* **89**, page 098701, 2002.
- [66] I. Bediaga, E.M.F. Curado y J. Miranda. nonextensive thermodynamical equilibrium approach in  $e^+e^- \rightarrow$  hadrons. *Physica A* **286**, page 156, 2000.
- [67] A. Upadhyaya, J.P. Rieu, J.A. Glazier y Y. Sawada. Anomalous diffusion and non-Gaussian velocity distribution of Hydra cells in cellular aggregates. *Physica A* **293**, page 549, 2001.
- [68] O.A. Rosso, M.T. Martin y A. Plastino. Brain electrical activity analysis using wavelet based informational tools. *Physica A* **313**, page 587, 2002.
- [69] C. Tsallis, D. Prato y A.R. Plastino. Nonextensive statistical mechanics: Some links with astronomical phenomena. *en Proc. XIth United Nations / European Space Agency Workshop on Basic Space Science*, page 259, 2004.
- [70] L.C. Malacarne, R.S. Mendes y E.K. Lenzi. q-Exponential distribution in urban agglomeration. *Phys. Rev. E* **65**, page 017106, 2002.
- [71] S. Abe yÑ. Suzuki. Itineration of the Internet over nonequilibrium stationary states in Tsallis statistics. *Phys. Rev. E* **67**, page 016106, 2003.
- [72] <http://tsallis.cat.cbpf.br/biblio.htm>.
- [73] V. Latora, M. Baranger, A. Rapisarda y C. Tsallis. The rate of entropy increase at the edge of chaos. *Phys. Lett. A* **273**, page 97, 2000.
- [74] L.G. Moyano, A.P. Majtey y C. Tsallis. Weak chaos and metastability in a symplectic system of many long-range-coupled standard maps. *Enviado a Eur. Phys. J. B*, 2006.

- [75] C. Tsallis, S.V.F. Levy, A.M.C. Souza y R. Maynard. Statistical-Mechanical Foundation of the Ubiquity of Lévy Distributions in Nature. *Phys. Rev. Lett.* **75**, pages 3589–3593, 1995.
- [76] S. Buldyrev, A. Goldberger, S. Havlin, C-K. Peng, M. Simons y H.E. Stanley. Generalized Lévy Walk Model for DNA Nucleotide Sequences. *Phys. Rev. E* **47**, pages 4514–4523, 1993.
- [77] A.R. Plastino y C. Tsallis L. Borland. Information gain within nonextensive thermostatics. *J. Math. Phys.* **39**.
- [78] C. Tsallis. Generalized entropy-based criterion for consistent testing. *Phys. Rev. E* **58**, page 1442, 1998.
- [79] W.H. Press, B.P. Flannery, S.A. Teukolsky y W.T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing (Hardcover), 2nd ed.* Cambridge University Press, 1984.
- [80] P. Bernaola-Galván, J.L. Oliver y R. Román-Roldán. Decomposition of DNA Sequence Complexity. *Phys. Rev. Lett.* **83**, page 5181, 1999.
- [81] J. Christensen y P. Ressel C. Berg. *Harmonic Analysis on Semigroups.* Springer-Verlag, New York, 2nd ed., 1984.
- [82] I.J. Schoenberg. Metric spaces and positive definite functions. *Trans. Am. Math. Soc.* **44**, pages 522–536, 1938.
- [83] C. Adami. What is complexity? *BioEssays*, **24**, pages 1085–1094, 2002.
- [84] M. Gell-Mann y S. Lloyd. Information measures, effective complexity, and total information. complexity. *Complexity*, **2(1)**, pages 44–52, 1996.
- [85] S.A. Kauffman. *The Origins of Order: Self-Organization and Selection in Evolution.* Oxford University Press, 1993.
- [86] R. Solé y B. Goodwin. *Signs of Life: How Complexity pervades Biology.* Basic Books, 2000.
- [87] D. Feldman y J. Crutchfield. Measures of statistical complexity: why? *Phys. Lett. A* **238**, pages 244–252, 1998.
- [88] A.N. Kolmogorov. Three approaches to the concept of the amount of information. *Prob. Info. Trans.*, **1**, pages 1–7, 1965.
- [89] A. Lempel y J. Ziv. ver. *IEEE Transaction on Information Theory*, **22(1)**, page 7581, 1976.

- [90] S. Lloyd y H. Pagels. Complexity as thermodynamic depth. *Ann. Phys.* **188**, pages 186–213, 1988.
- [91] J.P. Crutchfield y K. Young. Inferring statistical complexity. *Phys Rev Lett*, **63**, pages 105–108, 1989.
- [92] P. Grassberger. Toward a quantitative theory of self-generated complexity. *Int J. Theor. Phys.* **25**, pages 907–928, 1986.
- [93] DW. McShea. Metazoan complexity and evolution: is there a trend? *Evolution* **50**, pages 477–492, 1996.
- [94] DW. McShea. Functional complexity en organisms: Parts as proxies. *Biology and Philosophy* **15**, pages 641–668, 2000.
- [95] G. Bell y A.O. Mooers. Size an complexity among multicellular organisms. *Biol J Linnean Society* **60**, pages 345–363, 1997.
- [96] C. Adami yÑ.J. Cerf. Physical complexity. *Physica D* **137**, pages 62–69, 1999.
- [97] D.R. Chialvo. Critical brain networks. *Physica A* **340**, pages 756–765, 2004.
- [98] P. Bak, C. Tang y K. Wiesenfeld. Self-organized criticality: An explanation of  $1/f$  noise. *Rev. Lett.* **59**, pages 381–384, 1987.
- [99] M.Ñewman. Power laws, Pareto distributions and Zipf’s law. *Contemporary Physics*, **46**, pages 323–351, 2005.
- [100] P. Bernaola-Galván, P. Carpena, R. Román-Roldán y J. Oliver. Study of statistical correlations in dna sequences. *Gene* **300**, pages 105–115, 2002.
- [101] Bernaola-Galván, J. Oliver y R. Román-Roldán. Decomposition of DNA sequence complexity. *Phys. Rev. Lett.* **83**, pages 3336–3339, 1999.
- [102] R. Román-Roldán, P. Bernaola-Galván y J. Oliver. Sequence compositional complexity of DNA through an entropic segmentation method. *Phys. Rev. Lett.* **80**, pages 1344–1347, 1998.
- [103] W. Li. The complexity of DNA. *Complexity* **3(2)**, pages 33–37, 1997.
- [104] H. Herzel y I. Große. Correlations in DNA sequences: The role of protein coding segments. *Phys. Rev. E* **55**, pages 800–810, 1997.
- [105] B. Lewin. *Genes VI*. Oxford University Press, 1987.
- [106] M.A. Huynen y P. Bork. Measuring genome evolution. *Proc. Natl. Acad. Sci. USA* **95**, pages 5849–5856, 1998.

- [107] C.K. Peng, S.V Buldyrev, A.L. Goldberger, S. Havlin, F. Sciortino, M. Simons, H.E. Stanley. Long-range correlations in nucleotide sequences. *Nature* **356**, pages 168–170, 1992.
- [108] J. Oliver. Curso de verano, Modelos matemáticos en Biología: un viaje de ida y vuelta. *Resúmenes*, 2005.
- [109] J Felsenstein. PHYLIP (Phylogeny Inference Package).
- [110] X. Chen, S. Kwong y M. Li. A compression algorithm for DNA sequences and its applications in genome comparison. *R. Shamir, S. Miyano, S. Istrail, P. Pevzner y M. Waterman (eds). Proc. of the Fourth Annual International Conference on Computational Molecular Biology (RECOMB). Tokio, Japan*, pages 131–142, 2000.
- [111] Y. Cao, A. Janke, P.J. Waddell, M. Westerman, O. Takenaka, S. Murata, N. Okada, S Pääbo y M. Hasegawa. Conflict Among Individual Mitochondrial Proteins in Resolving the Phylogeny of Eutherian Orders. *J. Mol. Evol.* **47**, page 307322, 1998.
- [112] N. Saitou y M. Nei. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, pages 406–425, 1987.
- [113] A. Wheeler y W. Zurek. *Quantum Theory and Measurement*. Princeton University Press, Princeton, 1983.
- [114] C. Fuchs y J. van de Graaf. Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Trans. Inf. Theory*, **45**, pages 1216–1227, 1999.
- [115] D. Brady y L. Hughston. *Geometric Issues on the Foundations of Science*. Eds. S.A. Huggett, L. Mason, K. Tod, S.-T. Tsou and N. Woodhouse, Oxford University Press, 1990.
- [116] A. Bhattacharyya. On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Calcutta Math. Soc.*, **35**, pages 99–110, 1943.
- [117] J. Anandan y Y. Aharonov. Geometry of quantum evolution. *Phys. Rev. Lett.* **65**, pages 1697–1700, 1990.
- [118] G. Lindblad. Entropy, information and quantum measurements. *Comm. Math. Phys.* **33**, pages 305–322, 1973.
- [119] S. Abe. Nonadditive generalization of the quantum Kullback-Leibler divergence for measuring the degree of purification. *Phys. Rev. A* **68**, pages 032302–4, 2003.
- [120] A. Wehrl. General properties of entropy. *Rev. Mod. Phys.* **50** (2), pages 221–260, 1978.



- 
- [121] M.J. Donald. Math. proc. camb. phil. soc. **101**. page 363, 1987.
- [122] R.F. Werner. Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model. *Phys. Rev. A*, **40**, pages 4277–4281, 1989.
- [123] A.S. Holevo. *Probl. Inform. Transmission* **9**, page 177, 1973.
- [124] E. Schrödinger. Die gegenwertige situation in der quantenmechanik. *English translation: J.D. Trimmer, Proc. of the Am. Phil. Soc.*, **124**, pages 323–38, 1980.
- [125] V. Vedral, M. Plenio, M.A. Rippin, y P. Knight. Quantifying entanglement. *Phys. Rev. Lett.* **78**, pages 2275–2279, 1997.
- [126] C.H. Bennett. How to define complexity in physics, and why. pages 137–148, 1990.
- [127] B. Schumacher. Quantum coding. *Phys. Rev. A*, **51**, pages 2738–2747, 1995.
- [128] A.P. Majtey, P.W. Lamberti, M.T. Martin y A. Plastino. Wootters' distance revisited: a new distinguishability criterium. *Eur. Phys. J. D* **32**, pages 413–419, 2005.



# Apéndice A

## Palabras en secuencias

Una secuencia  $\mathcal{S}$  de longitud  $L$  se define como una cadena lineal de  $L$  símbolos que corresponden a algún alfabeto finito  $\mathcal{A}$  de longitud  $k$ . Un segmento de  $n$  símbolos, con  $n \leq L$ , se denomina una  $n$ -*tupla*. El conjunto  $W_n$  es el conjunto de todas las  $n$ -*tuplas* que se pueden extraer de  $\mathcal{S}$  y tiene  $R$  elementos

$$W_n = \{w_{n,1}, w_{n,2}, \dots, w_{n,R}\} \\ R = k^n \quad (\text{A.1})$$

Para identificar y contar computacionalmente las  $n$ -*tuplas* se desliza una ventana de longitud  $n$  a través de la secuencia, desde la posición 1 hasta la  $L - n + 1$ .

$$C_n^S = (c_{n,1}^S, \dots, c_{n,R}^S) \quad (\text{A.2})$$

Con esto se pueden calcular las frecuencias de las palabras,  $f_n^S$ , para estimar las probabilidades  $p_{n,i}^S$  de encontrar una palabra específica  $w_{n,i}^S$ , y a partir de éstas construir un vector de probabilidades de  $n$ -*tuplas*:

$$P_n^S = (P_{n,1}^S, P_{n,2}^S, \dots, P_{n,R}^S) \quad (\text{A.3})$$

Las frecuencias  $f_n^S$  se obtienen calculando la abundancia relativa de cada palabra.

De este modo, se puede asignar a cada secuencia un vector, y trabajar con este objeto matemático.

Veamos un ejemplo: si tomamos una secuencia de ADN, tenemos  $\mathcal{A} = \{A, C, G, T\}$ ,  $k = 4$ . Podemos estudiar estas secuencias por ejemplo con palabras de longitud  $n = 3$ , una palabra tipo sería la palabra o 3-*tupla*  $ACT$ . Para una secuencia  $\mathcal{S} = ATATAC$ , con  $L = 6$ , el vector  $P_3^S$  se estima utilizando las frecuencias relativas de todos los trinucleótidos. Las frecuencias se determinan deslizando una ventana de tres letras, hasta la posición  $L - n + 1 = 4$ :

$$\begin{aligned}W_3 &= \{ATA, TAT, TAT, TAC, AAA, \dots\} \\C_3^S &= (2, 1, 1, 0, \dots) \\f_3^S &= (0, 5, 0, 25, 0, 25, 0, \dots)\end{aligned}$$

Los vectores  $C_3^S$  y  $f_3^S$  tienen longitud  $R = 4^3 = 64$ , y las coordenadas nulas corresponden a palabras que no están en  $\mathcal{S}$ , en este caso, trinucleótidos ausentes.

# Apéndice B

## Complejidad máxima

Queremos que la medida de complejidad alcance el máximo para la secuencia que nosotros consideramos, la más compleja posible. Esto es, una secuencia con distribución de longitudes de segmentos que obedece una ley de potencia,

$$P(l) \sim \frac{1}{l^\mu}. \quad (\text{B.1})$$

con exponente  $\mu = 1$ .

La complejidad, definida como la entropía de la distribución de segmentos será máxima, cuando tengamos igual número de segmentos en todos los intervalos, para ello elegimos una partición particular del intervalo  $[1, L]$ . Se puede ver, que en el caso de tener  $n$  segmentos en cada intervalo, la complejidad alcanza un máximo local, pero si se toma una partición para la cual el número de segmentos por intervalo (de una secuencia que obedece una ley de potencia con exponente 1) es 1, la complejidad alcanza un máximo absoluto.

## Máximo local de la complejidad

Sea una secuencia  $\mathcal{S}$  de longitud  $L$  que después del proceso de segmentación presenta  $N_s$  segmentos cuya distribución de longitudes obedece una ley de potencia. Vamos a suponer que se obtienen  $n$  segmentos en cada uno de los  $J^*$  primeros intervalos. Evaluamos la complejidad de esta secuencia:

$$C_{\mathcal{S}} = H_{\mathcal{S}}[P] \quad (\text{B.2})$$

En este caso  $N_s = nJ^*$  y  $\tilde{p}_j = \frac{n}{N_s}$ ,  $j = 1, \dots, J^*$ . Luego,

$$C_{\mathcal{S}} = - \sum_{j=1}^{J^*} \tilde{p}_j \log(\tilde{p}_j) = \log(J^*) \quad (\text{B.3})$$

A partir de esta secuencia  $\mathcal{S}$  construimos una secuencia  $\mathcal{S}'$  de igual longitud y que resulta de cambiar 2 segmentos del intervalo superior,  $[A_{J^*-1}, A_{J^*})$  de la siguiente forma:

- 1 segmento aumenta su longitud hasta pasar al intervalo siguiente,  $[A_{J^*}, A_{J^*+1})$ .
- 1 segmento disminuye su longitud en igual cantidad (de tal forma que  $L' = L$ ) y queda en el mismo intervalo

Para que esto sea posible, basta que existan en  $[A_{J^*-1}, A_{J^*})$ , 2 segmentos en el semi-intervalo superior. Para el cálculo de las probabilidades necesarias para evaluar  $C_{S'}[P']$ , debemos tener en cuenta que:

1. Hay  $n$  segmentos en los  $J^* - 1$  primeros intervalos
2. Hay  $n - 1$  segmentos en  $[A_{J^*-1}, A_{J^*})$ ; y
3. 1 segmento en  $[A_{J^*}, A_{J^*+1})$

Entonces

$$C_{S'}[P'] = \log(J^*) + \frac{n}{N_s} H \left[ \frac{1}{n}, \frac{n-1}{n} \right] \quad (\text{B.4})$$

$$= C_S[P] + \frac{n}{N_s} H \left[ \frac{1}{n}, \frac{n-1}{n} \right] \quad (\text{B.5})$$

$$> C_S[P] \quad (\text{B.6})$$

## Máximo global

Pensemos ahora en una secuencia que tiene 1 segmento en los  $J^*$  primeros intervalos. Para esta secuencia,

$$C_S = \log(J^*) \quad (\text{B.7})$$

Veamos que no es posible encontrar otra secuencia  $S'$ , con  $L' = L$ , que tenga una complejidad mayor.

Para tener 1 segmento en el intervalo  $[A_{J^*}, A_{J^*+1})$ , se deben usar al menos dos segmentos de otros intervalos que quedarán vacíos. De esta forma habremos ocupado el intervalo  $J^* + 1$ , a costa de otros dos intervalos vacíos. Tendremos  $J^* - 1$  intervalos ocupados con un segmento cada uno.

$$C_{S'} = \log(J^* - 1) < C_S \quad (\text{B.8})$$

Si por el contrario, se tiene una secuencia  $S'$  con más segmentos que  $S$ , cortando algunos de los iniciales, se tendrá una distribución de longitudes con un número igual o menor de intervalos ocupados. En cualquier caso será  $C_{S'} < C_S$ .

Por lo tanto la complejidad máxima de una secuencia  $S$  de longitud  $L$  será:

$$C_S = \log(J_{max}^*(L)), \quad (\text{B.9})$$

donde  $J_{max}^*(L)$  es el número de primeros intervalos que contienen un segmento de una secuencia de longitud  $L = \sum_j^{J^*} l_j$ .

### Este máximo no es alcanzable siempre

La longitud  $L$  de una secuencia con los  $J^*$  primeros intervalos ocupados está comprendida entre los siguientes extremos:

- todos los segmentos tienen la longitud del extremo inferior del intervalo al que pertenecen:

$$L_{min}(J^*) = \sum_{j=1}^{J^*} A_j$$

- todos los segmentos tienen la longitud del extremo superior del intervalo que ocupan:

$$L_{max}(J^*) = \sum_{j=2}^{J^*+1} (A_j - 1)$$

$$L_{max}(J^*) = L_{min}(J^*) + A_{j+1} - (J^* + 1)$$

Con esto podemos calcular el intervalo de longitudes de secuencias, para las que no existe  $J^*_{max}(L)$ :

$$\Delta L(J^*) = L_{min}(J^* + 1) - L_{max}(J^*) = J^* + 1$$

Este intervalo es relativamente pequeño para  $J^*$  suficientemente grande.





# Índice de figuras

3.1.	$JS_q$ en función de $q$ , para distribuciones de probabilidad correspondientes a un alfabeto binario $P^{(1)} = (t, 1 - t)$ y $P^{(2)} = (1 - t, t)$ , para $t = 0.3, 0.6$ . Los pesos se tomaron $\pi^{(1)} = \pi^{(2)} = 1/2$ . . . . .	34
3.2.	$\langle JS_q \rangle$ en función de la posición del curso móvil para ambas elecciones de los pesos. En este caso se concatenaron 2 secuencias generadas con probabilidades, $P^{(1)} = (0.2, 0.8)$ y $P^{(2)} = (0.7, 0.3)$ , y con longitudes $L_1 = 500$ y $L_2 = 1000$ respectivamente. El proceso de segmentación se realizó utilizando el parámetro $q = 1$ . En ambos casos se detecta el punto de fusión de las secuencias. Las curvas corresponden a un promedio sobre 2000 secuencias. En el inserto se muestra el mismo comportamiento para $q = 0.2$ . . . . .	36
3.3.	La $\langle JS_q \rangle$ para diferentes valores de $q$ mayores que 1. Nuevamente se concatenaron dos secuencias con probabilidades $P^{(1)} = (0.2, 0.8)$ y $P^{(2)} = (0.7, 0.3)$ , y longitudes $L_1 = 500$ y $L_2 = 1000$ . Las curvas corresponden al promedio sobre 2000 secuencias. En el inserto se muestra la $\langle JS_q \rangle$ para valores de $q < 1$ . . . . .	37
3.4.	$F_q$ en función de $\zeta$ , para valores fijos de $\psi$ y $\chi$ , y para $q = 2$ . . . . .	43
4.1.	$C_S$ en función del nivel de umbral $D_u$ para dos secuencias genómicas y una secuencia aleatoria generada artificialmente con idéntica composición que la ECO110k. Los círculos llenos corresponden a la secuencia ECO110k (111408 pares de bases, bp), los círculos vacíos corresponden a la secuencia de ADN humano HUMCTRADCV (98308 bp), y los triángulos corresponden a la secuencia aleatoria. El proceso de segmentación se realizó utilizando el alfabeto normal (alfabeto de 4 símbolos) para describir la secuencia. . . . .	57
4.2.	Perfiles de complejidad para las secuencias ECO110k ( $L_{ECO} = 111408$ bp) y HUMHBB ( $L_{HUM} = 73308$ bp). Los símbolos llenos corresponden a la complejidad para las secuencias completas, y los símbolos vacíos corresponden a la suma pesada de las complejidades para dos subsecuencias arbitrarias de cada una de las secuencias. Las subsecuencias fueron elegidas de forma tal que su concatenación fuera igual a la secuencia completa ( $L_{E1} = 57120$ bp y $L_{E2} = 54288$ bp; $L_{H1} = 42720$ bp y $L_{H2} = 30588$ bp). . . . .	58
4.3.	$C_S$ para las regiones codificada y no codificada del cromosoma humano 22. . . . .	59

4.4. Perfiles de complejidad para la cadena de genes pesados de miosina para diferentes especies (longitud total, porcentaje de intrones): Humano (28438bp, 74%), Rata (25759bp, 77%), Pollo (31111bp, 74%), Mosca del vinagre (22663bp, 66%), Brugia (11766bp, 32%), Acathamoeba (5894bp, 10%), Caenorhabditis (10780bp, 14%), Levadura (6108bp, 0%) . . . . .	61
5.1. ADN mitocondrial (ADNmt) del humano. (Imagen tomada de Brown. Genomes2)	66
5.2. Los tres árboles posibles entre primates, ferungulados y roedores. . . . .	68
5.3. Árbol con los genomas mitocondriales completos. . . . .	71
5.4. Árbol con los genes sin modificar y los intergenes desordenados. . . . .	72
5.5. Árbol con los símbolos correspondientes a los genes desordenados, se puede observar que la información filogenética ha sido destruida. . . . .	73
5.6. Matriz distancia entre los diferentes árboles. . . . .	73
6.1. $d_M^{JS}$ y $\frac{(d_M^W)^2}{2}$ para $P^{(1)} \sim P^{(2)}$ . Se eligen $P^{(1)} = (t, 1 - t)$ y $P^{(2)} = (x, 1 - x)$ y se evalúan las correspondientes distancias en función de $x$ para $t = 0.5$ . . . . .	86
6.2. $\sqrt{2d_{\mathcal{H}}^{JS, \tilde{\phi}}(\tilde{p}^{(2)}, \tilde{p}^{(1)})}$ en función de $\theta$ para $\varphi = 0,5$ y $\varphi = 0,8$ . . . . .	88
6.3. Gráfico 3D de $\sqrt{2d_{\mathcal{H}}^{JS, \tilde{\phi}}(\tilde{p}^{(2)}, \tilde{p}^{(1)})}$ en función de $\theta$ y $\varphi$ . Se puede apreciar claramente que esta función está acotada por el plano $z = \varphi$ . . . . .	88
6.4. Los objetos se hacen menos distinguibles cuando se dispone de información parcial sobre ellos. . . . .	91
6.5. Interpretación gráfica de las condiciones sobre la distancia $D$ ; ver texto para detalles . . . . .	100

# Publicaciones

- [M1] Imágenes Selectivas Rápidas por Resonancia Cuadrupolar Nuclear. E.P. Danieli, A.P. Majtey, J.F. Perlo, F. Casanova y D. Pusiol, *Anales AFA* **11**, 151 (1999).
- [M2] Pseudo-Quasielastic Component in the Neutron Scattering Cross Section. C.A. Condat, A.P. Majtey y S.A. Menchón, *Phys. Rev. B* **68**, 092301-4 (2003).  
Este trabajo fue seleccionado para su publicación en el Virtual Journal of Biological Physics Research, 15 de septiembre de 2003.
- [M3] Non Logarithmic Jensen-Shannon Divergence. P.W. Lamberti y A.P. Majtey, *Physica A* **39**, 81-90 (2003).  
Ver capítulos 2 y 3.
- [M4] A monoparametric family of metrics for statistical mechanics. A.P. Majtey, P.W. Lamberti y A. Plastino. *Physica A* **344**, 547-553 (2004).  
Ver capítulo 3.
- [M5] Ubiquity of metastable-to-stable crossover in weakly chaotic dynamical systems. F. Baldovin, L.G. Moyano, A.P. Majtey, A. Robledo y C. Tsallis. *Physica A* **340**, 205-218 (2004).
- [M6] Wootters' distance revisited: a new distinguishability criterium. A.P. Majtey, P.W. Lamberti, M.T. Martín y A. Plastino. *Eur. Phys. J. D* **32**, 413-419 (2005).  
Ver capítulo 6.
- [M7] Jensen-Shannon divergence as a measure of distinguishability between mixed quantum states. A.P. Majtey, P.W. Lamberti y D.P. Prato. *Phys. Rev. A* **72**(1), 052310-15 (2005).  
Este trabajo fue seleccionado para ser publicado en el Virtual Journal of Quantum Information, noviembre de 2005.  
Ver capítulo 6.
- [M8] Weak chaos in large conservative system infinite-range coupled standard maps. L.G. Moyano, A.P. Majtey y C. Tsallis. COMPLEXITY, METASTABILITY AND NONEXTENSIVITY, Proceedings of the 31st Workshop of the International School of Solid State Physics Erice, Sicily, Italy, 20-26 de julio de 2004.

- [M9] Weak chaos and metastability in a symplectic system of many long-range-coupled standard maps. L.G. Moyano, A.P. Majtey y C. Tsallis. *Eur. Phys. J. B* **52**, 493-500 (2006).
- [M10] Evaluación de la complejidad de secuencias genómicas. A.P. Majtey, R. Román-Roldán y P.W. Lamberti. *Actas de la Academia de Ciencias. Biomat* (2005).  
Ver capítulo 5.
- [M11] A complexity measure for symbolic sequences and applications to DNA. A.P. Majtey, R. Román-Roldán y P.W. Lamberti. Enviado a *Phys. Rev. E* (2006).  
Ver capítulo 5.

# Agradecimientos

No recuerdo el día en que el aula 27 se transformó en “el aula de la suerte”. Ahí discutimos, planteamos problemas y metas, surgieron ideas, y pocas veces salimos vencidos por alguna cuenta que los dos enfrentamos tiza en mano, intercambiando nuestros roles, unas veces alumno otras maestro. A Walter le agradezco todo lo que me enseñó. Le agradezco el brindarme con coraje la posibilidad maravillosa de trabajar en cosas nuevas, nuevas para ambos; el dejarnos seducir por el laberinto del conocimiento y siempre encontrar, más tarde o más temprano, alguna salida. Le agradezco la curiosidad que me contagió, la paciencia que me tuvo, su tiempo inagotable, el -sí, pasé -cada que vez que golpeé su puerta.

A lo largo de estos años, como un *árbol* que estira sus raíces dentro de la tierra para alimentarse, fui conociendo gente que de una u otra manera hoy son parte de esta tesis.

Gracias a Ramón Román-Roldán y a Constantino Tsallis por sus valiosas discusiones, las ideas, los consejos, su hospitalidad cada vez que me recibieron. Gracias a Carlos Condat, Ángel Plastino y Domingo Prato con quienes tuve la posibilidad de trabajar en estos años. Fue muy importante poder interactuar con todos ellos que me transmitieron su calidez, su experiencia y su pasión.

Cómo hubiese podido sobrevivir a las siestas infranqueables de no haber sido por los mates de mis compañeros de la 324, y también de los colados. Gracias por los mates, las galletitas, de salvado, obvio, las charlas, las consultas, las tortas y los criollos, las cervecitas, en fin, gracias por el día a día.

Gracias a Rio por Luis y Pablo y la eterna manía de cuestionarlo todo.

Gracias a mi mamá por cuidar mi semilla y celebrar conmigo cada brote.

A mi hermana por ser sol que me abraza cada día.

A mis amigas del alma por apuntalarme después de alguna tormenta y estar a mi lado siempre.

Gracias a Damián por ser el viento que mece mis hojas, el pájaro que anida en mis ramas, la savia que nutre mi tronco.

Gracias Damián, por ser el bosque en el que habito.