

Facultad de Astronomía, Matemática y Física

Caracterización de Secuencias Simbólicas

Trabajo especial de Licenciatura en Física

Leonardo Esteban Riveaud

Director: Pedro Walter Lamberti

CORDOBA-ARGENTINA

-2015-



Caracterización de Secuencias Simbólicas por Leonardo Esteban Riveaud se distribuye bajo una [Licencia Creative Commons Atribución 2.5 Argentina](https://creativecommons.org/licenses/by/2.5/argentina/).

Palabras Claves

Teoría de la Información

Análisis de secuencias

Divergencia de Jensen-Shannon

Agradecimientos

Los caminos largos y empinados se suelen no recorrer en soledad. El más sincero agradecimiento al profe Walter Lamberti por permitirme aprender y trabajar con él en el transcurso de estos dos últimos años y en los cinco que vendrán. A Diego Mateos por prestarme su ayuda siempre. Al profe Bustos por la paciencia en cada consulta que le hacía. Al profe Budde por sus consejos y charlas sobre Prigogine que quedarán en la memoria por siempre. Los agradecimientos a mis familiares y amigos son extricatamente personales.

Resumen

Existe una clara relación entre sistemas dinámicos y secuencias simbólicas. Esto ha conducido tanto al desarrollo de modos de asignación de la evolución de un sistema dinámico a cadenas simbólicas, como a la introducción de cantidades que describan esa dinámica. Un ejemplo de esto último es el método utilizado por Yang, Yien y Hseu [1]. En el presente trabajo estudiaremos un método mixto, basado en el mapeo utilizado por Yang y colaboradores, y en la Divergencia de Jensen Shannon, para caracterizar y diferenciar cadenas de caracteres con distribución de símbolos no estacionaria. Este método se usará para investigar la diferencia entre señales caóticas y ruidosas, representadas matemáticamente por series temporales. El método se aplica a distintos tipos de ruidos y mapas caóticos, en los cuales se puede observar la robustez del método propuesto. También, y por último, se compararán nuestros resultados con los obtenidos por medio de la propuesta de distancia de Yang y colaboradores [1].

Caracterización de Secuencias Simbólicas

L. E. Riveaud¹,

1 L. E. Riveaud FaMAF, Córdoba, Argentina

* E-mail: leoriveaud@gmail.com

Contenidos

1. INTRODUCCIÓN	4
2. CONCEPTOS BÁSICOS	6
2.1. Breve marco teórico para el análisis de Series Temporales	6
2.1.1. Proceso Estocastico	6
2.1.2. Consecuencias estadísticas de tener una sola secuencia de muestra .	13
3. ELEMENTOS DE TEORIA DE LA INFORMACION	15
3.1. Entropía	15
3.1.1. Entropía de Información	16
3.2. Relaciones de Similitud entre Cadenas Simbólicas	19
3.2.1. Divergencia de Kullback-Leibler	19
3.2.2. Divergencia Jensen-Shannon	22
4. ASPECTOS METODOLOGICOS	26
4.1. Correspondencia entre Serie Temporal y cadena simbolica binaria	26
4.2. Método de Análisis	28
4.2.1. Algoritmo de “La Ventana Móvil”: El intento de entender una estructura en base a una estadística	30
4.2.2. ¿A qué nos referimos con cambios significativos?; el significado de la Significancia	34
4.3. Medida de disimilitud propuesta por Yang y colaboradores	35
5. CAOS, RUIDO Y EL TEOREMA DE WOLD	37
5.1. Visión del determinismo y alatoreidad desde el proceso estocástico, Teorema de Wold	39
5.1.1. Proceso completamente aleatorio: Ruido Blanco	39
5.2. Propiedades de un proceso Random y su diferencia con un procesos Caotico, mas específico	42
5.3. Sistemas Caóticos	42
5.3.1. Tent-Map	47
5.3.2. Mapa Logístico	49
5.4. Ruido Rosa	52

6. RESULTADOS PROPIOS	55
6.1. Diferencias entre caos y ruido con JSD	55
6.1.1. Ruido Blanco y Tent-Map	56
6.1.2. Ruido Rosa y Tent-Map	57
6.1.3. Ruido Blanco y Ruido Rosa	59
6.1.4. Ruido Blanco y Mapa Logístico	61
6.1.5. Tent-Map y Mapa Logístico	61
6.1.6. Ruido Rosa y Mapa Logístico	62
6.2. Matriz Distancia	64
6.2.1. Otra propuesta de distancia	65
7. CONCLUSIONES	66
8. Apéndice A: Repaso de Teoría de Probabilidades	68
9. Apéndice B: Convergencias y Desigualdades	73
10. Apéndice C: Algunas cuestiones sobre Teoría de la Información	76
10.1. Entropía Condicional	77
10.2. Desigualdad de Jensen “finita”	79
11. Apéndice D: Proyecciones entre Procesos Estocásticos	80
11.1. Nuestro espacio de Hilbert	80
11.2. Teorema de Proyección	82
12. Apéndice E: Representación Espectral	84
13. Apéndice F: Código fuente del algoritmo de la “Ventana Móvil”	86
14. Apéndice G: Código fuente del algoritmo de D_{YHY}	91
15. Apéndice H: Código fuente para el cálculo del ruido Rosa	104

1. INTRODUCCIÓN

La distinción entre procesos estocásticos (PE) y sistemas caóticos (SC) ha recibido atención por parte de los investigadores dedicados al estudio de las series temporales.

Aunque muy distintas en su origen físico, las series temporales provenientes de SC comparten con las generadas por los PE varias propiedades que las hacen prácticamente indistinguibles:

- Un amplio espectro de potencias
- Una función correlación deltiforme
- Un comportamiento irregular de las señales medidas

Esta lista no agota las características similares entre un PE y los SC. Estas similitudes hacen posible que, en muchas aplicaciones técnicas, los SC sean reemplazados por PE y viceversa. Por otro lado las series temporales con estructuras complejas aparecen con frecuencia tanto en sistemas naturales como artificiales. Por ello el interés detrás de esta distinción reside en develar las causas de impredecibilidad que gobierna esos sistemas.

En este sentido hubo intentos de diferenciación de ambos sistemas a través de sus propiedades dinámicas. En el caso de sistemas caóticos se observaron dimensiones fractales finitas (no enteras) mientras que para un PE, esa dimensión es infinita. Osborne y Provenzale [11] observaron para una no-convergencia en la dimensión de correlación (como una estimación de la dimensión fractal). Ellos mostraron que las series temporales generadas por espectros de leyes de potencia inversa y fases aleatorias son caminos fractales aleatorios con dimensión de Hausdorff finita, y consecuentemente, con dimensión de correlación finita. Varios otros trabajos han estado dedicados a esta problemática. Basta mencionar el trabajo de Sugihara y May [12] basado en una previsión no-lineal en el cual ellos comparan trayectorias predichas con las verdaderas y hacen una distinción tentativa entre caos dinámico y medición de error.

Otras referencias relevantes para esta temática son los trabajos de Kaplan y Glass [13], Kantz y colaboradores [14], etc.

Los SC muestran “sensitividad a las condiciones iniciales”, lo cual manifiesta inestabilidad generalizada en el espacio de las fases. Esto indica falta de predictibilidad más allá del carácter determinístico de la trayectoria temporal. En un sistema caótico dos puntos vecinos en el espacio de las fases se apartan exponencialmente rápido (exponente de Lyapunov). Por ello hay valores del tiempo en que ambos puntos son inexorablemente distinguibles. La noción de distinguibilidad está asociada a la noción de información y a su vez esta, a la noción de entropía. Es detrás de este simple esquema que aparece el concepto de entropía de Kolmogorov y Sinai [15].

Una propuesta sumamente interesante la han realizado recientemente O. Rosso y colaboradores en la cual hacen uso de una medida de complejidad [17]. Esta medida es una

modificación de una propuesta de Lopez-Ruiz, Mancini y Calbet [18], la cual contiene una medida de desequilibrio cuantificada a través de una distancia entre las distribuciones de probabilidad.

Nuestra tesis pretende encontrar una manera de distinguir los dos tipos de procesos (PE y SC), mediante la Teoría de la Información. La manera de abordar la problemática será estudiar las series temporales que producen estos sistemas. Para poder implementar las herramientas de la Teoría de la Información será necesario mapear estas series temporales a secuencias simbólicas, mediante el método utilizado por Yang y colaboradores [1]. Para simplicidad y síntesis del trabajo solo se trabajarán con cuatro series temporales, dos caóticas y dos aleatorias. Las series caóticas elegidas fueron el Mapa Logístico y el Temp-Map, y las aleatorias el Ruido Rosa y el Ruido Blanco. Sus propiedades se detallarán en una sección exclusiva para poder tener una visión adecuada de los resultados obtenidos.

Teniendo ya las secuencias simbólicas, tomaremos dos caminos diferentes, aunque relacionados. Primero intentaremos distinguir, en base a la Divergenia Jensen-Shannon, cuando una cadena simbólica presenta características caóticas o aleatorias. Explícitamente lo que haremos es pegar dos secuencias simbólicas y mediante un algoritmo distinguir la transición de una secuencia a otra. Los detalles del programa se darán en la sección 4 (“Aspectos Metodológicos”).

El segundo objetivo será formar una “Matriz Distancia” entre las series mencionadas. Es decir, calcular la distancia entre cada una de las series dadas. La misma, la construiremos con dos funciones diferentes, la Divergenia Jensen-Shannon y la distancia de Yang y Yein [1].

2. CONCEPTOS BÁSICOS

2.1. Breve marco teórico para el análisis de Series Temporales

Una serie temporal es un conjunto de valores observados (experimentalmente o generados por algún algoritmo) de una dada variable, x_t , representando t el tiempo en que ese valor se observa. Diremos que la serie temporal es discreta si el conjunto de valores de tiempos t , es discreto. En los últimos años ha crecido notablemente el número de métodos de estudio de las propiedades estadísticas de las series temporales. Ellos se originan en conceptos clásicos de la estadística, en la teoría de la información, y otras áreas de la matemática. A su vez la importancia de las series temporales es que una gran variedad de fenómenos naturales pueden representarse por medio de ellas. Secuencias genéticas, registros electrofisiológicos (EEG y ECG), dinámica de motores moleculares, modelos climáticos, etc., son claros ejemplos de esta presencia ubicua y de la importancia de las series temporales, en las ciencias naturales. Desde un punto de vista formal, una serie temporal se puede representar como un proceso estocástico. Es por ello que a continuación repasaremos las principales propiedades de este tipo de procesos.

2.1.1. Proceso Estocástico

Al ser los valores x_t de carácter impredecible es natural que los tratemos como si fueran una variable aleatoria (VA) X_t (ver Apéndice A, para una correcta introducción a los conceptos de teoría de probabilidades). Se define como Proceso Estocástico a la familia de variables aleatorias $\{X_t; \forall t \in T\}$; definas en un espacio de probabilidad $(\Omega, \psi, \mathbb{P})$. El parámetro t no pertenece al espacio muestral, es decir, no es un evento. Esto significa que para cada t uno tendrá una variable aleatoria $X_t : \Omega \rightarrow \mathfrak{R}$; que vista como una aplicación es $X : T \times \Omega \rightarrow \mathfrak{R}$.

Uno puede ver a la serie temporal $\{x_1, x_2, \dots, x_n\}$ como una realización “particular” finita del proceso estocástico $\{X_t; t \in T\}$. Viéndolo así sería :

$$X_{-\infty}, \dots, X_{-1}, \underbrace{X_0, \dots, X_n}_{x_0, \dots, x_n}, X_{n+1} \dots \quad (1)$$

La sucesión de arriba es un proceo estocástico, y la de abajo una serie temporal, el hecho que tenga las letras en minúsculas es porque son un valor específico (particular), dentro del rango de la variable aleatoria, que ha tomado esta.

A continuación se darán unos ejemplos para visualizar mejor los conceptos a exponer. En el primero de ellos se muestra claramente el concepto de VA, pero en el segundo no; es este el que merece mayor atención para esta tesis.

Ejemplo 1: *Función periódica con amplitud y defasaje aleatorios*

Sea A y Θ dos VA. EL rango de A cumple con $Rango(A) \geq 0$ y el rango de Θ es distribuido uniformemente en $[0, 2\pi]$. Podemos definir un proceso estocástico $\{X_t\}$ de la forma

$$X_t(\omega) = A(\omega)\cos(t + \Theta(\omega)); \text{ con } \omega \in \Omega, t \in \mathbb{T} \quad (2)$$

Se puede ver claramente que toma elementos del conjunto T y del espacio muestral Ω

Ejemplo 2: *Un proceso binario de igual probabilidad*

Sea una secuencia construida a base de un conjunto de VA independientes; $X_t; t = 1; 2; 3; \dots$, donde la probabilidad para cada t es

$$P(X_t = 1) = P(X_t = 0) = 1/2 \quad (3)$$

En este caso no es para nada obio que exista un espacio probabilístico $(\Omega, \psi, \mathbb{P})$ en donde el proceso estocástico este definido, es más tampoco podemos decir que está definida la distribución de probabilidad conjunta

$$P(X_1 = x_1; \dots; X_n = x_n) = 2^{-n} \quad (4)$$

donde $\{x_i; i = 1; \dots, n\}$ tomarán el valor 1 o 0.

El segundo ejemplo es un claro caso en el cual uno no esta provisto del proceso estocástico en si ($\{X_t; t \in \mathbb{T}\}$) si no de su probabilidad para cada t . Lo que buscamos es algo que nos garantice que ese espacio probabilístico va a existir. Esto es común en muchas aplicaciones de procesos estocásticos, es más hay algunos textos que se contentan con solo verificar que existe el espacio probabilístico y nunca lo muestran explícitamente, solo trabajan con sus distribuciones de probabilidad. Para poder verificar la existencia de tal procesos en el ejemplo 2 necesitaremos de la ayuda de un teorema propuesto por Kolmogorov a principios del siglo XX. Este importantísimo teorema lo describimos en el Apéndice A. También se dará algunas definiciones y conceptos necesarios para poder interpretar correctamente el teorema.

Dentro de la variedad de procesos estocásticos posibles existen un grupo que le prestaremos especial atención. Estos procesos son los llamados Estacionarios y Estrictamente Estacionarios. Estos procesos son de vital importancia ya que los mayores reultados generales de series temporales (como el teorema de descomposición de Wold) se cumplen solo para este tipos de procesos. Los resultados para procesos que no sean Estacionarios no son generales (en la mayoría de los casos) y dependen del modelo que uno proponga.

Para poder defenir este tipo especial de procesos estocásticos necesitaremos de la definición de la Covarianza y de la Autocovarianza.

Definición: COVARIANZA

Sean X e Y dos variables aleatorias definidas en el mismo espacio probabilístico $(\Omega, \psi, \mathbb{P})$, se define como covarianza a;

$$Cov(X; Y) \doteq E[(X - E(X))(Y - E(Y))] \quad (5)$$

donde $E(X)$ es el primer momento de la variable aleatoria X (ver la definición en el Apéndice A). Teniendo esto podemos definir la Autocovarianza de un proceso estocástico.

Definición: AUTOCOVARIANZA

Sea $\{X_t; t \in T\}$ un proceso estocástico definido en un espacio probabilístico $(\Omega, \psi, \mathbb{P})$, tal que $Var(X_t) < \infty$ para cada $t \in T$, se define entonces a la función autocovarianza $\gamma_{X_t}(\cdot)$ del proceso X_t como

$$\gamma_{X_t}(h) = Cov(X_t; X_{t+h}) = E[(X_t - E(X_t))(X_{t+h} - E(X_{t+h}))] \quad (6)$$

Dado esto, podemos definir lo que es un Proceso Estacionario;

Definición: PROCESO ESTACIONARIO

Se le dice Estacionario al Proceso Estocástico $\{X_t; t \in T\}$, con $T = 0, \pm 1, \pm 2, \dots$ cuando satisface las siguientes propiedades;

1. $E|X_t|^2 < \infty; \forall t \in \mathbb{Z}$
2. $E(X_t) = \mu$
3. $\gamma_{X_t}(h) = \gamma_{X_t}(t+h); \forall t, h \in T$

Definición: PROCESO ESTRICTAMENTE ESTACIONARIO

Llamaremos de esta manera a los procesos $\{X_t; t \in T\}$ que satisfagan ;

$$P(X_{t_1}, \dots, X_{t_k}) = P(X_{t_1+h}, \dots, X_{t_k+h}) \quad (7)$$

para todo entero positivo k y $\forall t_1, \dots, t_k, h \in T$.

Se puede ver fácilmente que si un proceso estrictamente estacionario tiene segundo momento finito (una de las condiciones de estacionariedad) y primer momento, entonces también Estacionario. No es necesario recurrir a contraejemplos muy complicados para demostrar que la recíproca no se cumple, es decir, estacionario no implica generalmente estrictamente estacionario.

Volvamos al Ejemplo 2; pero con una distribución de probabilidad distinta. Es decir, supongamos que no disponemos del proceso estocástico, pero si conocemos su distribución de probabilidad. Intentemos hacer un experimento conceptual. Supongamos que nos dan la siguiente distribución de probabilidad;

$$F(X_t = 1) = \frac{1}{t};$$

$$F(X_t = 0) = 1 - \frac{1}{t};$$

Vemos claramente que el proceso subyacente no es estrictamente estacionario ya que la distribución de probabilidad dependen del tiempo. Primero nos aseguramos que la distribución dada es precisamente un medida de probabilidad para eso verificamos que se cumplan las condiciones propuestas en el Apéndice A; vemos entonces que dado un espacio muestral Ω ; tenemos $\Omega \times T = \{X_t = 0; X_t = 1\} ; \forall t \in T$; donde $T = \mathbb{Z} \geq 1$

1. $F(X_t = 0; X_t = 1) = F(X_t = 0) + F(X_t = 1)$
2. $F(\Omega) = F(\{X_t = 0; X_t = 1\}) = F(X_t = 0) + F(X_t = 1) = 1$; ya que los eventos son disjuntos.
3. $0 \leq F(Y) \leq 1$, donde Y representa cualquiera de los dos eventos, sea cara ($X_t = 0$) o cruz ($X_t = 1$)

Teniendo estas probabilidades se podría construir las cadenas simbólicas, que no serán otra cosa que cadenas de ceros y unos, un ejemplo de una de ellas podría ser:

$$1110110111000111... \quad (8)$$

En un análisis informal se podemos que irán apareciendo cada vez menos 1 a lo largo de la cadena. Vemos por como están definida la distribución de probabilidad la cadena siempre tiene que empezar con el caracter 1 ya que $P(X_1 = 1) = 1$ y $P(X_1 = 0) = 0$.

Dado el teorema de Kolmogorov podemos asegurar que el espacio muestral existe y que la probabilidad conjunta $P(X_1, \dots, X_n)$ existe. Al ser cada variable aleatoria independiente una a una, significa que la probabilidad conjunta será

$$P(X_1, \dots, X_n) = \prod_{k=1}^{k=n} P(X_k) \quad (9)$$

obiamente dependerá del parámetro t .

Un ejemplo más simple pero muy útil también es el de un proceso estrictamente estacionario, es decir, supongamos que nos dan a una distribución de probabilidad de la forma

$$F(X_t = 0) = \alpha; \forall t \in T$$

$$F(X_t = 1) = \beta; \forall t \in T$$

donde α y β con constantes que no dependen del parámetro t .

El hecho de que nuestra F tenga que cumplir con la condición de medida de probabilidad nos exige que α y β sean positivas y que cumplan con $\alpha + \beta = 1$ para cada $t \in \mathbb{T}$. Se puede ver que en este caso las cadenas pueden empezar tanto con 0 como con 1. Un ejemplo clásico de lo dicho recién es el experimento de tirar monedas.

Ejemplo *El experimento teórico de tirar una moneda*

Supongamos que tenemos una moneda la cual tiene de una lado un 1 y del otro un 0, y empezamos a arrojarla hacia arriba y observar de que lado cae. Sabemos, porque alguien nos lo dijo (hipótesis), que la probabilidad de que salga un 1 es r y la que salga un 0 es q , tal que $r + q = 1$. Consideremos el lanzamiento simultaneo (pero independientes uno de los otros) de n monedas. La probabilidad de que salgan k veces el 1 sin importar el orden (es equivalente a pedir que salgan $n - k$ veces el 0 sin importar el orden) tendríamos que encontrar la probabilidad conjunta de la unión de los eventos de que en una cierta cadena (es decir un orden determinado) ocurra k veces el 1. Es decir, cada orden determinado (con k veces 1) es un evento distinto, y si nosotros buscamos la probabilidad conjunta de la unión de todos estos eventos y por definición de medida de probabilidad, es la suma de las probabilidades. Esto nos dice que la probabilidad conjunta de todas las cadenas posibles con k veces uno, es la suma de cada probabilidad conjunta referida a cada cadena en particular. En total existen $\binom{n}{k}$ cadenas distintas con k veces 1, ya que son las distintas maneras de colocar k objetos en n cajas. Por lo tanto la probabilidad conjunta total que buscamos es;

$$P = \binom{n}{k} p(X_1, \dots, X_n)$$

donde $p(X_1, \dots, X_n)$ es la probabilidad conjunta de que ocurra una cadena (con k veces 1) con un orden determinado. Pero como cada moneda es independiente una de otra, esta probabilidad se transforma en $p(X_1, \dots, X_n) = r^k q^{n-k}$, entonces la distribución de probabilidad del experimento de tirar monedas no es más que ;

$$P = \binom{n}{k} r^k q^{n-k}$$

Se puede calcular facilmente utilizando la función característica, la media, la varianza y la desviación estandar de esta distribución, pero nosotros mostraremos el resultado solamente, y así para poder dedicarle más tiempo a los teoremas del apartado siguiente. El primer momento de la variable aleatoria $x_n = \frac{k}{n}$ definida como la frecuencia relativa de aparición del caracter 1 es

$$M \doteq E[x_n] = E \left[\frac{k}{n} \right] = r \quad (10)$$

la varianza y la desviación estandar son;

$$D^2[x_n] = \frac{rq}{n}; \quad (11)$$

$$\sigma \doteq \sqrt{D^2} = \sqrt{\frac{rq}{n}}. \quad (12)$$

Este planteo es equivalente al hecho en el Apendice A con la Variable Aleatoria Indicadora. Desde ahora en más se usarán los dos de forma indistinta. En los dos casos hemos supuesto que nuestra cadena es infinita. Obviamente esto no es así en las aplicaciones, como es el caso de nuestra tesis, pero se podrá aproximar de buena manera teniendo en cuenta ciertos teoremas limites, como se verá mas adelante.

Intentemos ahora resolver el problema de forma inversa, es decir, la cadena (obviamente finita) y deseamos obtener la distribución de probabilidad y el proceso estocástico. Este no es un problema menor, ya que debemos nosotros mismos decidir cual es una medida de probabilidad adecuada.

La teoría de la medida nos provee de las condiciones que tiene que satisfacer una distribución de probabilidad para que esta sea una medida. Pero no nos brinda la distribución de probabilidad propiamente dicha, es decir, explícitamente hablando. Esto en la práctica es un problema ya que trabajar en abstracto imposibilitaría la aplicación de la teoría. El problema fue resuelto por Bernulli hace mucho tiempo. Gracias al teorema que lleva su nombre, uno puede aproximar la medida de probabilidad con la Frecuencia Relativa del evento. Este es un logro importante en la Teoría de Probabilidad ya que nos permite relacionar la medida de probabilidad con el concepto intuitivo de probabilidad, la frecuencia relativa.

El teorema de Bernulli es un caso particular de la Ley débil de los grandes números. Es por esto que primero comentaremos el resultado general para luego particularizar y obtener el resultado de Bernulli. En este apartado se utilizarán conceptos y desigualdades escritas en el Apéndice B.

Teorema: LEY DÉBIL DE LOS GRANDES NÚMEROS

Sea $\{X_1, X_2, \dots\}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas (i.i.d) ,es decir, variables aleatorias que tengan la misma distribución de probabilidad y sean independientes dos a dos; tales que $Var(X_i) < \infty$. Sea S_n ; con $n \geq 1$, una variable aleatoria definida como $S_n \doteq \sum_{i=1}^n X_i$. Entonces para cualquier $\varepsilon > 0$, se tiene;

$$\lim_{n \rightarrow +\infty} P \left(\left| \frac{S_n}{n} - E(X_i) \right| > \varepsilon \right) = 0 \quad (13)$$

Demostración: Usando la desigualdad de Chebyshev demostrada en el Apéndice B, y aplicandola a la variable aleatoria S_n . Usando la linealidad del operador $E(\cdot)$ y que las variables son i.i.d. tenemos que

$$E \left(\frac{S_n}{n} \right) = \frac{1}{n} E \left(\sum_{i=1}^n X_i \right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = E(X_1) \quad (14)$$

Pero como las variables aleatorias $\{X_1, X_2, \dots\}$ son independientes tenemos que

$$Var \left(\frac{S_n}{n} \right) = \frac{1}{n^2} Var \left(\sum_{i=1}^n X_i \right) = \frac{1}{n^2} \sum_{i=1}^n Var(X_i) = \frac{Var(X_1)}{n} \quad (15)$$

Luego teniendo en cuenta la “Pequeña desigualdad de Chevyshev”, nos queda que;

$$P\left(\left|\frac{S_n}{n} - E(X_1)\right| > \varepsilon\right) \leq \frac{Var(X_1)}{n\varepsilon_2} \quad (16)$$

Como $Var(X_1) < \infty$ el lado derecho de la desigualdad tiende a 0 cuando $n \rightarrow \infty$.

Una consecuencia importante del teorema anterior es cuando el conjunto de variables aleatorias que pertenecen a la sucesión son dicotómicas, es decir, pueden tomar dos valores. En otras palabras sería una sucesión de variables aleatorias indicadoras. Y su importancia radica en la interpretación de la probabilidad como frecuencia relativa de un evento “Éxito”. Primero daremos el concepto de Convergencia en probabilidad cero, que mantiene el espíritu a las convergencias definidas en el Apéndice B.

Definición: CONVERGENCIA EN PROBABILIDAD A CERO

Sea X_t una variable aleatoria definida en un espacio probailístico $(\Omega, \psi, \mathbb{P})$, se dice que converge en probabilidad a cero, si para todo $\varepsilon > 0$;

$$P(|X_n| > \varepsilon) \rightarrow 0 \text{ si } n \rightarrow \infty$$

y lo denotaremos como $X_n \xrightarrow{P} 0$.

Teorema BERNULLI

Estableciendo ya que las variables $\{X_j\}$ serán dicotómicas podemos ver que la variable aleatoria $\frac{S_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i$ toma otro significado, el de *Frecuencia Relativa*, ya que será la razón entre la cantidad de veces que ocurre el evento *éxito* sobre el total de ensayos realizados. Y como la variable X_1 es indicadora (o dicotómica) se fácilmente que

$$E(X_1) = 0 \cdot (1 - P(\text{éxito})) + 1 \cdot (P(\text{éxito})) = P(\text{éxito}) \quad (17)$$

y si lo traducimos con la notación que usamos para el experimento de monedas tenemos;

$$E(X_1) = 0 \cdot (1 - r) + 1 \cdot r = r \quad (18)$$

entonces nos queda que

$$P\left(\left|\frac{S_n}{n} - r\right| > \varepsilon\right) \leq \frac{r(1-r)}{n\varepsilon_2} \quad (19)$$

Y el miembro derecho tenderá a 0 cuando $n \rightarrow \infty$. Con esto vemos que la probabilidad de que la Frecuencia Relativa del evento *éxito* (o A) se desvíe de la probabilidad del evento *éxito*, $P(\text{éxito})$, en una cantidad mayor de la cantidad prefijada ε , se hace cada vez menor mientras aumenta n .

2.1.2. Consecuencias estadísticas de tener una sola secuencia de muestra

Aunque es posible realizar ciertos experimentos controlados (repetibles varias veces bajo idénticas condiciones de partida o iniciales) que permitan obtener varias realizaciones particulares del mismo proceso estocástico, en general, esto es casi imposible. Es decir, en la mayoría de los casos uno lleva a observar el proceso estocástico una única vez. Casi siempre la serie temporal se refiere a un período muestral que tan solo es una parte de la historia del proceso estocástico que subyace dicha serie. No obstante, si las circunstancias naturales del período muestral al que se refiere la serie observada se mantienen relativamente estables después de dicho período, entonces se espera que las conclusiones obtenidas del análisis de dicha serie sean aplicables también a momentos posteriores, al menos a corto plazo. Esta idea justifica el empleo de un modelo elaborado con una muestra dada para describir la evolución temporal de un proceso estocástico después del período muestral considerado.

En lo que sigue haremos un resumen de la estrategia utilizada en los casos de tener solo una secuencia pero de la perspectiva del Análisis de Series Temporales y no de la Teoría de la Información. Esto nos ayudará a entender porque elegimos la Teoría de la Información para la investigación en esta tesis.

En el Análisis de Series Temporales es de interés poder inferir el valor que tomará la variable aleatoria X_N dado los datos observados (x_1, \dots, x_{N-1}) . Cuando uno tiene una sola cadena se ve obligado hacer ciertas suposiciones tales como de estacionariedad y ergodicidad del procesos subyacente a la serie temporal.

Cuando el proceso es estacionario los parámetros $\mu, \gamma_0, \gamma_1, \gamma_2, \dots$ pueden estimarse a partir de una sola cadena de observaciones. Para una muestra x_0, \dots, x_n asociada al proceso estocástico $\{X_t\}$ los estimadores son

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=0}^n x_i \quad (20)$$

$$\hat{\gamma}_k = \frac{1}{n} \sum_{i=0}^{n-k} (x_i - \hat{\mu}_X)(x_{i+k} - \hat{\mu}_X) \quad (21)$$

Primero se ve que si el proceso no fuese estacionario la autocovarianza dependerá del tiempo, pero nosotros no conocemos esta dependencia, por lo tanto, esta suposición se amolda perfectamente a la situación. Si uno quisiese modelar procesos no estacionarios con una sola cadena de muestra las cosas se complican bastante, aunque en algunos casos uno puede transformalas en estacionarias imponiendo otras condiones. Nosotros no profundizaremos en estos aspectos.

La otra pregunta a hacerse es si los estimadores son consistentes con los valores reales de la media (μ) y de la autocovarianza (γ_k , para cualquier k). Esta pregunta se responde si imponemos la condición de ergodicidad en el procesos, es decir, exigir que

$$\gamma_k \rightarrow 0 \text{ cuando } k \rightarrow \infty \quad (22)$$

esto trae como consecuencia que el error con el que el estimador $\hat{\mu}_n$ difiere de la verdadera media μ sea muy pequeño cuando n crece, formalmente esto es que

$$E(\hat{\mu}_n - \mu)^2 \rightarrow 0 \quad (23)$$

es decir, $\hat{\mu}_n$ converge en media cuadrática a μ . Ya suponiendo todo esto, los estimadores de los párrafos anteriores son consistentes. Pero no nos sirve de nada si no podemos describir el proceso subyacente a la muestra. Este problema se soluciona imponiendo un modelo. Esto claramente nos hace perder generalidad pero si uno tiene cierto conocimiento teóricos de como puede ser el proceso podremos decir cual modelo es el conveniente. En este trabajo elegiremos a modo de ejemplo el modelo AR(p) (Autoregresivo de orden p).

Nuestro modelo AR(p) no permite escribir la variable aleatoria X_t como una combinación lineal de las variables aleatorias anteriores más una variable aleatoria que tiene como distribución la normal. Esto es;

$$X_t - \mu = Z_t + \sum_{i=1}^p \phi_i X_{t-i} \quad (24)$$

Para poder describir nuestro proceso necesitamos encontrar el vector estimado de $\Phi = (\phi_1, \dots, \phi_p)$ o sea $\hat{\Phi}$. Para eso utilizaremos la ecuación de Yule-Walker que dice que

$$\Gamma_p \Phi = \gamma_p \quad (25)$$

donde Γ_p es la matriz de covarianza $[\gamma_{(i-j)}]_{i,j=1}^p$ y $\gamma_p = (\gamma_1, \dots, \gamma_p)$. La estimación se hará resolviendo el siguiente sistema de ecuaciones

$$\hat{\Gamma}_p \hat{\Phi} = \hat{\gamma}_p \quad (26)$$

Nosotros no resolveremos este problema, solo es un bosquejo de como se encararía un problema desde la perspectiva del Análisis de Series temporales. En esta visión nosotros buscamos describir un procesos estocástico subyacente de la única cadena de muestra, e intentar inferir el futuro. Esto es de gran utilidad para aplicaciones financieras como la bolsa de comercio, o climáticas para la predicción de huracanes por ejemplo. Pero resulta difícil desde estas técnicas poder analizar los “cambios” o “estructura” de la cadena de muestra (serie temporal). Es por esta razón que nosotros elegimos a la Teoría de la Información como nuestra herramienta principal para nuestra investigación. De ahora en más no nos preocuparemos por encontrar el proceso estocástico subyacente, nos contentaremos con conseguiri estimar su distribución de probabilidad y así aplicar las herramientas de la teoría de la información.

3. ELEMENTOS DE TEORIA DE LA INFORMACION

En esta sección se presentarán algunos resultados importantes de la Teoría de la Información (TI). En el Apéndice C se complementará algunos de los tópicos aquí expuestos. Dos conceptos importantes que se tratarán por separado son la Entropía de Shannon y la Divergencia de Jensen-Shannon (D_{JS})

3.1. Entropía

El concepto de Entropía ha sido estudiado en diferentes áreas de la ciencia. A mediados del siglo XIX Rudolf Clausius estudió detalladamente la entropía desde la termodinámica. Entendiendo a esta como una medida extensiva y aditiva que depende solo del estado termodinámico del sistema. Clausius la pudo cuantificar en la inecuación:

$$dS \geq \sum_{i=1}^n \frac{\delta Q_i}{T_i} \quad (27)$$

donde el subíndice i se refiere que se encuentra a una temperatura T_i y de la cual se extrae uno se entrega una cantidad de calor δQ_i . Si el proceso es reversible y adiabático entonces la inecuación se convierte en ecuación. Podemos ver que Clausius solo obtuvo la diferencia de entropía entre dos estados y no el valor absoluto en cada estado. Esto se conseguiría años después.

Fue el físico austriaco Ludwig Boltzmann en la década de los setenta del siglo XIX quien pudo darle una interpretación estadística a la entropía. Su formulación fue

$$S = k \cdot \ln(\alpha); \quad (28)$$

donde k es la constante de Boltzmann y α es el número de microestados posibles del sistema, supone que todos los microestados tienen la misma probabilidad. Esta última condición no es más que imponer una condición de equilibrio sobre el sistema. Si en estas condiciones se cumple el teorema de Bernoulli uno puede aproximar a la probabilidad de que ocurra un estado de la siguiente forma $p_i = \frac{\text{rank}(i)}{\alpha}$ ($\text{rank}(i)$ significa el número de veces que aparece el microestado i), pero como dijimos en el equilibrio todos los estados son equiprobables y por lo tanto esto implica que $p_i = \frac{1}{\alpha}$. Si reemplazamos $1 = \sum_{i=1}^{\alpha} 1/\alpha$, por lo tanto,

$$S = -k \sum_{i=1}^{\alpha} p_i \ln(p_i) \quad (29)$$

3.1.1. Entropía de Información

Unos de los logros más importantes del siglo pasado en el ámbito de las matemáticas la Teoría de la Información creada por Claude E. Shannon cerca de 1945. Aunque su motivación fue la formalización de la Teoría de las Comunicaciones sus resultados se fueron esparciendo por muchísimos ámbitos científicos. Uno de los conceptos fundamentales de esta teoría es la noción de entropía de información. Esta se puede interpretar como la información promedio transmitida por una variable aleatoria X .

Pensemos en el siguiente sistema; supongamos que uno hace n mediciones de la variable aleatoria X . Estas observaciones pueden ser tomadas como un código escrito en una cadena de símbolos. La caracterización estadística de la cadena va a estar dada por la distribución de probabilidad de sus caracteres. Si uno lo ve desde la perspectiva de que lo que estamos leyendo es un mensaje enviado por alguien entonces la caracterización del mensaje se dará por la distribución de probabilidad de la fuente de información. La entropía de información se puede entender también como una medida de incertidumbre al leer el mensaje ya que, como veremos en los párrafos siguientes, la entropía para fuentes con distribuciones que contienen mucho ruido (en el sentido de probabilidad) es mayor; siendo máxima para un fuente completamente aleatoria. Esto parece contradictorio ya que en el primer párrafo la definimos como el promedio de la información, pero no, es complementario. Es decir, si la Entropía caracteriza la incertidumbre nuestra al leer el mensaje reflejará por lo tanto la información del mensaje. Si nuestra incertidumbre es máxima significa que la información promedio que transmite el código es máxima. Esto se entenderá mejor cuando desarrollemos con más detalles el concepto de información intrínseca de una observación.

Definición: INFORMACIÓN

Para poder entender el significado de *promedio de información*, deberemos obviamente, comprender a que nos referimos con información. Uno de las suposiciones más fuertes que hizo Shannon fue la de imponer que la información que trae una variable aleatoria discreta sea una función (llamémosla $I(p)$) que dependa solamente de la distribución de probabilidad de esa variable. Es decir, si buscamos la información que nos brinda una cierta observación, esta dependerá solo de la probabilidad de aparición de esta observación. Este punto es esencial ya que habla de una connotación estadística de la información. Shannon escribió en su trabajo las propiedades que esta función debía satisfacer;

1. $I(p) \geq 0$; La información es una cantidad no negativa
2. $I(1) = 0$; La información transmitida por un evento que siempre ocurre es nula.
3. $I(p_{1,2}) = I(p_1) + I(p_2)$; es una cantidad aditiva para eventos independientes.

En el último item la función $p_{1,2}$ se refiere a la probabilidad conjunta del evento 1 y 2

El procedimiento exacto con el cual Shannon pudo encontrar esta función excede esta tesis, nosotros nos conformaremos en dar el resultado al que llegó y verificaremos que cumple con las tres propiedades antes mencionadas. La función información de una observación es

$$I(p) = \log\left(\frac{1}{p}\right); \quad (30)$$

El logaritmo cumple con la aditividad para eventos independientes ya que la probabilidad conjunta de eventos independientes es el producto de las marginales. Y como $0 \leq p \leq 1$ siempre va a ser no negativa. La base del logaritmo no es de importancia conceptual, pero si se tendrá en cuenta a la hora de hacer un calculo explícito. Si la base es de 2 entonces la unidad de información es un *bit*.

Definición: ENTROPÍA DE INFORMACIÓN

Sea X una variable aleatoria con valores discretos (x_1, \dots, x_n) y con distribución de probabilidad $P(X) = \{p(x_1), \dots, p(x_n)\}$. Definimos a la entropía de información

$$H(X) = E[I(X)] = E[-\log(P(X))] = \sum_{i=1}^n -p(x_i)\log(p(x_i)); \quad (31)$$

En su artículo Shannon demostró como llegar a la ecuación 31 de una forma rigurosa. También demostró propiedades adicionales que cumple el funcional $H(P(X)) = H(p_1, \dots, p_n)$; estas son:

1. $H(P(X))$ debe ser continua respecto los p_i , es decir, que para cambios pequeños en la distribución de probabilidad trae como consecuencia pequeños cambios en la H .
2. Es simétrica respecto a las probabilidades, esto es $H(p_1, p_2, \dots) = H(p_2, p_1, \dots)$.
3. Si se cumpliera la ley debil de los grandes números (especificamente el teorema de Bernulli), es decir si podemos especificar a cada probabilidad con $p_i = \frac{\text{rank}(i)}{n}$; y si aparte todos los eventos son equiprobables, osea $p_i = \frac{1}{n}; \forall i$ entonces H es monotonamente creciente con n . Y como aclaramos en la sección anterior su máximo valor va a estar dado por $\max(H) = \log_b(n)$, dependiendo de b que es la base del logaritmo .

Vemos que el funcional tiene una similitud con la definición de entropia estadística, salvo por la constante k. Por esto es que uno puede interpretar ciertos hecho naturales desde la perspectiva de la Teoría de la Información.

Como ejemplo esclarecedor pensemos en lo siguiente: Tomemos n observaciones de un ensayo de Bernulli. Supongamos que tenemos una moneda “perfecta”, es decir, que la probabilidad de que salga cruz y cara son las mismas, o sea $p_1 = p_2 = \frac{1}{2}$. Por lo dicho anteriormente si tenemos eventos equiprobables la entropía es máxima ya que en el próximo lanzamiento no tendremos información sobre que lado de la moneda caerá. Pero

si tuviésemos una moneda que tenga un lado más pesado que el otro las probabilidades ya no serán iguales, lo que implicaría que la entropía de información H ya no sea máxima porque un lado de la moneda caerá con más probabilidad que el otro. Esto no significa que adivinaríamos el resultado del siguiente lanzamiento. Esto es consecuencia de que definimos a la información con una connotación probabilística. Pero si tuviésemos la certeza de que la moneda caiga del lado “cara” la entropía se anula ya que la incertidumbre es cero.

Relacion entre la Entropía y la Informacion Mutua En el Apendice C; se ha demostrado la identidad $H(Y|X) = H(X, Y) - H(X)$. Esta relación nos permite interpretar a $H(Y|X)$ como la incertidumbre que se produce solamente por la variable Y , interpretando implícitamente que $H(X, Y)$ es la incertidumbre total producida por la medición de las dos variables aleatorias. Podemos reescribir esta ecuación de la siguiente manera,

$$H(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y); \quad (32)$$

sumando y restando $H(Y)$ tenemos que

$$H(X, Y) = H(X) + H(Y) - H(Y) + H(Y|X); \quad (33)$$

Teniendo esto en cuenta podemos definir a la Información Mutua como

$$I(X : Y) \doteq H(Y) - H(Y|X) = H(X) - H(X|Y); \quad (34)$$

pudiendo así describir a la Entropía Conjunta como

$$H(X, Y) = H(X) + H(Y) - I(X : Y); \quad (35)$$

La ecuación (35) nos da un indicio de como se “distribuye” la incertidumbre entre las dos variables aleatorias. La cantidad $I(X : Y)$ viene a representar la incertidumbre compartida entre las dos variables. Vemos que acá también hay una diferencia con respecto a $P(X, Y)$ ya que esta le da una medida de probabilidad al conjunto intersección; y $H(X, Y)$, por mas que usa las probabilidades conjuntas, mide la incertidumbre total de las dos variables.

Se puede ver que si uno tomase la misma variable aleatoria, es decir, calcular la auto-información mutua, y sabiendo que si X es una variable aleatoria $H(X|X) = 0$ ya que no hay información que pueda aportar aparte de que la que aporta, entonces

$$I(X : X) = H(X) - H(X|X) = H(X); \quad (36)$$

Esta ecuación muestra la relación que estábamos buscando. De todos modos los significados son muy distintos para cada lado de la ecuación (36).

3.2. Relaciones de Similitud entre Cadenas Simbólicas

Como hemos descripto en apartados anteriores, nuestro propósito no se centra en la relación entre los Procesos Estocásticos si no entre las distribuciones de probabilidad de los mismos. El hecho de no tener en nuestras manos los estimadores del primer momento y de la variancia de las cadenas simbólicas hace que si uno busca “la relación” entre dos cadenas tenga que recurrir a herramientas distintas que la Correlación o Covarianza. Para bien nuestro la Teoría de la Información nos brinda la solución. Son muchas las funciones que nos pueden dar, en algun sentido, “la relación” entre dos cadenas simbólicas. Es aquí donde el nombre de Distancia juega un papel importante. Hay muchas distancias y divergencias y el uso de una o de otra depende del propósito que se tenga.

Primero haremos una descripción matemática de los conceptos de Divergencia y Distancia, daremos algunos ejemplos significativos para nuestra tesis, y luego en la siguiente sección desarrollaremos con mas detalle la herramienta que utilizaremos en esta tesis, la Divergencia de Jensen-Shannon (JSD).

Definición: DISTANCIA

Sea \mathcal{X} un conjunto abstracto, y sea d una función definida sobre este conjunto. Se le llamará distancia a la función d si cumple

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

donde x e y son elementos de \mathcal{X} . Pero si además la función d satisface la desigualdad triangular, es decir,

$$d(x, z) \leq d(x, y) + d(y, z) \quad (37)$$

se la llamará Métrica asociada al espacio \mathcal{X} . Algo a recordar es que nuestro espacio abstracto \mathcal{X} no es el espacio de cadenas simbólicas si no el espacio de distribuciones de probabilidades, ya que las distancias o métricas que usaremos no distinguen cadenas que tengan la mismas distribuciones de probabilidad. Esto se aclarará cuando se defina la D_{JS} , y se verá que su dominio son las distribuciones de probabilidad y no la cadena simbólica en sí.

Se dirá que una función de f es Divergencia si cumple con las primeras dos condiciones, es decir, no es simétrica. Como se verá luego, la D_{JS} , cumple con la condición de distancia, aunque se la llame divergencia.

3.2.1. Divergencia de Kullback-Leibler

Como vimos recientemente uno puede definir una distancia entre cadenas simbólicas y así caracterizar ciertas propiedades de las mismas. Pero también puede ocurrir que existan

características que se puedan medir con funciones que no son necesariamente distancias o métricas del espacio.

Definición: ENTROPÍA RELATIVA O DIVERGENCIA K-L

Sean $P(X)$ y $Q(X)$ dos distribuciones de probabilidad de la misma variable aleatoria discreta. Definimos como Divergencia de K-L a

$$D_{KL}(P\|Q) \doteq \sum_{x \in X} P(x) \cdot \log \left(\frac{P(x)}{Q(x)} \right) = E_P \left[\log \left(\frac{P(X)}{Q(X)} \right) \right] \quad (38)$$

Se toma como convención que si $0 \cdot \log(0/Q) = 0$ y que si $P \cdot \log(P/0) = \infty$. También se puede ver que si $P = Q$ entonces $D(P\|Q) = 0$.

De la desigualdad de Jensen descrita en el Apéndice C, podemos ver que

$$f(E[X]) \leq E[f(X)] \quad (39)$$

Si consideramos que nuestra $f(Y) = -\log(Y)$, donde Y es una nueva variable aleatoria discreta definida como $Y \equiv \frac{Q(X)}{P(X)}$, si $P(X) > 0$ y $Y = 0$ si $P(X) = 0$. Vemos que la función $[-\log(\cdot)]$ es convexa y como $\sum_i P(x_i) = 1$ (cumple con la condición impuesta a los γ_i en el Apéndice C), entonces tenemos que

$$\begin{aligned} D_{KL}(P\|Q) &= E_P \left[\log \left(\frac{P(X)}{Q(X)} \right) \right] = E_P \left[-\log \left(\frac{Q(X)}{P(X)} \right) \right] \\ &\geq -\log \left(E_P \left[\frac{Q(X)}{P(X)} \right] \right) \\ &= -\log \left(\sum_{P(x) \neq 0} Q(x) \right) \\ &\geq 0 \end{aligned} \quad (40)$$

la última desigualdad es cierta ya que $\sum_{i \in I} Q(x_i) \leq 1$ ya que el conjunto I es del conjunto tal que $P(X)$ este normalizada. Vemos claramente que si el conjunto I es tal que $\sum_{i \in I} Q(x_i) = 1$ entonces la última desigualdad se convierte en igualdad. Es decir que cumple con la primer condición de Distancia que habíamos descripto anteriormente.

Es bueno aclarar que la Divergencia K-L no es una distancia, y tampoco será una métrica del espacio de probabilidades, ya que no satisface algunas de las propiedades mencionadas anteriormente;

1. $D_{KL}(P\|Q) \neq D_{KL}(Q\|P)$, no es simétrica
2. No cumple la desigualdad triangular
3. No es acotada para $Q(X) \equiv 0$

4. Necesitamos que la medida de probabilidad $P(X)$ sea absolutamente continua respecto de $Q(X)$, esto es, que si $P(X) = 0$ para todo el rango de la variable aleatoria X (también se puede tomar que es para $\beta \in \mathcal{B}$, donde es el conjunto de las σ -álgebra) entonces $Q(X) = 0$ para todo el rango de X , [19].

En busca de una interpretación de esta divergencia, se puede ver que

$$D_{KL}(P\|Q) = \sum_{i \in I} P(x_i) \log(P(x_i)) - \sum_{i \in I} P(x_i) \log(Q(x_i)) = -H(P(X)) + H(P, Q) \quad (41)$$

donde el último término es la Entropía Cruzada. Si utilizamos que $D_{KL}(P\|Q) \geq 0$, se puede ver que

$$-\sum_{i \in I} P(x_i) \log(Q(x_i)) \geq -\sum_{i \in I} P(x_i) \log(P(x_i)) = H(P) \quad (42)$$

En este sentido se puede entender a la $D(P\|Q)$ como la cantidad de bits que necesitamos de más por usar la distribución $Q(X)$ en vez de la $P(X)$, ya que el lado derecho no crece, dado que la distribución P es fija.

Otra forma de interpretar a la $D(P\|Q)$ es relacionarla con la información mutua de dos variables aleatorias. Si reescribimos la ecuación (34), sabiendo que $P(y)P(x|y) = P(x, y)$ y que $\sum_{x \in \mathcal{X}} P(x, y) = P(Y)$ tenemos

$$\begin{aligned} I(X : Y) &= H(Y) - H(Y|X) \\ &= -\sum_{y \in \mathcal{Y}} P(y) \log(P(y)) + \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} P(x) P(y|x) \log(P(y|x)) \\ &= -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(x, y) \log(P(y)) + \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(x, y) \log(P(xy|x)) \\ &= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(x, y) \log\left(\frac{P(y|x)}{P(y)}\right) \\ &= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(x, y) \log\left(\frac{P(x, y)}{P(y)P(x)}\right) \end{aligned} \quad (43)$$

Por lo tanto la información mutua de las variables X e Y resulta ser igual a la D_{KL} entre las probabilidades conjunta y la multiplicación de las dos marginales esto es

$$I(X : Y) = D(P(X, Y)\|P(X)P(Y)) \quad (44)$$

Así podemos interpretar a la información mutua como cuán lejos está la distribución de probabilidad conjunta de la distribución conjunta de variables independientes

3.2.2. Divergencia Jensen-Shannon

Viendo las falencias de la Divergencia K-L, J.Lin propuso una Divergencia que está íntimamente relacionada con la D_{KL} y le da el nombre de Divergencia Jensen-Shannon (D_{JS}), su definición es

$$D_{JS}(P(X), Q(X)) \doteq \frac{1}{2} \left(D_{KL} \left(P, \frac{P+Q}{2} \right) + D_{KL} \left(Q, \frac{P+Q}{2} \right) \right) \quad (45)$$

Esta medida de similitud corrige, en algún sentido, las falencias presentadas por la Divergencia anterior. En esta sección nos ocuparemos de desarrollar las propiedades de esta cantidad.

Usando la definición de la D_{KL} podemos ver que

$$D_{JS}^{(1/2)}(P, Q) = \frac{1}{2} \left[\sum_{i \in I} P(x_i) \cdot \log \left(\frac{2P(x_i)}{P(x_i) + Q(x_i)} \right) + \sum_{j \in I} Q(x_j) \cdot \log \left(\frac{2Q(x_j)}{Q(x_j) + P(x_j)} \right) \right] \quad (46)$$

Como los índices i y j barren todo el intervalo I entonces podemos reemplazar j por i , tenemos

$$\begin{aligned} D_{JS}^{(1/2)} &= \frac{1}{2} \left[\sum_{i \in I} P(x_i) \cdot \log (P(x_i)) - \sum_{i \in I} P(x_i) \cdot \log \left(\frac{P(x_i) + Q(x_i)}{2} \right) \right] \\ &\quad + \frac{1}{2} \left[\sum_{i \in I} Q(x_i) \cdot \log (Q(x_i)) - \sum_{i \in I} Q(x_i) \cdot \log \left(\frac{P(x_i) + Q(x_i)}{2} \right) \right] \end{aligned}$$

Utilizando las definiciones de Entropía de información podemos ver que

$$D_{JS}^{(1/2)}(P, Q) = H \left(\frac{P+Q}{2} \right) - \frac{1}{2} [H(P) + H(Q)] \quad (47)$$

Surgen entonces tres propiedades importante.

1. $D_{JS}^{(1/2)}(P, Q) \geq 0$; Por la desigualdad de Jensen, al ser $D_{KL} \geq 0$ positiva para cualquier par de distribuciones de probabilidad.
2. $D_{JS}^{(1/2)}(P, Q) = D_{JS}^{(1/2)}(Q, P)$; es simétrica como se ve fácilmente en la ecuación (47).
3. Es acotada ya que la entropía lo es.

Una formulación más general todavía de la D_{JS} , la dio luego J. Lin asignándole pesos a las distribuciones de probabilidad:

$$D_{JS}^{(\pi_1, \pi_2)}(P, Q) = H(\pi^{(1)}P + \pi^{(2)}Q) - \pi^{(1)}H(P) - \pi^{(2)}H(Q) \quad (48)$$

donde $\pi^{(1)}$ y $\pi^{(2)}$ son los pesos asignados a P y Q respectivamente y cumplen $\pi^{(1)} + \pi^{(2)} = 1$. Vemos facilmente que si $\pi^{(1)} = \pi^{(2)} = 1/2$ queda la ecuación (47) que deducimos anteriormente. Vale aclarar que esta definición de la D_{JS} es equivalente a la propuesta en la ecuación (45), por eso no tiene el supra indice 1/2. La elección de los pesos se aclarará en el transcurso de la sección. Haremos un resumen de las propiedades de la D_{JS} (general), y especificaremos cuales se agregan si es la $D_{JS}^{1/2}$:

1. Si P(X) y Q(X) son distribuciones de probabilidad para una misma variable aleatoria, dado que la D_{KL} es siempre positiva, se sigue de la definición (45) que

$$D_{JS}(P, Q) \geq 0; \quad (49)$$

Con $D_{JS}(P, Q) = 0$ si y solo si $P = Q$.

2. Está siempre bien definida, es decir, no necesitamos que las distribuciones de probabilidad sean continuamente absolutas .
3. Es acotada;

$$0 \leq D_{JS} \leq \ln(2) \quad (50)$$

Pero si consideramos a los pesos iguales, osea que $\pi^{(1)} = \pi^{(2)} = 1/2$, la D_{JS} tiene dos propiedades más (aunque ahora se debería llamar $D_{JS}^{(1/2)}$)

1. Es simétrica respecto de sus argumentos $D_{JS}^{(1/2)}(P, Q) = D_{JS}^{(1/2)}(Q, P)$
2. Es el cuadrado de la métrica de nuestro espacio de probabilidades, esto es

$$d_{JS}(P, Q) = \sqrt{D_{JS}^{1/2}(P, Q)}; \quad (51)$$

y se puede verificar que la ecuación (51) cumple con la desigualdad triangular (como toda métrica), es decir que,

$$d_{JS}(P, Q) \leq d_{JS}(P, W) + d_{JS}(W, Q); \quad (52)$$

donde P(X), Q(X) y W(X)son distribuciones de probabilidad definidad para la misma variable aleatoria.

Si uno tuviera más de dos distribuciones de probabilidad uno podría generalizar la función D_{JS} la cual quedaría como

$$D_{JS}(P^{(1)}, \dots, P^{(m)}) = H \left(\sum_{i=1}^m \pi_i P^{(i)} \right) - \left(\sum_{i=1}^m \pi_i H(P^{(i)}) \right) \quad (53)$$

Teniendo la forma mas general de la divergencia J-S podemos dar lugar a su interpretación en la teoría de la información.

Interpretación de la DJS Como hicimos con la D_{KL} , intentaremos dar una interpretación dentro del marco de la teoría de la información. Primero haremos una introducción a la explicación en la cual definiremos formalmente cada variable.

Consideremos una secuencia \mathcal{S} de N símbolos escogidos de un alfabeto $\mathcal{A} = \{a_1, \dots, a_k\}$ y denotemos como p_i a la probabilidad de encontrar al símbolo a_i dentro de la cadena \mathcal{S} en una posición fija pero arbitraria. Supongamos ahora que dividimos la secuencia original en m subsecuencias $\mathbb{S} = \mathcal{S}^{(1)}, \dots, \mathcal{S}^{(m)}$ de longitudes $n^{(1)}, \dots, n^{(m)}$; y denotamos como $p_i^{(j)}$ a la probabilidad de encontrar el símbolo a_i en una posición fija pero arbitraria dentro de la subsecuencia $\mathcal{S}^{(j)}$.

Definamos ahora dos variables aleatorias, una será “el símbolo” y la llamaremos “A” y la otra será “la subsecuencia” y la llamaremos “S”, es decir, la variable “A” toma valores del conjunto $\mathcal{A} = \{a_1, \dots, a_k\}$ y la variable “S” toma valores del conjunto $\mathbb{S} = \mathcal{S}^{(1)}, \dots, \mathcal{S}^{(m)}$. Lo que haremos es elegir arbitrariamente (con probabilidad uniforme) un casillero de la secuencia \mathcal{S} , y anotaremos que símbolo sale (variable aleatoria “A”) y en que subsecuencia está contenido (variable aleatoria “S”). Repitiendo el procedimiento (y teniendo en cuenta la ley de los grandes números descripta anteriormente) podremos construir la probabilidad de que salga elegido (de un sorteo arbitrario) el símbolo a_i en la subsecuencia \mathcal{S}_j , esta la probabilidad conjunta de las dos variables “a” y “s”, que es p_{ij} . Obviamente se interpreta a p_{ij} como $p(A = a_i, S = \mathcal{S}^{(j)})$.

Teniendo en cuenta la definición de probabilidad marginal podemos ver que la probabilidad de que salga elegido el símbolo a_i pero en cualquier posición de la cadena original \mathcal{S} es $p_i = \sum_{j=1}^m p_{ij}$. De la misma forma podemos preguntarnos cual será la probabilidad de que el casillero que salga sorteado (con probabilidad uniforme, o sea arbitrariamente) se encuentre en la subsecuencia $\mathcal{S}^{(j)}$ sin importar el símbolo, esto es $\pi_j = \sum_{i=1}^k p_{ij}$. Pero este resultado lo podemos deducir por otro camino ya que conocemos los largos de cada subsecuencia. Supongamos que sorteamos casilleros arbitrariamente, un número grande de veces, y quisiésemos calcular la probabilidad de que salga sorteado un casillero que pertenezca a la subsecuencia $\mathcal{S}^{(j)}$. También sabemos que esta subsecuencia tiene un largo $n^{(j)}$. Si la probabilidad del sorteo es uniforme, es decir, es un sorteo “bueno” o arbitrario como se suele decir, y si tenemos en cuenta la ley de los grandes números podemos decir que la probabilidad π_j es el número de veces que salió un casillero perteneciente en la subsecuencia $\mathcal{S}^{(j)}$ dividido la cantidad total de veces que se hizo el sorteo. Para facilitar los cálculos podemos decir que el sorteo se haga un factor de veces el número total de elementos de la cadena, o sea, N . Esto significa hacer el sorteo, por ejemplo, $1000000 \cdot N$, tal que $\pi_j = \frac{\#\mathcal{S}^{(j)}}{1000000 \cdot N} = \frac{n^{(j)}}{N}$. Resumiendo la probabilidad de aparición de un casillero en una subsecuencia $\mathcal{S}^{(j)}$ sea cual sea el símbolo, es $\pi_j = \frac{n^{(j)}}{N}$. Ya teniendo las definiciones formales de las variables, se nos presenta el siguiente problema. Supongamos que se hace un sorteo (con probabilidad uniforme o arbitrario) del cual se extrae un casillero de la cadena original \mathcal{S} y se anota el valor de las variables A y S , pero solo se nos comunica el valor de la variable A y nuestra tarea es averiguar el valor de la variable S . Es decir, nos dicen el símbolo y tenemos que averiguar en que subsecuencia se encuentra. Entonces

la pregunta a hacerse es; ¿contiene información la variable A sobre la variable S ?. La teoría de la Información nos brinda una herramienta, la información mutua. Recordando la ecuación (43), resulta

$$I(A : S) = \sum_{a \in \mathcal{A}} \sum_{s \in \mathcal{S}} p(a, s) \log \left(\frac{p(a, s)}{p(a)p(s)} \right) = \sum_{i=1}^k \sum_{j=1}^m p_{ij} \log \left(\frac{p_{ij}}{p_i \pi_j} \right) \quad (54)$$

También sabemos que la probabilidad condicional de tener el símbolo a_i dada la subsecuencia $\mathcal{S}^{(j)}$ es $p_i^{(j)} \equiv p(a_i | \mathcal{S}^{(j)}) = \frac{p_{ij}}{\pi_j} = \frac{p(A=a_i, S=\mathcal{S}^{(j)})}{p(S=\mathcal{S}^{(j)})}$, podemos reescribir la ecuación (54) en

$$\begin{aligned} I(A : S) &= \sum_{i=1}^k \sum_{j=1}^m \pi_j p_i^{(j)} \log \left(\frac{p_i^{(j)}}{p_i} \right) \\ &= \sum_{i=1}^k \sum_{j=1}^m \pi_j p_i^{(j)} \log \left(p_i^{(j)} \right) - \sum_{i=1}^k \sum_{j=1}^m \pi_j p_i^{(j)} \log (p_i) \\ &= \sum_{j=1}^m \pi_j \sum_{i=1}^k p_i^{(j)} \log \left(p_i^{(j)} \right) - \sum_{i=1}^k \left(\sum_{j=1}^m \pi_j p_i^{(j)} \right) \log (p_i) \end{aligned} \quad (55)$$

Haciendo uso de que $p_i = \sum_{j=1}^m p_{ij} = \sum_{j=1}^m \pi_j p_i^{(j)}$, tenemos que la ecuación (55) queda

$$\begin{aligned} I(A : S) &= - \sum_{j=1}^m \pi_j H \left(P^{(j)} \right) + H \left(\sum_{j=1}^m \pi_j P^{(j)} \right) \\ &= D_{JS} \left(P^{(1)}, \dots, P^{(m)} \right) \end{aligned} \quad (56)$$

donde $P^{(j)} \equiv P(A | \mathcal{S}^{(j)})$ es la distribución de probabilidad condicional de la variable aleatoria A dada la subsecuencia $\mathcal{S}^{(j)}$, que también se puede interpretar directamente como la distribución de probabilidad de la subsecuencia $\mathcal{S}^{(j)}$. Como podemos ver de la ecuación (56) la información mutua (o que comparten) las variables A y S es igual a la D_{JS} (generalizada) entre las distribuciones de probabilidad condicional de todas las subsecuencias. Algo que vale la pena enfatizar de esta deducción es que nos brinda una definición de los pesos para la D_{JS} generalizada, que surgió de forma natural sin hacer ninguna suposición. O si se quiere ver de otra manera, si queremos interpretar a la D_{JS} como se la interpretó recientemente en la ecuación (56) deberemos también interpretar o definir a los pesos como $\pi_j = \frac{n^{(j)}}{N}$, que no es más que la fracción de caracteres que pertenecen a la subsecuencia $\mathcal{S}^{(j)}$ (o como también se dijo, la probabilidad de que si se sortea un casillero sea cual sea el símbolo pertenezca a la subsecuencia $\mathcal{S}^{(j)}$). Esta interpretación se revisará en los siguientes capítulos especificando para el caso que investigaremos nosotros.

4. ASPECTOS METODOLOGICOS

En esta sección desarrollaremos el método utilizado en nuestra investigación. Esto abarcará aspectos teóricos como también los algoritmos computacionales usados. Las herramientas teóricas serán dos, la D_{JS} y la distancia propuesta por Yang y Yein. Ambas tienen, como se vio, aspectos diferentes, y por lo tanto, la metodología también los tendrá.

Primero describiremos la forma en que mapearemos una serie temporal a una cadena simbólica binaria y luego la técnica con la que analizamos dicha cadena.

4.1. Correspondencia entre Serie Temporal y cadena simbólica binaria

Supongamos que tenemos una serie temporal como las ejemplificadas en el inicio de este trabajo, y la representamos con $\mathcal{S} = \{x_1, \dots, x_N\}$. Como tenemos una única serie las limitaciones en el análisis son obias. Uno podría estimar, suponiendo condiciones de estacionariedad, los momentos y varianzas para luego poder inferir nuestro proceso estocástico en el tiempo $(N+1)$, o sea X_{N+1} . Pero si nuestro propósito es distinto, como por ejemplo deducir propiedades que tienen que ver con la estadística en general y no con la inferencia de un valor futuro, la herramienta a considerar será la teoría de la información. Esto lleva a cambiar nuestro esquema de análisis ya debemos adaptar nuestro problema a la teoría de la información.

Para poder analizar series temporales dentro del marco de la teoría de la información deberemos primero transformar estas a cadenas simbólicas. ¿Por qué? ¿Que beneficios nos da hacer esto? Las cadenas simbólicas son sucesiones de símbolos, que pertenecen a un alfabeto finito y conocido, ordenados por un parámetro que llamaremos tiempo “t”. Las series temporales, en cambio, son sucesiones de números ordenados pero que no pertenecen a un alfabeto finito y conocido. Como mucho, se podrá saber el rango en el cual estarán esos números, pero nunca los podremos ubicar en un alfabeto finito. Esto es un problema si queremos usar, por ejemplo, la divergencia J-S ya que esta es un funcional que tiene a las distribuciones de probabilidad discretas como dominio. Por lo tanto, será necesario transformar nuestra serie temporal a una sucesión de símbolos que pertenezcan a un alfabeto finito y conocido.

El método que eligiremos es el presentado por Yang, Yein y Hseu en su trabajo del 2003 [1]. Es importante ver que sea cual sea el método que elijamos, tendrá que ser capaz de que la cadena simbólica resultante exprese las propiedades que tenía la serie temporal. Surge la pregunta: ¿Que propiedades queremos que se transmitan de \mathcal{S} a \mathcal{C} ?. Se sobrentiende que del hecho de reducir el alfabeto hace que no haya una relación unívoca entre la cadena simbólica resultante y la serie temporal. Esto plantea un problema de subjetividad, en el cual, lo que pesará será nuestros objetivos. Para nuestra investigación será importante que la cadena simbólica \mathcal{C} represente de la mejor manera posible la estructura de la serie temporal, es decir, no nos interesa en si los valores $\mathcal{S} = \{x_1, \dots, x_N\}$ sino la estructura

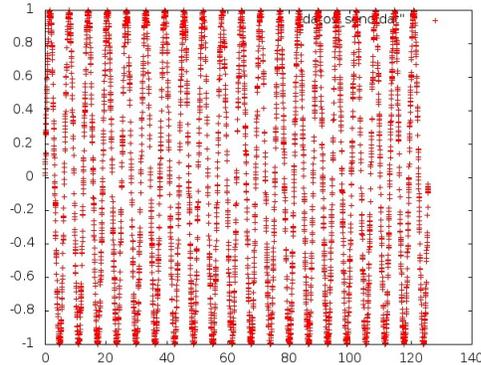


Figura 1. La función seno de 5000 puntos elegidos al azar (que fueron ordenados de menor a mayor)

general de la serie. Y si nos ponemos minuciosos lo que buscamos es como se relacionan estos valores, no su valor en sí. El método presentado por Yang [1], mantiene este punto de vista, y es por eso que lo elegimos.

Sea $\mathcal{S} = \{x_1, \dots, x_N\}$ nuestra serie temporal, definimos al elemento “n-ésimo” de la cadena binaria $\mathcal{C} = \{a_1, \dots, a_{N-1}\}$, como

$$c_n = \begin{cases} 0 & ; \text{if } x_n \leq x_{n+1} \\ 1 & ; \text{if } x_n > x_{n+1} \end{cases} \quad (57)$$

La cadena \mathcal{C} tiene un elemento menos, es decir, llegará hasta a_{N-1} . Se ve claramente que lo que se busca con esta asignación es darle importancia al crecimiento y decrecimiento de la serie temporal, con lo que se logra en consecuencia mapear la estructura general de la serie. Como ejemplo transformaremos una serie temporal creada por una función periódica (el seno por ejemplo) valuada en puntos arbitrarios del dominio y luego verificaremos que la estructura periódica se mantiene.

Supongamos que elegimos al azar 5000 puntos pertenecientes al intervalo $I = [0; 40*\pi]$. A esos puntos los ordenamos de menor a mayor, para luego aplicarles la función $\sin(\cdot)$. Si graficamos los puntos obtenemos la Figura 1

Si ahora utilizamos la técnica mencionada anteriormente, podremos mapear estos puntos a una cadena binaria. Se pondrá una porción de la cadena tal que se pueda interpretar

enteras (suponiendo condiciones de estacionareidad) estaremos analizando solamente su crecimiento y decrecimiento de forma estadística y general. Pero si nosotros queremos entender su comportamiento de una manera más local para poder distinguir “Cambios Significativos” de la serie (por ejemplo; que una serie temporal tenga un comportamiento aleatorio y de repente se vuelva constante por cierta cantidad de casilleros) debemos utilizar estas herramientas de una forma local. Es aquí donde la implementación de un método es necesaria. *Necesitamos un método para que con la teoría estadística se pueda captar cambios estructurales de una serie.* Nuestro método, que lo llamaremos “La Ventana Móvil”, nos permitirá implementar a la D_{JS} de una forma local y dinámica. El valor de la D_{JS} se interpretará como una medida de disimilitud. Cuando los valores sean “grandes” (el significado de “grande” viene dado por el concepto de significancia, que será descrito en capítulos posteriores) implicará que a ese tramo de la cadena hay un “Cambio Significativo”. El porque interpretamos a los valores de la divergencia de esta manera se explicará en el detalle del algoritmo, en el capítulo siguiente. Nuestra principal herramienta teórica será la D_{JS} . Trabajaremos con cuatro series temporales, generadas por distintos procesos, caóticos y aleatorios. Estas secuencias son, Mapa Logístico, Temp-Map, Ruido Rosa y Ruido Blanco. Para esta primera etapa uniremos de a pares una secuencia con otra secuencia y formando una cadena doble. Nuestro propósito es poder detectar, con la ayuda de la D_{JS} , el punto de unión de las dos subsecuencias. En los ejemplos dados en la tesis, el punto de cambio es conocido, pues nosotros hemos construido la secuencia completa. El interés, desde el punto de vista de las aplicaciones, está en poder distinguir transiciones caóticas o ruidosas en una secuencia arbitraria. En esta sección analizaremos en detalle el algoritmo que esta hecho en base a la D_{JS} . Como contamos anteriormente la D_{JS} es un funcional que va desde el espacio de las distribuciones de probabilidad a los reales. Nosotros utilizaremos la ecuación (48), o sea, la D_{JS} escrita en su versión de pesos para dos distribuciones de probabilidad. Con el fin de encontrar cambios sugerentes en la distribución de probabilidad, hemos diseñando un algoritmo que sea capaz de distinguir estos cambios en una escala diez veces (en orden de magnitud) inferior al largo de la cadena doble. El término “Cambios Significativos” se detallará en una sección futura, pero es bueno que uno sepa que existe una subjetividad inherente en la definición de este término, es nuestro deber dejar en claro porque se elige lo que se elige. Es por esto que se dedica en esta tesis una sección a la definición detallada de este termino.

No será necesario para entender conceptualmente el algoritmo, la definición detallada de “Cambios Significativos”, es por esto que primero presentaremos el algoritmo y después si el lector desea podrá leer la definición de Singificancia en el capítulo posterior.

4.2.1. Algoritmo de “La Ventana Móvil”: El intento de entender una estructura en base a una estadística

La primera pregunta que uno debe hacerse antes de desarrollar un algoritmo, es ¿Qué buscamos?. La respuesta más específica en este caso será: Un algoritmo que sea capaz de distinguir cambios locales en la distribución de probabilidad (que refieren cambios en la “estructura” de la serie temporal) de una cadena simbólica y que se base en herramientas estadísticas, en nuestro caso en la teoría de la información.

Este algoritmo consta principalmente de tres partes. La primera consiste en construir una cadena binaria formada por la unión de dos de las series temporales mencionadas más arriba. En la segunda parte, se construye una cadena simbólica distinta a partir de la cadena binaria anteriormente obtenida. Esta cadena se la llamará cadena alfabetizada, ya que la construiremos en base a valores determinados por *Palabras* de un largo “L” fijado con anterioridad. La estadística se hará en base a esta última cadena simbólica. Por último, introduciremos una *Ventana* que tiene de cada lado un largo de “ \mathcal{L} ”, y que recorrerá la cadena alfabetizada. Luego se calculará la distribución de probabilidad de cada lado (de largo \mathcal{L}) de la ventana. Teniendo esto se aplicará la D_{JS} a las dos distribuciones obtenidas de cada lado de la ventana. Es visible que si nosotros movemos la ventana casillero por casillero las distribuciones irán cambiando casillero por casillero, y en consecuencia el valor de la D_{JS} . Nuestro destino es probar que el valor más grande de D_{JS} se da cuando la ventana se encuentra en la cercanías de la mitad de la cadena alfabetizada. A continuación haremos una descripción detallada del algoritmo.

Creación de la Cadena simbólica (binaria) doble Elijamos dos de las cuatro series temporales que están a nuestra disposición, y llamémoslas \mathcal{S}_1 y \mathcal{S}_2 . Ahora “peguemos” el casillero final de la \mathcal{S}_1 con el primer casillero de la \mathcal{S}_2 . Si suponemos que cada una de ellas tiene largo N, entonces el largo de la serie temporal doble será $2N$. Llamaremos “Serie Total” (\mathcal{S}_T), a la unión de las dos Series (1 Y 2). Este método transformará nuestra serie temporal Total a una cadena binaria de largo $(2N - 1)$, y la llamaremos Cadena Binaria Total, \mathcal{C}_T . El hecho que tenga un elemento menos es consecuencia directa del método elegido, si uno no lo recuerda, será preferible volver a la sección 4.1. Para poder esclarecer el proceso descrito anteriormente, haremos un ejemplo con pocos casilleros.

Tomemos los últimos casilleros de una serie temporal producida por el mapa logístico de largo N y los primeros casilleros de una serie temporal producida por un proceso de ruido blanco entonces tenemos;

$$\begin{aligned}
 \text{Mapa Log} &= \left\{ \begin{array}{l} \dots \\ 0,904065 \\ 0,346926 \\ 0,906274 \\ 0,339767 \\ 0,897301 \\ 0,368606 \\ 0,930943 \\ 0,257154 \\ 0,764103 \\ 0,720999 \\ 0,804638 \\ 0,628782 \\ 0,933661 \end{array} \right. \\
 \text{Ruido Blanco} &= \left\{ \begin{array}{l} 0,003740 \\ 0,146042 \\ 0,247883 \\ -0,040866 \\ 0,229108 \\ 0,044863 \\ 0,004388 \\ 0,207709 \\ -0,127349 \\ -0,162176 \\ -0,109847 \\ -0,097632 \\ 0,128830 \\ \dots \end{array} \right.
 \end{aligned}$$

El tramo de la Cadena Binaria Total correspondiente a los 2×13 valores anteriores, es

$$\dots 1010101010101001011011000 \dots$$

vemos claramente que en este tramo tenemos 25 valores en vez de 26 como en el tramo de la \mathcal{S}_T detallado anteriormente.

Una vez obtenida por completo la Cadena Binaria Total, \mathcal{C}_T , es posible implementar las herramientas de la Teoría de la Información de una manera más simple y compacta, ya que el alfabeto ahora es finito, $\mathcal{A}_{\mathcal{C}_T} = \{0; 1\}$.

Es bueno aclarar que una de las ventajas del método de Yang es que nos permite comparar series temporales que tomen valores completamente sistintos. Es está relacionado a la intención de interpretar propiedades que no se relacionan con los valores del espacio de las faces, si no con los cambios sugerentes de la serie misma.

Construcción de la cadena alfabetizada Supongamos que tenemos la cadena binaria total, \mathcal{C}_T , de $(2 \cdot N - 1)$ casilleros. Ahora tomemos un segmento de largo L , que llamaremos *Palabra*, y leemos los símbolos que tiene en su interior. Este segmento se irá moviendo de a un casillero a la vez, hasta detenerse en el último casillero. Nuestros resultados fueron obtenidos con $L = 8$. En los siguientes párrafos se detallaran cada uno de los pasos en la construcción de la cadena alfabetizada:

1. Tomemos los primeros L casilleros y leamos sus símbolos. La idea será corresponder a este conjunto ordenado de L símbolos un solo número natural perteneciente al intervalo $I = [0, 2^{L+1} - 1]$.
2. La elección del número que asignaremos a la *Palabra* de L elementos será de la siguiente manera:

$$a_i = \sum_{j=0}^{L-1} c_{i+j} \cdot 2^j \quad (58)$$

Donde a_i es el elemento i -ésimo de la cadena alfabetizada, \mathcal{C}_A , y c_{i+j} es el elemento $(i+j)$ -ésimo de la cadena total binaria, \mathcal{C}_T . Vemos que la sumatoria es hasta $(L - 1)$, esto es porque empieza en cero y la palabra tiene largo L . Nuestro primer elemento de la cadena alfabetizada será para $i = 0$. Entonces tendremos que

$$a_0 = c_0 2^0 + c_1 2^1 + \dots + c_{L-1} 2^{L-1} \quad (59)$$

3. El siguiente paso es mover la *Palabra* de largo L un casillero a la derecha, entonces nuestro siguiente elemento de la Cadena Alfabetizada es

$$a_{i=1} = c_1 2^0 + c_2 2^1 + \dots + c_L 2^{L-1} \quad (60)$$

Un punto a tener en cuenta es que hay una relación unívoca entre la palabra contenida en la ventanita y el número correspondiente a la Cadena Alfabetizada.

4. Seguiremos repitiendo este procesos hasta el casillero $[(2N - 1) - (L - 1)]$, es decir, hasta cuando la *Palabra* toque el último casillero. Esto significa que el primer casillero de la *Palabra* estará en el casillero $[(2N - 1) - (L - 1)]$ de la Cadena Total Binaria, \mathcal{C}_T .
5. Luego la Cadena Alfabetizada, \mathcal{C}_A tendrá $(L - 1)$ casilleros menos que la Cadena Total Binaria, o sea contendrá $[(2N - 1) - (L - 1)]$ casilleros.

Uno podría preguntarse porque recurrir a este artilugio, es decir, porqué construir otra cadena si ya tenemos una secuencia de alfabeto finito en la cual se puede implementar fácilmente la teoría de la información. Más allá de que la elección del valor L es algo subjetivo del investigador, unas de las razones más importantes, es poder tener en cada elemento de la Cadena Alfabetizada una noción o representación de los cambios de la serie temporal original. Como dijimos anteriormente los caracteres 0 y 1 de la cadena \mathcal{C}_T representan si creció o decreció con respecto al valor anterior de la serie. Es decir, que en la cadena \mathcal{C}_T se tiene en cuenta los cambios y no el valor absoluto de la serie. Entonces en la Cadena Alfabetizada, \mathcal{C}_A , tiene en cuenta los cambio en un cierto período, definido por nosotros. Por lo tanto, \mathcal{C}_A representará cambios locales de la serie temporal original (locales de largo L) en vez de cambios puntuales como en la cadena \mathcal{C}_T . Hacer estadística sobre la cadena \mathcal{C}_A significará ver que período de cambios locales son más probables que otros. Es decir, uno podría encontrar que en un período (nos referimos período en este caso al tiempo transcurrido) de largo L , es muy probable sucedan crecimientos locales grandes. Esto es equivalente a decir que es muy probable encontrar caracteres a_i de valor elevado.

Tener en un número la representación del crecimiento o decrecimiento local (de largo L) nos permite hacer estadística sobre esas propiedades.

Análisis de la Cadena Alfabetizada Teniendo ya la cadena \mathcal{C}_A estamos en condiciones de implementar Teoría de la Información. Para esto crearemos una *Ventana* de largo $2 \cdot \mathcal{L}$ que se moverá desde el casillero $[\mathcal{L} - 1]$ (recordar que la cadena empieza en el casillero 0) hasta el casillero $[(2N - 1) - (L - 1)] - \mathcal{L}$, es decir, hasta que la *Ventana* toque el último casillero; moviéndose al paso de un casillero por vez. Lo que haremos ahora es analizar, para cada valor del puntero , los símbolos que se encuentran de cada lado de la *Ventana*. Para no generar confusiones , vale remarcar que el valor del puntero es el número del casillero que se encuentra en el medio de la *Ventana* pero al cual no se le lee el símbolo que tiene en su interior. Sus posibles valores van desde \mathcal{L} hasta $[(2N - 1) - (L - 1)] - \mathcal{L}$.

Buscamos ahora calcular la probabilidad de aparición de cada símbolo (pertenecientes al alfabeto $I = [0; 2^{L+1} - 1]$), es decir, queremos encontrar la distribución de probabilidad de cada lado de la *Ventana*, ya que cada lado tendrá su distribución de probabilidad particular. El método para calcular estas distribuciones será utilizar la forma frecuentista de la misma, teniendo en cuenta la verificación de la ley de los grandes números. Nuestro objetivo es encontrar “cambios significativos” en la estructura de la serie temporal, esto se traduce a encontrar “cambios significativos” en la distribución de probabilidad de la cadena alfabetizada. Como dijimos anteriormente esta cadena representa la forma o estructura que tiene la serie temporal subyacente, no así los valores que toma la misma. Por lo tanto analizar estadísticamente la cadena alfabetizada estaremos analizando estadísticamente su estructura. Teniendo ya las distribuciones de probabilidad de cada lado de la *Ventana*, lo que haremos será compararlas utilizando la Divergencia de Jensen-Shannon. Para simplificar el procedimiento construiremos una cronología del mismo.

1. Primero posamos el puntero que se encuentra entre los dos lados de la *Ventana*

(que son de largo \mathcal{L}) en el casillero $(\mathcal{L} - 1)$. Calculamos, de forma frecuentista la distribución de probabilidad de cada lado. A estas las llamaremos $P_{\mathcal{L}-1}^D$ y $P_{\mathcal{L}-1}^I$ distribución de probabilidad en el puntero $\mathcal{L} - 1$ a la derecha y a la izquierda respectivamente.

2. El siguiente paso será calcular la D_{JS} entre las distribuciones recién calculadas. El valor de la divergencia para este valor del puntero se lo llamará $D_{JS}[\mathcal{L} - 1]$
3. Ahora se procederá a mover la *Ventana* un casillero a la derecha, es decir, mover el puntero un casillero a la derecha, a la posición \mathcal{L} . De igual manera, calcularemos las distribuciones de probabilidad de cada lado ($P_{\mathcal{L}}^D$ y $P_{\mathcal{L}}^I$) para luego calcular el valor $D_{JS}[\mathcal{L}]$
4. Este procedimiento se repetirá hasta que el puntero tome el valor $[(2N - 1) - (L - 1)] - \mathcal{L}$, y la divergencia será $D_{JS} [[(2N - 1) - (L - 1)] - \mathcal{L}]$

Teniendo todos los valores correspondientes de la divergencia se construirá un histograma dando el valor de la D_{JS} para cada valor del puntero. Lo que buscamos son los valores del puntero para los cuales la D_{JS} toma los valores más altos. Esos casilleros (o sea los valores del puntero) corresponderán a las posiciones de la cadena alfabetizada donde ocurrirán los mayores cambios en la distribuciones de probabilidad. Qué se corresponderá con los mayores cambios “estructurales” de la serie temporal. Un ejemplo puede ser que un un lugar la serie temporal crezca y decrezca periódicamente y con período constante y después de un cierto casillero crece y decrece con un período muy diferente.

4.2.2. ¿A qué nos referimos con cambios significativos?; el significado de la Significancia

Nos dedicaremos ahora a repasar la noción Significancia utilizada en las publicaciones de Grosse y colaboradores [6]. El objetivo de este apartado es dar una descripción simple y no detallada de la función, ya en el transcurso de la tesis no la utilizaremos. Pero si servirá para completar los conocimientos en las aplicaciones de la teoría de la información en secuencias simbólicas. En nuestra tesis no fue necesaria utilizarla porque nosotros conocíamos el punto de fusión de las dos secuencias simbólicas y en los casos que hemos estudiado (como se verá más adelante) los valores de la D_{JS} en cercanías a este punto son considerablemente elevados comparados con los del resto de la secuencia.

Supongamos que observamos un valor z determinado de D_{JS} . Consideremos la probabilidad de conseguir al azar ese valor o uno inferior, en una secuencia generada a partir de una única distribución. A esa probabilidad la llamaremos Significancia

$$S(z) = Prob [D_{JS} \leq z] \quad (61)$$

La Significancia da la probabilidad de obtener el valor z al calcular la D_{JS} entre las distribuciones de probabilidad de las subsecuencias que están de cada lado del cursor,

suponiendo que la secuencia total (la secuencia de los dos lados de la *Ventana* pegadas) fue generada por una única distribución.

Como la Significancia acumula la probabilidad de todos los valores menores a z , si el valor de la significancia esta muy cerca de 1 significa que es muy poco probable encontrar un valor de D_{JS} mayor z con una única distribución que genere la secuencia total. Por lo tanto z es significativo. Es decir aquí se toman en cuenta las fluctuaciones estadísticas.

El cálculo de la distribución de probabilidad de D_{JS} no lo incluiremos en este apartado. El lector interesado puede consultarlo en [6].

4.3. Medida de disimilitud propuesta por Yang y colaboradores

Hemos decidido describir esta medida de similitud en este capítulo ya que su definición esta intimamente relacionada con la construcción de la misma. Es por eso que se la colocó en un capítulo en donde se analizan los aspectos metodológicos del trabajo.

Cuando decimos “medida” no nos estamos refiriendo al sentido matemático estricto sino mas bien en el sentido de “cantidad” que representa la similitud entre las dos secuencias. Yang y colaboradores en [1] idearon esta función con el propósito de distinguir series temporales generadas por electrocardiogramas, es decir, señales cardíacas. Para poder simplificar el análisis utilizaron un método que mapea una serie temporal en una secuencia simbólica, donde los posibles valores ahora son un alfabeto finito. Esta estrategia es la utilizada por nosotros tambien. Este mapeo se describió detalladamente en la sección 4.1 (“Correspondencia entre serie temporal y cadena binaria”). Como se puede deducir la cadena simbólica que utilizaremos será una cadena binaria.

Teniendo ya la cadena binaria de largo N , $\mathcal{C} = \{c_0, c_1, \dots, c_{(N-1)}\}$, idearon en método de análisis de palabras en secuencias simbólicas esto consiste en:

Se define una *ventana* de largo m que no es más que un conjunto de casilleros, que en nuestro caso será un conjunto de 4 casilleros. Teniendo la *ventana* posicionada, se registran los símbolos que tiene cada uno de los casilleros, a este conjunto de símbolos lo llamaremos *palabra*. Como se sabe hay 2^m posibles palabras. A cada una de ellas las enumeraremos, es decir, la etiquetaremos con un nombre. Nombraremos a w_k a una de esas 2^m posibles palabras. El hecho de enumerarlas no significa que el orden en que lo haga importe es solo para darle un nombre. Lo q importante es que una vez que el nombre w_k se refiera a una cierta palabra no se modifique.

Iremos corriendo nuestra *ventana* de m casilleros y leyendo la palabra que se forma. Esto nos dice que si la *ventana* tiene largo m entonces habrá $\binom{N}{m} - (m - 1)$ palabras por cada cadena de N estados. El primer paso a realizar será leer los primeros m casilleros y verificar que *palabra* es. Luego se mueve la *ventana* de largo m al siguiente casillero, es decir, ahora tendríamos que leer los casilleros (a_1, \dots, a_{m+1}) . Y así sucesivamente iremos contabilizando todas las posibles palabras que puedan haber en la secuencia, es decir, construiremos un ranking de aparición de la palabra w_j en la secuencia S_1 . Cada secuencia tendrá un ranking distinto de palabras.

La medida de similitud presentada por Yang y colaboradores es:

$$D_{YHY}^{(m)}(S_1, S_2) = \frac{\sum_{k=1}^{2^m} |R_1(w_k) - R_2(w_k)| p_1(w_k) p_2(w_k)}{(2^m - 1) \sum_{k=1}^{2^m} p_1(w_k) p_2(w_k)} \quad (62)$$

donde $p_1(w_k)$ es la probabilidad de aparición de la palabra w_k en la secuencia 1, y $R_1(w_k)$ es el ranking que ocupa la palabra w_k en la secuencia S_1 . De la misma manera se define la probabilidad $p_2(w_k)$ y $R_2(w_k)$ para la secuencia S_2 . Como se ve la medida de similitud está definida para un cierto largo de palabra m . Se puede verificar que $D_{YHY}^{(m)}$ da valores que pertenecen al intervalo $[0; 1]$.

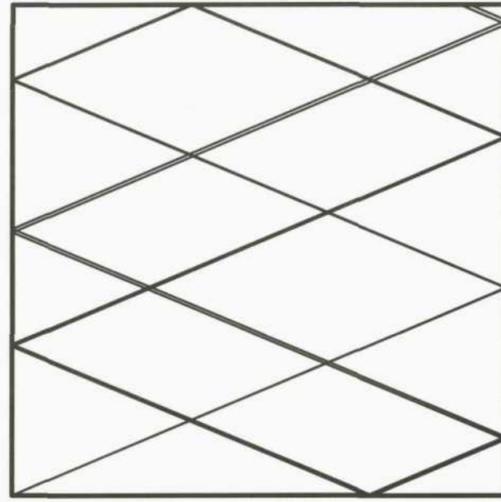


Figura 3. Cuando un sistema no es sensible a las condiciones iniciales las trayectorias se separarán en tiempos muy grandes

5. CAOS, RUIDO Y EL TEOREMA DE WOLD

En esta sección daremos las nociones que creemos necesarias referidas a los procesos Caóticos y Aleatorios para poder interpretar de forma correcta los resultados obtenidos en las siguientes secciones. Primero describiremos de forma general los conceptos de ruido y caos para luego detallar los procesos que elegimos en este trabajo: los ruidos Blancos y Rosa y los procesos caóticos del Mapa Logístico y el Tent-Map. Analizaremos las diferencias entre los procesos caóticos y aleatorios.

A las teorías no cuánticas (mecánica newtoniana y relativista) se las considera deterministas, pero esto no quiere decir que sean predecibles o computables. En este sentido Prigogine cuestiona el término determinista para la física clásica argumentando que el caos y las bifurcaciones impiden a uno determinar el futuro conociendo las condiciones iniciales. Este argumento fue muy discutido por científicos de la época diciendo que Prigogine confunde determinismo con predecibilidad. Y es aquí donde el concepto de caos toma fuerza. El caos está relacionado con la imposibilidad de conocer el devenir del sistema, aún conociendo las condiciones iniciales (Caos Determinista), es decir, el sistema es impredecible. Un sistema caótico-determinista es un sistema muy sensible a la variación de las condiciones iniciales, pero que depende explícitamente de ellas. Para poder interpretar correctamente usaremos como herramienta didáctica un ejemplo realizado por Francisco Torre en [3]. Supongamos que tenemos una mesa de pool y tiramos dos bolas desde el mismo lugar (abajo a la izquierda) pero con ángulos ligeramente distintos. Si tenemos en cuenta que no hay rozamiento y que las bolas rebotan conforme a la ley de reflexión lo que ocurrirá será algo parecido a la Figura 3.

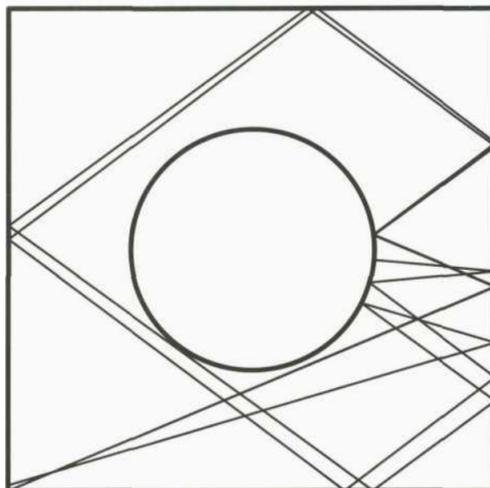


Figura 4. En este sistema pequeñas variaciones en la largada de las bolas produce grandes variaciones en la trayectoria de las bolas

Pero si en vez de tener una mesa de pool normal tuviéramos una mesa de pool de Sinái, es decir, con un obstáculo circular en el medio, las cosas cambiarían drásticamente. Lo que ocurre ahora es que si la bola de pool choca con el obstáculo el ángulo de reflexión es mucho mayor que el de incidencia. Esto producirá que las trayectorias diverjan rápidamente para pequeñas variaciones del ángulo de salida de las bolas, como se ve en la Figura (4) En un sistema donde uno no pueda manejar con absoluta precisión las condiciones iniciales van a ver situaciones parecidas a las comentadas anteriormente, en estos sistemas la predicibilidad será para tiempos muy pequeños. Ya habiendo entendido que el determinismo no implica necesariamente predicibilidad del devenir del sistema, es bueno preguntarse entonces ¿qué es la aleatoriedad? ¿qué relación tiene con el determinismo y con el caos, si es que la tiene?. Se dirá que un sistema es completamente aleatorio si la evolución no depende en absoluto de los valores iniciales del sistema. Siguiendo el ejemplo de la bola y la mesa de pool uno podría considerar un sistema aleatorio de la siguiente manera. Supongamos que nos preocupa el devenir de la bola que tiene un ángulo de largada conocido, pero nos dicen que la bola cuando toca la pared rebota con un ángulo indefinido ya que la pared es de forma desconocida. La trayectoria de la bola no dependerá en absoluto del ángulo de largada ya que uno desconoce la pared. Vemos entonces que si un sistema es aleatorio entonces también es impredecible, aunque de naturaleza completamente distinta al caos-determinista. Estas sutilezas a veces suelen confundir, nosotros intentaremos dejar claro los conceptos.

En el primer capítulo de esta sección se intentará caracterizar y diferenciar los conceptos de determinismo y aleatoriedad, utilizando procesos estocásticos que viven en $L^2[(\Omega, \psi, \mathbb{P})]$, concluyendo en el famoso teorema de Wold. En el siguiente capítulo se intentará profundizar en ciertos tipos de Determinismo, el Caos Determinista, describi-

endo con detalle los que utilizaremos en la tesis.

5.1. Visión del determinismo y alatoreidad desde el proceso estocástico, Teorema de Wold

Para iniciar en el estudio de las cadenas simbólicas caóticas y parcial o completamente aleatorias es bueno comenzar estudiando y caracterizando los procesos estocásticos que subyacen detrás de ellas. Lo que haremos será estudiar procesos estacionarios (lineales) ya que con ellos se puede ver muy bien la diferencia de estos dos conceptos.

Para empezar se definirá formalmente lo que se entienda por Proceso Estocástico Estacionario Determinista y Proceso Estocástico Estacionario Completamente Aleatorio esto ayudará después a definir procesos que tengan características compartidas.

5.1.1. Proceso completamente aleatorio: Ruido Blanco

Definición: RUIDO BLANCO Sea $\{Z_t\}$ un proceso estocástico que vive en un espacio de Hilbert $\mathcal{H} \doteq (\Omega, \psi, \mathbb{P}) \cup \mathbb{G} = L^2[(\Omega, \psi, \mathbb{P})]$ con varianza σ^2 (vemos que es independiente de t) se lo llamará ruido blanco si cumple con las siguientes condiciones:

1. $E(Z_t) = 0$, para todo t entero.
2. $\gamma(h) = \sigma^2 \cdot \delta(h)$.

La segunda es un poco tramposa ya que si el parámetro t es un entero se tendría que escribir de la forma

$$\gamma(h) = \begin{cases} \sigma^2 & ; \text{si } h = 0 \\ 0 & ; \text{si } h \neq 0 \end{cases} \quad (63)$$

Cuando el parámetro sea continuo, se debe tomar $\gamma(h)$ siempre acompañada de la integral, ya que se la tiene que entender en el sentido de distribución. Se suele resumir los parámetros del proceso poniendo

$$\{Z_t\} \sim WN(0, \sigma^2) \quad (64)$$

De esta definición se pueden sacar varias conclusiones importantes. La primera, es que es estacionario ya que el momento es finito y la autocovarianza no depende del parámetro t , y por hipótesis el espacio es L^2 , lo que implicaría $E|Z_t|^2 < \infty$. La segunda, por como está definida la autocovarianza, el ruido blanco es un proceso estocástico incorrelacionado, es decir, hay una independencia (por lo menos en el sentido lineal) entre cada proceso para cada tiempo t . Esto nos dice que el valor que pueda tomar la variable aleatoria Z_i es independiente del valor que haya tomado la variable aleatoria Z_{i-1} . Como habíamos dicho un proceso completamente aleatorio es aquel que su valor no depende en absoluto

del valor que haya tomado el sistema en el momento anterior. También se puede ver que si tomamos la definición de densidad espectral del Apéndice E, tenemos que

$$S(\nu) = \int_{-\infty}^{\infty} \gamma(t)_{WN} dt = \sigma^2 \quad (65)$$

Vemos que es constante, he aquí el porque del nombre blanco. ya que tiene una distribución constante en las frecuencias. También se puede notar que nunca se nombró a la distribución de probabilidad; esto se debe a que las condiciones son sobre el momento y la autocovarianza.

Como se vio en la introducción nosotros tenemos consecuencias importantes por tener solo una sola cadena de muestra. Viendo las complicaciones que pueden surgir en estas condiciones, el camino que elijeremos será el camino inverso. Es decir, si nosotros quisiésemos verificar que una cadena simbólica es un ruido blanco, tendríamos, a priori, que estimar la autocovarianza para poder verificar la independencia de cada tiro, pero para eso tendremos que suponer la estacionariedad y la ergodicidad del proceso. Pero como se dijo en el capítulo 2.1.4, en última instancia uno debe suponer un modelo. En esta tesis se encara el problema de manera inversa, dado una cadena simbólica generada por un ruido blanco deducir propiedades de ella. El hecho de poder extraer propiedades de la cadena simbólica sabiendo el proceso estocástico que la generó nos permitirá luego, comparar propiedades de una cadena generada por un proceso desconosido. Como vimos un proceso completamente aleatorio se lo entiende como uno en el cual cada variable aleatoria no guarda relación con las demás. En nuestra tesis usaremos procesos estocásticos que se los llamarán ruidos también pero no hay que confundirse, estos si pueden tener una autocovarianza distinta a la delta de Dirac, y presentar relación o dependencia con las variables aleatorias para distintos tiempos ($h \neq 0$). Un ejemplo de estos es el ruido rosa, que lo trabajaremos en la siguiente sección.

Determinismo en nuestro espacio de Hilbert $L^2[(\Omega, \psi, \mathbb{P})]$ Otra familia importantes de procesos son los llamados deterministas. Como su nombre lo dice el valor de la variable aleatoria X_i estará determinado o influenciado por las variables aleatorias para tiempos anteriores. Teniendo en cuenta el teorema de la Proyección (Apéndice D) podremos definir matemáticamente estos procesos. Para poder definirlo formalmente es necesario tener en cuenta algunas definiciones. También es bueno aclarar que los procesos considerados en los siguientes apartados son estacionarios, esto se debe a que el resultado del teorema de Wold es para procesos estacionario. Sea nuestro espacio de Hilbert usual, $\mathcal{H} \doteq (\Omega, \psi, \mathbb{P}) \cup \mathbb{G} = L^2[(\Omega, \psi, \mathbb{P})]$ y sea $\mathcal{M}_n = \overline{\text{sp}}\{X_t; -\infty < t \leq n\}$ el subespacio cerrado generado por los elementos que estan entre llaves. Llamaremos σ^2 al error de media cuadrática para un paso, es decir el margen de error que hay en un paso al siguiente, esto es

$$\sigma^2 = E|X_{n+1} - P_{\mathcal{M}_n} X_{n+1}|^2 \quad (66)$$

donde $P_{\mathcal{M}_n}X_{n+1}$ es la proyección del elemento X_{n+1} en el subespacio generado por los elementos X_i ; $-\infty < i \leq n$, es decir, la proyección de x_{n+1} en \mathcal{M}_n . Diremos que un proceso estocástico estacionario (la generalización para procesos no estacionarios no se discutirá en esta tesis) es determinista si y solo si $\sigma^2 = 0$.

Como se puede ver los procesos deterministas son, si se los quiere llamar así, “antagónicos” de los procesos aleatorios como el ruido blanco. El ruido blanco tenía una independencia absoluta respecto de las variables para tiempos anteriores. En cambio, el proceso determinista es generado por estas.

Un resultado importante en el Análisis de Series Temporales es la descomposición de Wold.

Teorema: *Descomposición de Wold*

Si $\sigma^2 > 0$ entonces podemos expresar a cualquier proceso estocástico estacionario X_t como

$$X_t = V_t + \sum_{j=0}^{\infty} \psi_j Z_{t-j} \quad (67)$$

donde

1. $\psi_0 = 1$ y $\sum_{j=0}^{\infty} \psi_j^2 < \infty$
2. $\{Z_t\} \sim WN(0, \sigma^2)$
3. $Z_t \in \mathcal{M}_t$ para cada $t \in \mathbb{Z}$
4. $E(Z_t V_s) = 0$ para todo $s, t \in \mathbb{Z}$
5. $V_t \in \{\bigcap_{n=-\infty}^{\infty} \mathcal{M}_n\}$
6. El proceso estocástico $\{V_t\}$ es Determinista

Los procesos estocásticos $\{X_j\}$ y $\{V_j\}$ y los coeficientes $\{\psi_j\}$ están unívocamente determinados por la ecuación (67) y las condiciones (1–6). Dejaremos de lado la demostración, que vale decir es muy recomendable para aquellos que desean interiorizarse más en el tema, ya que desviará demasiado nuestra atención.

El teorema es un resultado teórico importantísimo, aunque en la práctica, uno debería calcular infinitos términos. Una aproximación válida es aproximar el modelo $MA(\infty)$ por un $MA(q)$, es decir, aproximar un proceso de media móvil infinito por uno finito pero con un q grande. Sabiendo que

$$Z_t = X_t - P_{\mathcal{M}_{t-1}}X_t \quad (68)$$

y que los coeficientes no son más que

$$\psi_j = \frac{\langle X_t, Z_{t-j} \rangle}{\sigma^2} \quad (69)$$

Pero aún así, el hecho de tener una sola secuencia de símbolos nos obliga a aproximar los momentos y los coeficientes con los estimadores correspondientes.

Por más que no utilizaremos este teorema nos surgió la necesidad de expresar este resultado ya que nos muestra las consecuencias de tener un proceso estacionario. *El hecho de poder descomponer un proceso estocástico (ya que para todo tiempo t se pueden descomponer las variables aleatorias X_t) en la suma de dos procesos completamente antagónicos nos afirma de nuevo lo importante que es estudiar los procesos deterministas (y en nuestro caso caóticos) y aleatorios.* Con esta motivación extra que nos dio este teorema abordaremos la tarea de describir los procesos que usaremos en la tesis

5.2. Propiedades de un proceso Random y su diferencia con un proceso Caótico, más específico

En la sección anterior nos ocupamos de interiorizar al lector sobre las nociones de determinismo y aleatoriedad, en las secciones siguientes describiremos los procesos deterministas que usaremos. Como también de estudiar un proceso llamado *Ruido Rosa* que se lo considera como un proceso que no es completamente aleatorio ya no que no es incorrelacionado.

Los procesos deterministas que son utilizados en la Descomposición de Wold son procesos que viven en un espacio de Hilbert, es decir, V_t pertenece a la intersección de subespacios lineales. Por ende, es un proceso lineal. Un caso más problemático son los procesos que no viven en $L^2[(\Omega, \psi, \mathbb{P})]$. Las soluciones de estos procesos no se podrán encontrar con el principio de superposición. A continuación describiremos un breve resumen de las propiedades que tienen los procesos caóticos para luego dar lugar al análisis particular de los casos que usaremos.

5.3. Sistemas Caóticos

Un buen comienzo para el estudio de sistemas caóticos es definir con claridad lo que es un sistema dinámico, ya que los procesos caóticos son un caso particular de ellos. Uno puede entender un Sistema Dinámico como un sistema que evoluciona en el tiempo. Un sistema dinámico puede estar descrito por un conjunto de ecuaciones diferenciales, ecuaciones integrales o ecuaciones en diferencias finitas o una mezcla de todas. Llamaremos *estado* del sistema a los valores que irá tomando mientras el tiempo transcurre y se le dirá *Espacio de Estados* \mathcal{E}_X al conjunto de esos valores. Este puede tener estructura de espacio topológico, o métrico o ser una variedad diferencial, en nuestro caso tomaremos espacios métricos, como la recta real por ejemplo. Utilizaremos el término *camino trazado* para referirnos al conjunto de valores que toma el sistema (o sea los estados) mientras evoluciona dada una condición inicial. Para terminar, se le dirá *Diagrama de Fase* al conjunto de todas las trayectorias posibles. Este nos ayudará a entender el comportamiento del

sistema dinámico dada cualquier condición inicial. Uno se ve tentado a relacionar un sistema dinámico con el concepto de serie temporal. Como pudimos ver una serie temporal no es más que una serie de valores ordenados por un parámetro. En este contexto diremos entonces que un camino (una serie de estados) es una serie temporal, con la particularidad que los valores para un tiempo $t = n$, o sea, x_n dependen explícitamente del valor inicial x_0 . Esto se debe por supuesto a que son sistemas deterministas. Una diferencia entre el análisis de series temporales y el estudio de sistemas dinámicos es que en estos últimos se los estudian en situaciones límite, es decir, trataremos de entender el comportamiento para valores grandes de tiempo.

Definición: SISTEMAS DINÁMICOS LLamamos Sistema Dinámico a la terna $(\mathcal{E}_X, \mathcal{T}, \mathbf{f})$, donde \mathcal{E}_X es el espacio de fase o de estados; \mathcal{T} es el conjunto de tiempo y \mathbf{f} es el flujo (en el caso discreto se lo llama mapeo) del sistema y cumple con la aplicación $f : \mathcal{T} \times \mathcal{E}_X \longrightarrow \mathcal{E}_X$. y con las propiedades:

1. $\mathbf{f}(\cdot, \cdot)$ es continua
2. $\mathbf{f}(0, x) = x$ para toda $x \in \mathcal{E}_X$
3. $\mathbf{f}(t, \mathbf{f}(s, x)) = \mathbf{f}(t + s, x)$ para todo $t, s \in \mathcal{T}$ y $x \in \mathcal{E}_X$

Diremos que el sistema dinámico es discreto si el conjunto \mathcal{T} es un subconjunto de los números enteros, y fluye de modo creciente y de un paso, o sea, $t : 1, 2, \dots, n$. Si \mathcal{T} es \mathbb{R}^+ o \mathbb{R} diremos que es continuo.

Como dijimos anteriormente un sistema dinámico va estar dado por una ecuación en diferencias o una ecuación diferencial depende del caso. Si estamos trabajando con un sistema discreto la función flujo \mathbf{f} será una relación recursiva del tipo

$$x_{k+1} = \mathbf{f}(x_k) \tag{70}$$

Como se puede ver concideramos sistemas *autonomos*, es decir, sistemas que no dependan explícitamente del tiempo. Esto nos permite considerar arbitrariamente el instante de tiempo inicial en $t = 0$ ya que el estado del sistema solo depende del estado anterior y no del instante de tiempo.

Si el sistema fuese continuo podría presentarse de la forma de una ecuación diferencial por ejemplo

$$\frac{dx}{dt} = \mathbf{f}(x(t)) \tag{71}$$

Nosotros nos centraremos en los sistemas discretos ya que trabajamos con secuencias simbólicas. A continuación daremos un conjuntos de definiciones que nos ayudarán en el futuro para poder interpretar correctamente los ejemplos que utilizaremos.

Definición: TRAYECTORIA Ya teniendo definido nuestro sistema dinámico podemos definir, para cada $x \in \mathcal{X}$, una *sucesión* de puntos $\mathcal{E}_{\mathcal{X}}$

$$\gamma(x, \mathbf{f}) = (x, \mathbf{f}(x), \mathbf{f}^2(x), \dots, \mathbf{f}^n(x), \dots) \quad (72)$$

y se define como Órbita al *conjunto* de los puntos de la *trayectoria*, es decir,

$$O(x, \mathbf{f}) = \{x, \mathbf{f}(x), \mathbf{f}^2(x), \dots, \mathbf{f}^n(x), \dots\} \quad (73)$$

donde $\mathbf{f}^k(x) = \overbrace{\mathbf{f} \circ \mathbf{f} \circ \dots \circ \mathbf{f}}^k$, es decir, la composición de la función \mathbf{f} , k veces. A la sucesión (72) se la conoce como *trayectoria de x bajo f* . La interpretación que uno le puede dar a esta sucesión $\gamma(x, \mathbf{f})$ es la siguiente: Supongamos que tenemos un objeto que se mueve bajo cierta leyes deterministas y para el tiempo $t = 0$ se encuentra en la posición x ; entonces en el tiempo $t = 1$, el mismo objeto se encontrará $\mathbf{f}(x)$, y para el tiempo $t = 2$, estará $\mathbf{f}(\mathbf{f}(x))$ y así sucesivamente.

Para el estudio adecuado de los sistemas dinámicos tendremos que analizar todas las órbitas posibles, es decir el mapa $\psi_t : \mathcal{E}_{\mathcal{X}} \rightarrow \mathbb{R}$. También es útil estudiar el estado asintótico de estas órbitas, esto es, estudiar cuando n tiende a infinito. De hecho un comportamiento asintótico que nos caracteriza es el estado estacionario del sistema, es decir, el estado para el cual el sistema permanece en tiempos suficientemente grandes comparados con las constante de tiempo características del sistema. Algunos de estos estados se los conoce como punto de equilibrio (punto fijo), órbitas periódica, eventualmente periódicas, estables y conjuntos límite. Muchos de estos comportamientos asintóticos son casos particulares de un grupo general llamo *Atractor*. Un Atractor es un *Conjunto Invariante*, es por esto que es adecuado a la cronología del conocimiento definir primero lo que es un conjunto invariante.

Definición: CONJUNTO INVARIANTE

Se dice que $\mathcal{V} \subseteq \mathcal{E}_{\mathcal{X}}$, siendo $\mathcal{E}_{\mathcal{X}}$ el espacio de los estados, se dice que es un conjunto invariante del sistema dinámico definido por \mathbf{f} , si $\mathcal{V} = \mathbf{f}(\mathcal{V})$. Esto es, si el resultado de la iteración del conjunto \mathcal{V} es el propio conjunto.

Definición: ATRACTOR

Entendemos intuitivamente al atractor como la región del espacio de los estados al cual convergen las orbitas dadas ciertas condiciones iniciales. Podemos definir a los atractores de un sistema dinámico $\mathbf{f}(\mathcal{X}, t)$, como un subconjunto invariante \mathcal{A} del espacio de los estados $\mathcal{E}_{\mathcal{X}}$ tal que

- Sea D un subconjunto del espacio de los estados, que contiene a \mathcal{A} , entonces ete convergerá al atractor, esto es:

$$\lim_{t \rightarrow \infty} |\mathbf{f}(D, t) - \mathcal{A}| = 0 \quad (74)$$

donde $A \subset D$.

- $D \supset f^t(t, D)$, para $t > 0$.

Vale aclarar que se utiliza como equivalente las notaciones $\mathbf{f}^n(\cdot)$ y $\mathbf{f}(n, \cdot)$, para remarcar la iteración n de nuestro sistema dinámico. El hecho de que puedan coexistir varios atractores da lugar al concepto de *Cuenca de atracción*, que no es más que el conjunto de todas las condiciones iniciales que conducen a un determinado atractor.

Daremos ahora algunos casos particulares de los atractores

Definición: PUNTO DE EQUILIBRIO

Un punto x_e se denomina punto de equilibrio si

$$\mathbf{f}(x_e) = x_e \quad (75)$$

Se puede ver que el punto de equilibrio es un estado estacionario del sistema. Diremos que x_e es estable si para todo $\epsilon > 0$, existe un $\delta > 0$ tal que

$$|x_0 - x_e| \leq \delta \implies |x_n - x_e| \leq \epsilon; \quad \forall n \geq 1 \quad (76)$$

Si el punto no es estable, lo llamaremos simplemente inestable. Un ejemplo simple podría ser la función $\mathbf{f}(x_t) = \frac{1}{x_t}$, aquí el estado $x_t = 1$ es un punto de equilibrio.

Otro concepto importante que deberemos tener en cuenta es la *periodicidad* de las órbitas. Como ya vimos la órbita relacionada al estado x_0 de un sistema dinámico discreto puede escribirse como

$$O(x_0, \mathbf{f}) = \{x_0, \mathbf{f}(x_0), \mathbf{f}^2(x_0), \dots, \mathbf{f}^n(x_0), \dots\} \quad (77)$$

Se dirá que $O(x_0, \mathbf{f})$ es periódica de periodo $p \in \mathbb{N}$ si $\mathbf{f}^p(x_0) = x_0$, es decir, que contendrá a los puntos $\{x_0, \mathbf{f}(x_0), \dots, \mathbf{f}^p(x_0)\}$. Notemos que si $p = 1$ entonces la órbita periódica se convierte en un punto de equilibrio. Se llamará *órbita eventualmente periódica* si para $\mathbf{f}^{p+n}(x_n) = x_n$ para algún $n > 1$ y $p > 2$.

Por último nos ocuparemos en definir a los *conjuntos límites*. Se llamará a z punto límite de la órbita $O(x_0, \mathbf{f})$ si existe una subsecuencia $\{x_{n_k} : k = 1, 2, \dots\}$ de $O(x_0, \mathbf{f})$ tal que $|x_{n_k} - z| \rightarrow 0$ si $k \rightarrow \infty$. El conjunto límite $L(x_0)$ es el conjunto de todos los puntos límites.

Es bueno aunque sea comentar sobre los atractores *Cuasiperiódicos* y los atractores *Extraños*. Los atractores cuasi-periódicos es una colección infinita de puntos concentrados en una región del espacio de las fases (o de estados), estas regiones tienen una geometría sencilla como por ejemplo una circunferencia o un toroide. Los atractores extraños también son una colección infinita de puntos concentrados en una región pero la diferencia es que su geometría es más complicada ya que posee *dimensión fractal*.

Es bueno aclarar diferencias sobre la asintoticidad y eventualidad de los atractores. La trayectoria de un punto concreto de un sistema dinámico puede no alcanzar nunca

exactamente un atractor en un tiempo finito, aunque se sitúe arbitrariamente cerca. Se dice entonces que son *asintóticamente* fijos, periódico, etc. También puede ocurrir que el punto inicial alcance el atractor después de un número finito de iteraciones, en este caso se le dirá *eventualmente* fijo, periódico, etc.

Las condiciones que debe cumplir un sistema dinámico para ser capaz de generar cada uno de los cuatro comportamientos anteriores son bien conocidas tanto para sistemas continuos o discretos. El único atractor posible para un sistema lineal autónomo es un punto fijo, y de uno no autónomo es un punto fijo y un ciclo límite. Como dijimos para que un sistema dinámico presente atractores causi-periódicos o caóticos es necesario que sea no lineal. Aunque esta condición es necesaria pero no suficiente. La existencia de caos depende también de la naturaleza (discreto o continuo) y también de su orden. Con *Orden* nos referimos a la cantidad de variable del sistema para poder describirlo. Por ejemplo, para un sistema autónomo de orden m lo describen m ecuaciones, y para un sistema no autónomo lo describen $(m + 1)$ ya que es necesario saber también el tiempo. Para un sistema continuo de orden 1 el único atractor posible es el Punto Fijo. En el caso de orden 2 (ver teorema de Poincare-Bendisson), los únicos comportamiento asintóticos estables (esto es, persiste a perturbaciones del sistema) son el Punto Fijo y los Ciclo Límites. Esto nos dice que para un sistema continuo no lineal es necesario que sea al menos de orden 3 para que existan atractores caóticos. En el caso discreto es diferente y depende si es invertible o no. Un mapa unidimensional no-invertible es posible encontrar atractores caóticos, como es el caso del Mapa Logístico (que desarrollaremos en capítulos siguientes). Si el mapa es invertible es necesario que sea al menos de orden 2, como es el caso (y que no desarrollaremos en esta tesis) del Mapa de Hénon, para poder generar señales caóticas.

Antes de empezar a describir los casos particulares de mapas que trabajaremos en esta tesis será útil tener la noción de caos. Es por eso que a continuación detallaremos esta definición.

Definición: SISTEMAS CAÓTICOS

Diremos que un sistema dinámico $(\mathcal{E}_X, \mathcal{T}, \mathbf{f})$ es caótico si cumple si presenta las siguientes propiedades

- Es sensible respecto a las condiciones iniciales
- Es topologicamente transitivo
- Sus puntos periódicos son densos en \mathcal{E}_X

Ahora nos ocuparemos en entender cada una de estas propiedades

Definición: TOPOLOGICAMENTE TRANSITIVO

La función $f : \mathcal{E}_X \rightarrow \mathcal{E}_X$, se llama topologicamente transitiva si para cualquier par de conjuntos abiertos $U, V \subset \mathcal{E}_X$ existe $k > 0$ tal que $\mathbf{f}^k(U) \cap V \neq \emptyset$. Es decir, que si

nuestro sistema es topologicamente transitivo, entonces a \mathcal{E}_X no se lo podrá descomponer en dos subconjuntos disjuntos invariantes. Una obserbación importante es que si nuestro sistema tiene una órbita densa es topologicamente transitivo. Esto es si para todo punto $q \in \mathcal{E}_X$ existe una sucesión $n \rightarrow \infty$ tal que $\mathbf{f}^n(p) = q$, para algún $p \in \mathcal{E}_X$.

Definición: SENSIBLE A LAS CONDICIONES INICIALES

La función $f : \mathcal{E}_X \rightarrow \mathcal{E}_X$ tiene dependencia sensible a las condiones iniciales si existe un $\delta > 0$, tal que para cualquier $x \in \mathcal{E}_X$ y para cualquier vecindad N de x existe $y \in N$ y $n \geq 0$ tal que $|\mathbf{f}^n(x) - \mathbf{f}^n(y)| > \delta$.

Definición: PUNTOS PERIÓDICOS DENSOS

Diremos que los puntos periódicos de un sistema dinámico $(\mathcal{E}_X, \mathbf{f})$ son *densos* si para cualquier subconjunto U de \mathcal{E}_X , siempre exista un punto periódico en U .

Teniendo estas definiciones en cuenta uno podría interpretar a los sistemas dinámicos caóticos como impredecibles, irreducibles pero también con un grado de *regularidad*, y esto último es concecuencia de que tiene puntos periódicos densos.

En las últimas décadas hubo grandes avances en la teoría del caos. En 1992 se demostró que la sensibilidad en las condiciones iniciales es consecuencias de las otras dos condiciones. En 1994 se demostró que un sistema dinámico definido en un intervalo por un función \mathbf{f} topologicamente transitiva, es caótico. Y finalmente en 1997 se demostró que un sistema dinámico \mathbf{f} es caótico si y solo si para cualquier subconjunto abierto U y V , existe una órbita periódica que visita ambos.

5.3.1. Tent-Map

En la presente sección reflejaremos la teoría estudiada recientemente en un ejemplo clásico de la literatura, “*La Tienda de campaña*” (también se la conoce como Tent-Map). Haremos primero un análisis general del mapa discreto para luego centrarnos en las regiones caóticas.

Se define al mapa “Tent-Map” como una aplicación $T : [0, 1] \rightarrow [0, \frac{\beta}{2}]$, con $0 < \beta \leq 2$, de la forma,

$$T(x) = \begin{cases} \beta x; & 0 \leq x < 0,5 \\ \beta(1 - x); & 0,5 \leq x \leq 1 \end{cases} \quad (78)$$

Para $\beta = 2$ tenemos la figura 5

Como se puede ver tiene un solo máximo en $x = 0,5$, y al tener dos intervalos, $E_1 = [0; 0,5)$ y $E_2 = [0,5; 1]$, se lo puede escribir de la siguiente manera

$$T(x) = x\beta\mathcal{Y}_{E_1} + \beta(1 - x)\mathcal{Y}_{E_2} \quad (79)$$

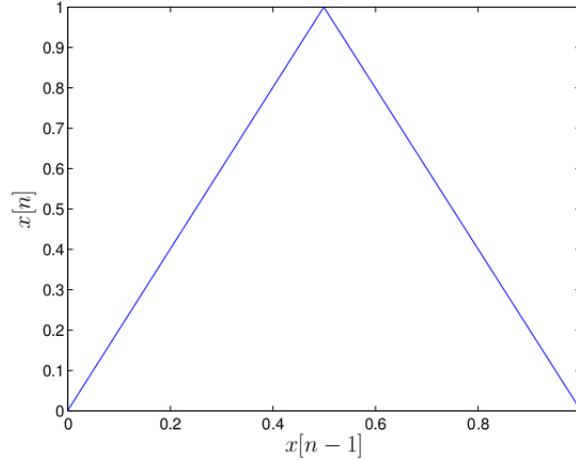


Figura 5. Se puede apreciar la simpleza de la función Tienda

Y si utilizamos la función *signo* también se lo puede expresar como

$$T(x) = x\beta \frac{1 - S(x - 0, 5)}{2} + \beta(1 - x) \frac{1 + S(x - 0, 5)}{2} \quad (80)$$

donde la función signo está definida como

$$S(x) = \begin{cases} 1; & x \geq 0 \\ -1; & x < 0 \end{cases} \quad (81)$$

Teniendo esto podemos reescribir la ecuación del mapa de la forma

$$T(x) = \frac{\beta}{2} (1 - 2x \cdot S(x - 0, 5) + S(x - 0, 5)) \quad (82)$$

Y recordando que $|x - 0, 5| = (x - 0, 5) \cdot S(x - 0, 5)$, la expresión que se usa habitualmente para el “Tent-Map” es

$$T(x) = \frac{\beta}{2} (1 - 2|x - 0, 5|) \quad (83)$$

Como se puede ver el comportamiento de $T(x)$ está definido por el parámetro β . En los mapas unidimensionales a este parámetro se lo suele denominar también como *parámetro de bifurcación* ya que determina el comportamiento asintótico. En función de β se pueden definir distintos comportamientos asintóticos del Tent-Map

1. ($0 < \beta < 1$): El atractor está compuesto por un único punto fijo atractivo en el origen ($x = 0$), de modo que cualquier condición inicial al cabo de algunas iteraciones converge hacia él.
2. ($\beta = 1$): Todos los puntos de la región $[0; 0, 5]$ son puntos fijos, mientras que los de la región $(0, 5; 1]$ se mapean en la región $[0; 0, 5]$ tras una única iteración, siendo por lo tanto puntos eventualmente fijos.
3. ($1 < \beta < \sqrt{2}$): Orbitas cuasi-periódicas. Aunque el espacio de las fases $[0, \beta/2]$, el atractor no ocupa el espacio completo, y permanece confinado en dos estrechas bandas dentro de la región $[\beta(2 - \beta)/2; \beta/2]$, entre las que va saltando alternativamente las muestras de las señales generadas. Ver figura 6
4. ($\sqrt{2} < \beta < 2$): Región caótica. De nuevo el comportamiento asintótico de las secuencias generadas se limita a la región $[\beta(2 - \beta)/2; \beta/2]$, aunque en este caso la cubre por completo y el comportamiento de la señales dentro de la misma es caótico.
5. ($\beta = 2$): Nos encontramos con caos completamente desarrollado, es decir, la secuencia simbólica generada cubre el espacio de las fases por completo. Se puede ver claramente en la figura 7

En la figura 8 se muestra el comportamiento asintótico de la Tent-Map para distintos valores de β . Esta gráfica se ha realizado tomando una condición inicial aleatoria, descartando las primeras 1000 iteraciones y dibujando las 1000 siguientes para cada valor de β . Esta gráfica se la conoce como *diagrama de bifurcación*, y muestra como varía el comportamiento del Tent-Map al modificarse el parámetro de bifurcación.

La demostración de porqué para $\beta = 2$ es un sistema dinámico excede esta tesis, nos conformaremos con decir que para este valor de parámetro se cumplen las tres condiciones dichas al principio. Es decir, existirán un conjunto denso de puntos periódicos. Es por esto que si observamos la Figura 8 los puntos periódicos completan el espacio de fases.

5.3.2. Mapa Logístico

El Mapa Logístico es sin duda el mapa polinómico mas conocido y simple. Este mapa es fruto de la discretización de la ecuación logística propuesta por el biólogo belga Verhulst. Esta se utiliza para modelar el crecimiento/decrecimiento de una población con recursos limitados. Formalmente hablando podemos decir que el Mapa Logístico es una aplicación $f : [0; 1] \longrightarrow [0; 1]$ con

$$f(x) = \lambda x(1 - x) \tag{84}$$

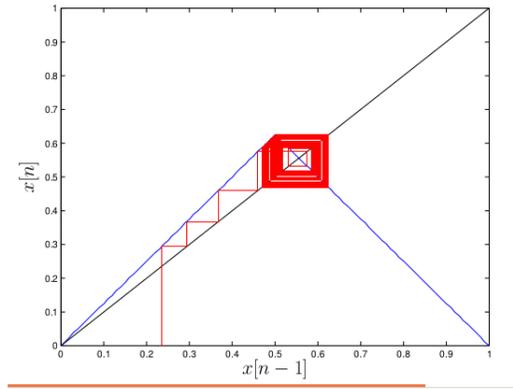


Figura 6. Estado causi-periódico del Tent-Map

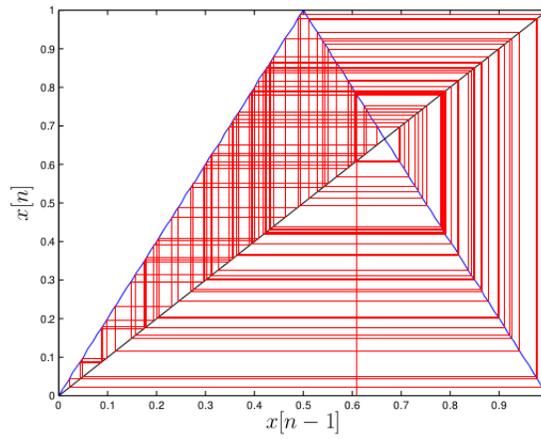


Figura 7. Se puede ver que las orbitas tocan todos los puntos del espacio de las fases

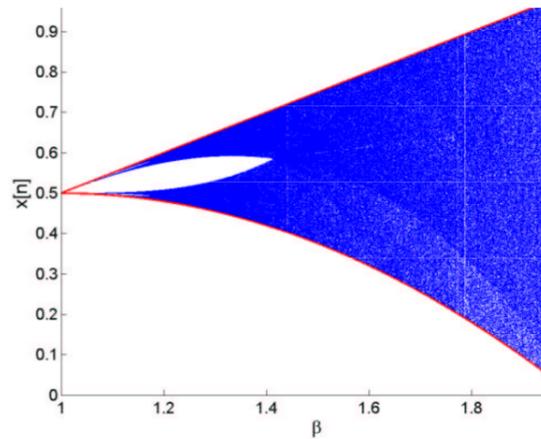


Figura 8. El comportamiento asintótico dependiendo de los diferentes valores del parámetro β

siendo λ un parámetro que varía de 0 a 4. En biología este mapa proporciona un valor normalizado de la población comprendido entre 0 (extinción) y 1 (máxima población posible). Este mapa puede presentar un comportamiento dinámico muy complejo que incluye puntos fijos, ciclo límite de todos los períodos, y caos, dependiendo únicamente del valor del parámetro λ . Se pueden distinguir varios comportamiento variando este parámetro:

- ($0 \leq \lambda < 1$): El atractor es un punto fijo en el origen.
- ($1 \leq \lambda < 3$): El atractor es un punto fijo en $x = 1 - \frac{1}{\lambda}$.
- ($3 \leq \lambda \leq 3,57$): Aparecen atractores periódicos de orden 2^n ($n = 1, 2, 3, \dots$) crecientes a medida que aumenta λ , y cada vez con menor separación entre ellos (es decir, la distancia entre el valor de λ correspondiente al periodo 2^n y al periodo 2^{n+1} es cada vez menor).
- ($3,57 \leq \lambda \leq 4$): Región caóticas intercaladas de regularidad (ciclo límite de período de no potencia de 2). Por ejemplo, en $\lambda = 3,6786\dots$ aparece el primer ciclo límite de período impar, y en $\lambda = 3,8284\dots$ aparece el ciclo límite de período 3, que por el teorema de Sarkovskii (que no detallaremos) implica la existencia de ciclos límite de todos los períodos para algún parámetro del mapa, y garantiza que el mapa sea capaz de generar secuencias caóticas para un cierto rango de su parámetro.

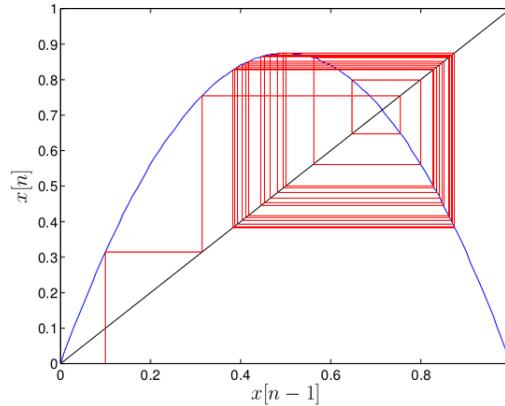


Figura 9. Región crítica, para el cual existen regiones caóticas

- ($\lambda = 4$): Caos completamente desarrollado, y la secuencia simbólica (caótica) generada cubre todo el espacio de fases por completo. Como se puede ver en la Figura 10

En la Figura 11 se muestra el Diagrama de bifurcaciones para el parámetro λ , dada una condición inicial aleatoria. Nosotros, como se especificará en la siguiente sección utilizaremos $\lambda = 4$.

Dejaremos a un lado en este trabajo la demostración de porqué el Mapa Logístico es caótico para $\lambda = 4$, ya que excede nuestras necesidades. Pero si podemos dejar en claro que los puntos periódicos para ese valor de parámetro forman un conjunto denso, como se ve en la Figura 11.

5.4. Ruido Rosa

Como se vio anteriormente el Ruido Blanco es un proceso estocástico $\{X_t\}$ donde cada una de las variables aleatorias (es decir, para cada t) están incorrelacionadas dos a dos. Es decir,

$$\gamma(h) = \sigma^2 \cdot \delta(h); \quad h = t - t' \quad (85)$$

Uno puede interpretar la ecuación anterior de la siguiente manera: Dada una medición al tiempo t de la variable aleatoria X , es decir, una vez obtenido el valor de X_t uno no puede predecir en ningún sentido el valor que tomará la variable aleatoria X_{t+1} . Uno podría considerar informalmente al Ruido Rosa como un caso “intermedio” entre el determinismo y el completo azar. Esto es porque no es completamente incorrelacionado.

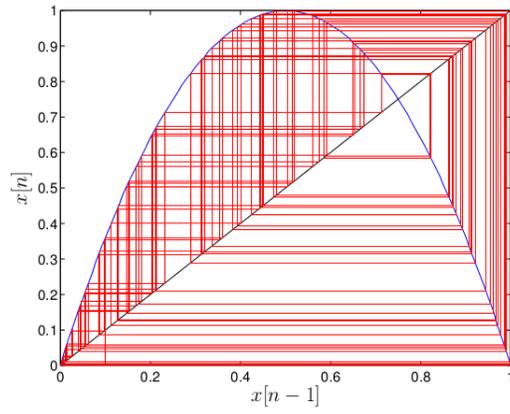


Figura 10. Situación caótica en el mapa logístico

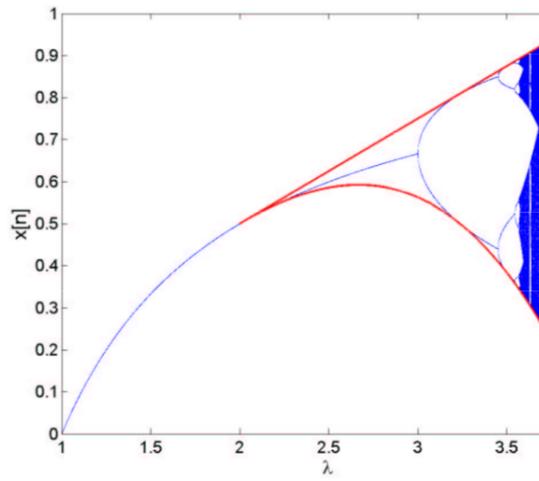


Figura 11. Diagrama de bifurcaciones para distintos valores de λ

El ruido rosa se presenta en diversas áreas de la ciencia por ejemplo, en sistemas físicos se lo puede encontrar en datos meteorológicos o en espectros de cuerpos astronómicos. En los dispositivos electrónicos es muy frecuente encontrarse con esta señal. En el área de la biología se lo encuentra en señales cardíacas, neuronales y también en el análisis estadístico de las secuencias simbólicas de ADN.

Podemos definir entonces al Ruido Rosa haciendo uso del análisis espectral,

$$f(\nu) \propto \frac{1}{\nu} \quad (86)$$

donde ν representa la frecuencia, y $f(\nu)$ la densidad espectral. Como se puede ver la principal diferencia con el ruido blanco es que su densidad espectral depende de la frecuencia. Esto nos lleva a decir que tiene una autocorrelación distinta a la del ruido blanco. Es por esto que se lo correlacionado como ruido correlacionado o “no-gaussiano”.

De la misma manera que con el ruido blanco, uno define al ruido rosa sin recurrir a la distribución de probabilidades y mucho menos al proceso estocástico subyacente. En varias publicaciones (como es el caso de Halley y colaboradores [5]) se intenta llegar a una descripción del ruido rosa mediante *Ruido Autoregresivo* dando como resultado

$$X_t \approx \mu + \sum_{j=1}^K A_t^{(j)} \quad (87)$$

La ecuación (87) se puede interpretar como la combinación de K procesos autoregresivos mas un “residuo” de bajas frecuencias. No utilizaremos la descripción (87) del proceso estocástico subyacente al ruido rosa, pero sirve para ver la estructura complicada que tiene. Nosotros tomaremos secuencias generadas por un algoritmo estandar que describiremos al final del trabajo.

6. RESULTADOS PROPIOS

En la primera parte de esta sección mostraremos los resultados de nuestras simulaciones en la distinción entre cadenas generadas por procesos estocásticos y sistemas caóticos. Es decir, la aplicación del algoritmo de “La Ventana Móvil” al método descrito en la sección 4. En la segunda parte de la presente sección construiremos la Matriz Distancia de la D_{JS} y la compararemos con la Matriz Distancia de la D_{YHY} . La comparación no será de una forma cuantitativa ya que lo que buscamos es la capacidad de distinguibilidad de cada función y no el valor absoluto.

6.1. Diferencias entre caos y ruido con JSD

Como dijimos en la sección 4 se pretende captar, utilizando el algoritmo de “La Ventana Móvil”, variaciones de las propiedades de una serie temporal. Algo relevante del método elegido, y que se ha dicho antes, es que nos permite comparar series que tomen valores en intervalos completamente distintos ya que lo que estamos midiendo son propiedades que no dependen del valor explícito que toma la serie temporal. Buscamos diferenciar series temporales, y la manera de hacerlo será comparar propiedades estructurales, como tasa de crecimiento, decrecimiento y periodicidad por dar algunos ejemplos. La propiedad de periodicidad la pueden tener series temporales que tomen valores completamente distintos ya que es una propiedad que cuantifica un orden de los valores y no el valor en sí.

Una de las características más importante del algoritmo es que utiliza la Divergencia de Jensen-Shannon. Implícitamente lo que estamos diciendo, es que la forma de diferenciar dos series temporales no será comparando los procesos estocásticos subyacentes, si no las distribuciones de probabilidad. El problema ahora será ¿qué distribución de probabilidad calculo?, ya que los espacios muestrales de las series pueden ser muy distintos. Es por esto que buscamos mapear las series temporales a secuencias simbólicas con un mismo alfabeto finito. Hay que recordar que por el método elegido para mapear la probabilidad que estimemos medirá cuán rápido crece o decrece una serie temporal, ya que los símbolos de la cadena alfabetizada representan eso.

En resumen; se mapeó una serie temporal a una secuencia simbólica, usando el método propuesto por Yang y colaboradores ([1]). Una vez obtenida la secuencia simbólica \mathcal{C} (binaria) correspondiente a una cierta serie temporal se la unió con la secuencia binaria de otra serie temporal. Teniendo la cadena binaria total (\mathcal{C}_T), resultado de la unión de las dos cadenas binarias elegidas, construimos la cadena alfabetizada llamada \mathcal{C}_A , con un largo de palabra de $L = 8$. A esta cadena se le aplica el algoritmo de la Ventana, que tendrá un largo $2 \times \mathcal{L}$ (se especificará más adelante el valor de \mathcal{L}). Vale recordar que el algoritmo medirá, utilizando la Divergencia Jensen-Shannon (D_{JS}), la disimilitud entre las distribuciones de probabilidad a cada lado de la Ventana. Este algoritmo nos devolverá los valores que toma la divergencia Jensen-Shannon para distintas posiciones del puntero. La

expectativa es que los valores de la divergencia tomen un máximo cerca del punto de fusión de ambas secuencias.

El Ruido Blanco utilizado es gaussiano y sus parámetros son la media $\mu = 0$ por definición y la varianza $\sigma^2 = 1$ ya que está estandarizado. Se ha confeccionado para que tome valores en el intervalo $(-1, 1)$. Se utilizó el generador de números aleatorios de Matlab.

El sistema dinámico Tent-Map que se usó es para valores del parámetro $\beta = 2$, es decir, cuando es caótico el sistema. Como dijimos anteriormente en la descripción de este sistema dinámico, en el estado de caoticidad se cubrirá todo el espacio de las fases, que es el intervalo $[0; 1]$.

El sistema dinámico Mapa Logístico con el cual trabajamos en esta tesis es con un parámetro de bifurcación $\lambda = 4$, es decir, cuando sus trayectorias son completamente caóticas. Al igual que el sistema Tent-Map, toma valores que se encuentran en el intervalo $[0; 1]$. La serie producida por el proceso estocástico del Ruido Rosa es generada por un algoritmo detallado en el Apéndice H.

El largo de *Palabra* que se usó en todos los casos es $L = 8$.

6.1.1. Ruido Blanco y Tent-Map

En este apartado mostraremos los resultados que al aplicar el algoritmo “La Ventana Movil” a las series temporales producidas por un Ruido Blanco y por el Tent-Map.

Como introducción a la simulación final mostraremos un ejemplo para un largo de cadena de $N = 5000$ es decir, que la serie temporal total es de $N = 10000$. Eligiéremos, aquí como en todos los casos, un largo de Palabra de $L = 8$ y un largo de Ventana de $2 \times \mathcal{L}$ con $\mathcal{L} = 500$.

En la Figura 12 se puede apreciar el gráfico de la D_{JS} producido por el algoritmo. Sobre el eje vertical tenemos los valores que toma la D_{JS} para distintas posiciones de la Ventana. Como se ve tiene un máximo cuando la posición del puntero se encuentra en las cercanías de la mitad de la cadena alfabetizada \mathcal{C}_A , lo cual es satisfactorio. Cuando el puntero se encuentra en la mitad de la cadena alfabetizada tiene de un lado de la Ventana símbolos que representan a una de las series temporales y del otro lado símbolos que representan a la otra serie. Decimos representan, ya que son símbolos pertenecientes a la cadena alfabetizada, y como detallamos en la sección 4 son valores de alfabeto finito, los cuales representan el crecimiento y decrecimiento de la serie temporal. Los valores más grandes de la D_{JS} significa una gran disimilitud entre las dos distribuciones de probabilidad de cada lado de la Ventana.

En la Figura 12 se puede ver que existen ciertas fluctuaciones estadísticas en los valores de D_{JS} , estas se suavizan si aumentamos el largo de cada cadena y de la Ventana, aunque dejando fijo el mismo largo de Palabra.

Si cada serie temporal tiene el largo de $N = 25000$, y tomamos un largo de Ventana de $2 \times \mathcal{L}$, con $\mathcal{L} = 10000$. En la Figura 13 se puede apreciar el cambio en las fluctuaciones

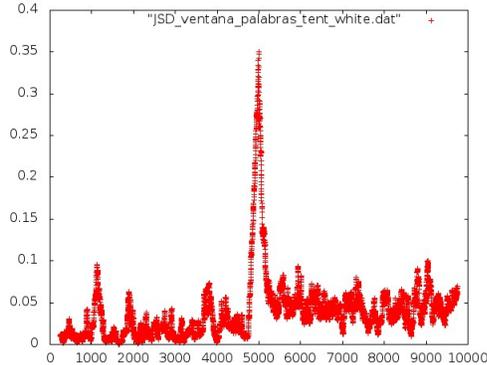


Figura 12. Valores de la D_{JS} para Ruido Blanco y Tent-Map con un largo de cadena de $N = 5000$ y un valor de $\mathcal{L} = 500$.

estadísticas de los valores de la D_{JS} del histograma.

Otra apreciación importante es que en la Figura 12 la D_{JS} se aproxima al máximo de una forma más rápida que en la segunda simulación (Figura 13). Esto se debe a que en proporción con la cadena alfabetizada, la Ventana de la primera simulación es mucho más pequeña que la Ventana en la segunda. En la primera simulación la Ventana es el undécimo del largo de la serie temporal ($\frac{500}{5000}$), y en la segunda, la Ventana es 2,5 veces más chica ($\frac{10000}{25000}$). Esto implica que la Ventana llegará antes (en sentido proporcional) a la otra serie temporal (a la parte de la cadena alfabetizada que representa la otra serie temporal), y por lo tanto los valores de la D_{JS} empezarán a crecer desde antes. La causa de utilizar dos Ventanas distintas se aclararán en los casos en que comparemos Ruido-Ruido y Caos-Caos. Otro punto que se debe mencionar es el valor máximo que toma la D_{JS} en la Figura 12 es menor que el valor para el segundo caso. Esto también está relacionado con lo mencionado anteriormente, ya que el hecho que la Ventana de la segunda simulación sea proporcionalmente más grande hace que la diferencia entre los dos lados de la Ventana sea más marcada. Esto es debido a que la D_{JS} , con el puntero en la mitad de la cadena alfabetizada, compara proporcionalmente más que si estuviéramos en el primer caso.

6.1.2. Ruido Rosa y Tent-Map

En este apartado mostraremos los resultados que resultan de aplicar el algoritmo para dos series temporales, una producida por Ruido Rosa y la otra por el sistema dinámico Tent-Map. Vale aclarar que el valor del parámetro de bifurcación del sistema dinámico Tent-Map que se eligió es $\beta = 2$ de modo que garantizamos una secuencia caótica.

La primera simulación se realizó con un largo de secuencia de $N = 5000$ y un largo de Ventana de $2 \times \mathcal{L}$, con $\mathcal{L} = 500$. Como se ve en la Figura 14 la máxima disimilitud se encuentra en la mitad de la cadena alfabetizada.

Como se vio en el ejemplo anterior podemos suavizar las fluctuaciones aumentando el

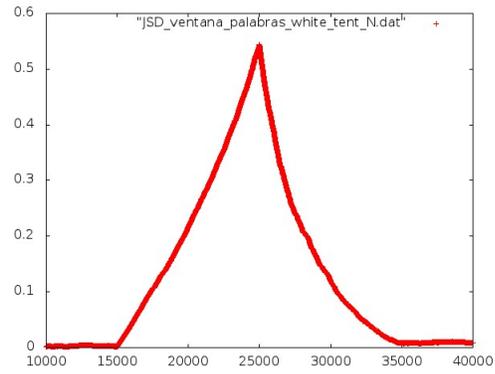


Figura 13. Valores de la D_{JS} para el Ruido Blanco y Tent-Map, con un largo de secuencia de $N = 25000$, y $\mathcal{L} = 10000$.

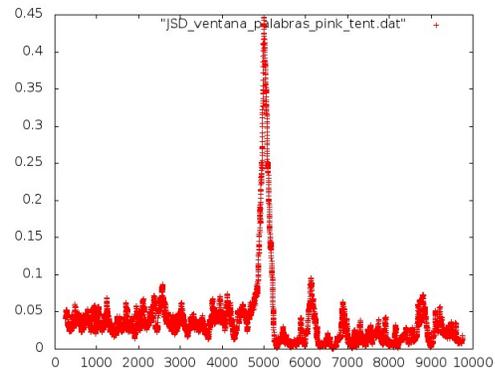


Figura 14. Valores de la D_{JS} para Ruido Rosa y Tent-Map, con un largo de secuencia de $N = 5000$ y $\mathcal{L} = 500$.

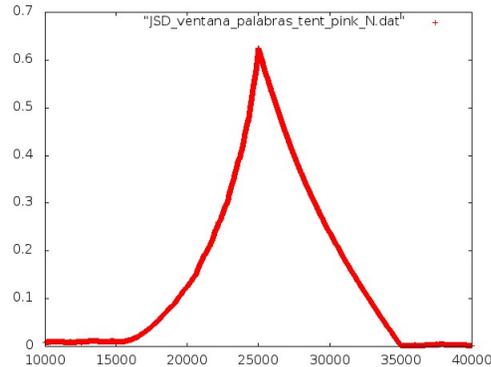


Figura 15. Valores de la D_{JS} para el Ruido Rosa y Tent-Map, con un largo de secuencia de $N = 25000$, y $\mathcal{L} = 10000$.

largo de las cadenas y de la Ventana. Así, si tomamos un largo $N = 25000$ y un largo de Ventana de $2 \times \mathcal{L}$ con $\mathcal{L} = 500$, como se ve en la Figura 15. Como en el apartado anterior la forma en que alcanza el máximo la D_{JS} es mucho más gradual y empieza aproximadamente cuando el puntero se encuentra en la posición $i = 15000$. Esto es consecuencia, de que existe una diferencia proporcional (en referencia al largo de la serie temporal) entre la Ventana de a primera simulación ($\mathcal{L} = 500$) y la segunda ($\mathcal{L} = 10000$). En la primera simulación la Ventana es 10 veces mas chica que la serie temporal y en la segunda 2,5 veces más chica. También como ocurre en el caso “Ruido Blanco y Tent-Map” el valor del máximo de la divergencia cambia para un ancho y otro. La causa de esto es que la divergencia, con el puntero posicionandose en la mitad de la cadena alfabetizada, toma proporcionalmente más valores (de cada lado) en la segunda simulación.

6.1.3. Ruido Blanco y Ruido Rosa

En este apartado mostraremos los resultados de aplicar nuestro algoritmo para dos series temporales, una producida por Ruido Rosa y la otra por Ruido Blanco. En la primera simulación se tomaron series de un largo de $N = 5000$ y una Ventana de $2 \times \mathcal{L}$ con $\mathcal{L} = 500$. Los valores que tomó la D_{JS} para distintas posiciones del puntero se pueden ver en la Figura 16. Aquí se puede apreciar que el algoritmo no es capaz de distinguir cuando pasamos de una serie a otra. Una interpretación posible es que los dos ruidos son localmente parecidos y necesitamos aumentar la escala para notar la diferencia. Es decir, necesitamos tomar una mayor cantidad de símbolos para que la D_{JS} pueda diferenciar entre las dos distribuciones de probabilidad. Pero si solo aumentamos el largo de la cadena y no el largo de la Ventana el resultado va a ser el mismo ya la divergencia tomará distribuciones de probabilidad calculadas con la misma cantidad de símbolos y la diferencia entre esas distribuciones no se notará al moverse el puntero. La segunda simulación se hizo con un largo de serie de $N = 25000$ y un largo de Ventana de $2 \times \mathcal{L}$,

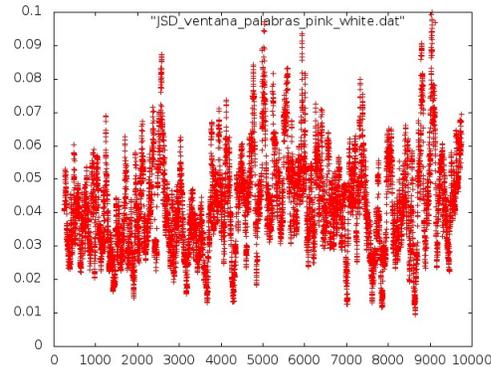


Figura 16. Valores de la D_{JS} para Ruido Rosa y Ruido Blanco, con un largo de secuencia de $N = 5000$ y $\mathcal{L} = 500$.

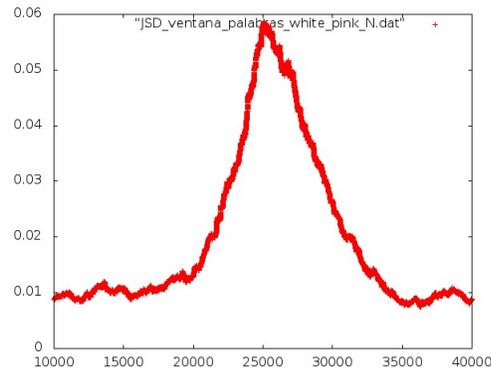


Figura 17. Valores de la D_{JS} para el Ruido Rosa y Ruido Blanco, con un largo de secuencia de $N = 25000$, y $\mathcal{L} = 10000$.

con $\mathcal{L} = 10000$. En la Figura 17 se puede apreciar el resultado. Es bueno aclarar que la diferencia proporcional (con respecto al largo de la serie) que existen entre la Ventana de la primera simulación y de la Ventana de la segunda simulación es porque necesitábamos una ventana de por lo menos un orden de magnitud más grande. Por supuesto esto requiere de mayor esfuerzo computacional.

Un apreciación importante de la Figura 17, es el valor del máximo de la d_{JS} de un orden menos de magnitud que en los casos que veníamos tratando (“Ruido Blanco y Tent-Map” y “Ruido Rosa y Tent-Map”). Esto nos muestra que sus estructuras locales son muy parecidas, y se necesitamos analizarlas de forma más global. Hay que recordar que el termino “estructura” hacer referencia a la forma en que crece o decrece la serie temporal ya que estamos haciendo estadística sobre la cadena alfabetizada.

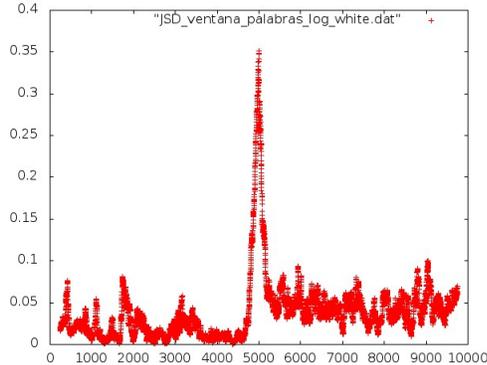


Figura 18. Valores de la D_{JS} para Mapa Logístico y Ruido Blanco, con un largo de secuencia de $N = 5000$ y $\mathcal{L} = 500$.

6.1.4. Ruido Blanco y Mapa Logístico

En este apartado mostraremos los resultados de aplicar el algoritmo llamado “La Ventana Móvil” para dos series temporales, una producida por Ruido Blanco y la otra por el sistema dinámico Mapa Logístico. Como dijimos se usara un valor de parámetro de $\lambda = 4$ para que sea un mapa caótico. Este caso es bastante similar a los dos primeros (“Ruido Blanco y Tent-Map” y “Ruido Rosa y Tent-Map”) ya que aplicamos nuestro algoritmo entre una cadena de ruido y otra caótica.

Como se hizo en todos los casos anteriores primero se realizó una simulación con un largode secuencia de $N = 5000$ y un largo de Ventana de $2 \times \mathcal{L}$, con $\mathcal{L} = 500$. El algoritmo pudo detectar satisfactoriamente el cambio de una serie temporal a otra, como se ve en la Figura 18.

En la segunda simulación se utilizó un largo de cadena de $N = 25000$ y un largo de Ventana de $2 \times \mathcal{L}$, con $\mathcal{L} = 10000$, como se ve en la Figura 19. Como sucedió en los primeros casos (“Ruido Blanco y Tent-Map” y “Ruido Rosa y Tent-Map”) la forma en que la D_{JS} se acerca al máximo es más lenta. También en este caso han desaparecido las fluctuaciones.

6.1.5. Tent-Map y Mapa Logístico

Otro ejemplo estudiado es el que resulta de concatenar dos mapas caóticos: el Tent-Map y el Mapa Logístico. Como veremos existen ciertas semejanzas en el que se concatenan dos ruidos (“Ruido Rosa y Ruido Blanco”).

La primera simulación que realizamos es con un largo de cada serie de $N = 5000$ y un largo de Ventana de $2 \times \mathcal{L}$, con $\mathcal{L} = 500$. Como se puede ver en la Figura 20 el algoritmo no puede registrar el cambio de una serie temporal a otra. Esto nos dice que en cierta escala de Ventana se comportan parecidas. Por lo tanto, su estructura es semejante. Para poder distinguir las diferencias entre las dos serie tendremos que aumentar la escala

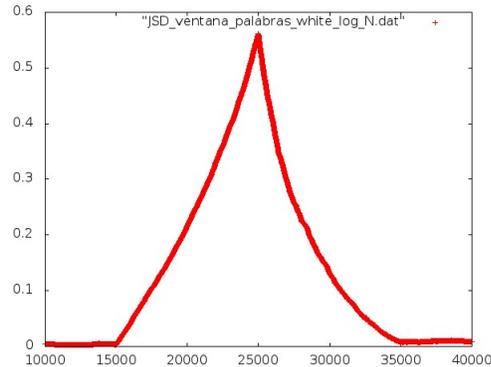


Figura 19. Valores de la D_{JS} para el Ruido Blanco y el Mapa Logístico, con un largo de secuencia de $N = 25000$, y $\mathcal{L} = 10000$.

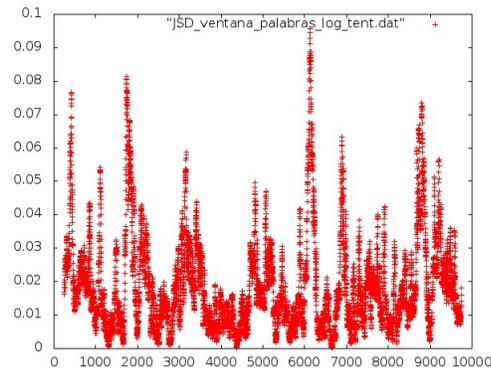


Figura 20. Valores de la D_{JS} para Mapa Logístico y Tent-Map, con un largo de secuencia de $N = 5000$ y $\mathcal{L} = 500$.

de análisis. Esto nos llevará a aumentar la cantidad de valores de la serie, y además aumentar el largo de la *Ventana*. Esto último es fundamental ya que las distribuciones de probabilidad (de la cadena alfabetizada) se distinguirán en una escala mayor y por lo tanto la D_{JS} necesitará mayor cantidad de símbolos que comparar. El resultado se muestra en la Figura 21.

Como se puede ver el máximo es del orden que el conseguido para el caso “Ruido Blanco y Ruido Rosa”, y de un orden de magnitud menor comparado con los demás casos.

6.1.6. Ruido Rosa y Mapa Logístico

Para concluir aquí analizamos el caso en que concatenamos un ruido rosa con un mapa logístico. Es de esperar entonces que el comportamiento del algoritmo sea similar que en los casos anteriores en donde se trató un ruido con un mapa caótico. En la primera

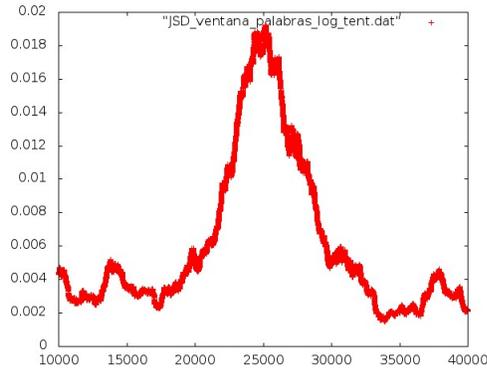


Figura 21. Valores de la D_{JS} para el Tent-Map y el Mapa Logístico, con un largo de secuencia de $N = 25000$, y $\mathcal{L} = 10000$.

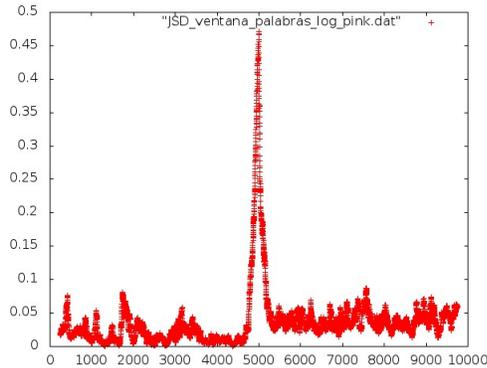


Figura 22. Valores de la D_{JS} para Mapa Logístico y Ruido Rosa, con un largo de secuencia de $N = 5000$ y $\mathcal{L} = 500$.

simulación introducimos series de largo $N = 5000$ cada una, con un largo de *Ventana* de $2 \times \mathcal{L}$, con $\mathcal{L} = 500$. Los distintos valores que toma la divergencia Jensen-Shannon se grafican en la Figura 22. Como ocurrió en los casos en que analizaron Caos con Ruido, el algoritmo ha podido distinguir la transición de una serie a otra. Cuando el puntero se encuentra cerca de la mitad de la cadena alfabética las distribuciones de probabilidad de cada lado son diferentes.

En la segunda simulación se incremento el largo de cada serie a $N = 25000$ y el largo de la *Ventana* en $2 \times \mathcal{L}$, con $\mathcal{L} = 10000$. En la Figura 23 se ven graficados los valores de la D_{JS} para las distintas posiciones del puntero. Como se puede ver la forma en que llega al máximo es más paulatina que en la primera simulación. En esta segunda simulación el máximo es más grande que en la primera, aunque los dos son del mismo orden de magnitud.

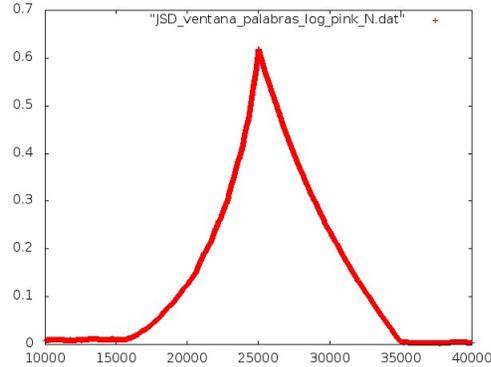


Figura 23. Histograma de la D_{JS} para el Ruido Rosa y el Mapa Logístico, con un largo de secuencia de $N = 25000$, y $\mathcal{L} = 10000$

6.2. Matriz Distancia

En esta sección mostraremos otra faceta de la investigación. Como se aclaró en las motivaciones del trabajo, nos interesa establecer una medida de disimilitud entre secuencias simbólicas generadas tanto por sistemas aleatorios o caóticos. En otras palabras, buscamos una matriz que en cada elemento contenga el valor de la medida de disimilitud entre esas dos secuencias. Se lo podría interpretar como la búsqueda de un catálogo de distancias entre secuencias simbólicas, y así comparar cuanto se asemeja una secuencia a otra. Nuestra medida de disimilitud es la Divergencia Jensen-Shannon, D_{JS} .

A diferencia de los resultados anteriores que aplicábamos la D_{JS} a la parte de la secuencia que se encontraba dentro de la *Ventana*, ahora lo haremos con la secuencia completa. El método es el siguiente: Primero tomamos una serie temporal de cualquiera de los cuatro procesos elegidos (Ruido Rosa, Ruido Blanco, Tent-Map y Mapa Logístico). A cada una de ellas, y por separado, las mapeamos a cadenas binarias (\mathcal{C}) con el método de Yang y colaboradores [1], desarrollado en la sección 4. Luego elegimos un tamaño de Palabra y construimos la cadena alfabetizada de la misma forma que en la sección 4. Teniendo esto, estimamos la probabilidad de aparición de cada elemento de la secuencia alfabetizada por la frecuencia relativa. Luego construimos la distribución de probabilidad. Con estas distribuciones de probabilidad podemos calcular la divergencia D_{JS} de dos secuencias distintas.

El resultado obtenido es

$$\begin{bmatrix} 0 & 0,385836 & 0,002489 & 0,342357 \\ 0,385836 & 0 & 0,391619 & 0,016423 \\ 0,002489 & 0,391619 & 0 & 0,340872 \\ 0,342357 & 0,016423 & 0,340872 & 0 \end{bmatrix}$$

La fila y columna 1 se refiere al Mapa Logístico. La fila y columna 2 se refiere al Ruido

Rosa. La fila y columna 3 se refiere a el Tent-Map. La fila y columna 4 se refiere al Ruido Blanco. Un ejemplo sería el elemento de matriz ($i = 2, j = 3$), que es la D_{JS} entre las distribuciones de probabilidad de la secuencia alabetizada del Ruido Rosa con la de la secuencia alfabetizada del Tent-Map. El número de caracteres de cada serie temporal es de $N = 5000$ y con un largo de Palabra $L = 8$.

Analizando los valores e la Matriz Distancia se puede ver que los órdenes de magnitud se mantienen si comparamos los valores con los máximos de la sección anterior, exepto la distancia entre T.Map y M.Logístico que es de un orden menor comparado con el máximo encontrado con el algoritmo “La Ventana Móvil” de la sección anterior.

Una observación importante es que de los valores de la matriz existen dos que son de ordenes de magnitud menores que los demás, estos son 0,02489, referido a T.Map-M.Logístico, y 0,016423, referido a R.Rosa-R-Blanco. Esto es consecuencia que estamos comparando una secuencia caótica con otra caótica, y una secuencia de ruido con otra de ruido. La simetría de la matriz se debe a que D_{JS} es una distancia.

6.2.1. Otra propuesta de distancia

En este apartado construiremos una matriz distancia usando la propuesta por Yang y colaboradores [1]. Como se dijo al inicio del trabajo, el objetivo no es compararla cuantitativamente con la matriz del apartado anterior, ya que los valores pueden estar defasados. Es decir, buscamos hacer un análisis cualitativo entre estas dos distancias. La matriz obtenida es,

$$\begin{bmatrix} 0 & 0,000187 & 0,000299 & 0,000234 \\ 0,000187 & 0 & 0,000187 & 0,000059 \\ 0,000299 & 0,000187 & 0 & 0,000207 \\ 0,000234 & 0,000059 & 0,000207 & 0 \end{bmatrix}$$

El ordenamiento de las filas y columnas es el mismo que el caso anterior.

Una observación importante es que la distancia entre la secuencia del Mapa Logístico y la de la secuencia del Tent-Map es la mayor en el caso de Yang y es la menor en la Matriz construida con la D_{JS} .

7. CONCLUSIONES

En esta Tesis presentamos un método algorítmico basado en la Divergencia Jensen-Shannon para la caracterización de series temporales. El método de la “Ventana Móvil”, como así lo llamamos, diferencia a las series temporales por su estructura o mejor dicho, por la forma en que crecen o decrecen. Esta manera de ver a las series temporales nos permitió comparar series de naturaleza muy distinta como las series generadas por un mapa caótico o un ruido. Para eso tuvimos que transformalas a secuencias de un espacio muestral finito y equivalente entre ambas. Hemos podido distinguir cada una de las transiciones posibles entre las serie trabajadas en la Tesis, lo que muestra la robustez del método.

De todas las simulación que se hicieron se pueden armar dos grupos. El primer grupo corresponde a las simulaciones hechas con una secuencia simbólica de ruido y otra de caos. Dentro de este grupo se encuentran las siguientes simulaciones: R.Rosa-T.Map, R.Rosa-M.Logístico, R.Blanco-T.Map y R.Blanco-M.Logístico. El segundo grupo corresponde a las simulaciones hechas entre secuencias de ruido con ruido , y entre secuencias de caos con caos. A este grupo lo componen las simulaciones : R.Rosa-R.Blanco y T.Map-M.Logístico. Existen similitudes entre los elementos del grupo, en los dos casos.

En el primer grupo de simulaciones el algoritmo pudo distinguir el cambio de una serie temporal a otra, más allá de la escala que se use, es decir, más allá del largo de la Ventana. Esto refleja que las series temporales de los ruidos son estructuralmente distintas a las de caos. Decimos “estructuralmente” ya que lo que estamos analizando es la cadena alfabetizada, y cada elemento de la misma representa el grado de crecimiento de la serie en un intervalo 8 casilleros (este intervalo es el largo de la Palabra, que es $L = 8$). De alguna manera se puede interpretar que la forma en que crecen y decrecen los valores en una serie temporal de ruido es distinta a la forma en que crecen y decrecen en un serie temporal hecha por un mapa caotico. Cuando decimos distinta tenemos que tener en cuenta que nuestra forma de cuantificar las disimilitudes entre las distribuciones de probabilidad de la cadena alfabetizada es por medio de la Divergencia Jesen-Shannon. Como se pudo ver a lo largo de las simulaciones de este grupo, todos los máximos valores que tomaba la D_{JS} eran del mismo orden de magnitud, y un orden de magnitud mayor que los máximos del otro grupo.

En el segundo grupo las cosas son distintas. Las primeras simulaciones para los dos casos en este grupo no pudieron distinguir la transición de una serie temporal a otra. como se ve en las Figuras 20 y 16. Uno podría interpretar estos resultados de la siguiente manera. A una cierta escala (es decir, para un largo de Ventana de $\mathcal{L} = 500$) las forma en que crecen y decrecen los valores de las series temporales son similares. Por lo tanto, las distribuciones de probabilidad de la cadena alfabetizada serán similares. Este problema fue solucionado aumentando la escala (es decir, aumentar el largo de la Ventana a $\mathcal{L} = 10000$) pudiendo así tomar una mayor cantidad de valores de la cadena alfabetizada. Si lo analizamos de esta manera uno podría decir que hay muy pocos “lugares” dentro de las series temporales

en donde crecen y decrecen de diferente manera, es por eso que necesitamos tomar la mayor cantidad de valores para comparar. Aún así el máximo valor de la D_{JS} (en los dos casos, R.Blanco-R.Rosa y T.Map-M.Logístico) es de un orden de magnitud menor a los máximos encontrados en las simulaciones del otro grupo. Situaciones similares, en cuanto a la capacidad de distinguibilidad entre PE y SC ha sido reportado por otros autores [17]

Por otro lado, se construyeron dos Matrices Distancias, la referida a la Divergencia Jesnsen-Shannon y la utilizada por Yang y colaboradores. Como se pudo ver, la matriz de D_{JS} tiene en cada elemento tres ordenes de magnitud mayor que los que los obtenidos para la matriz de D_{YHY} . Esto nos permite decir que el error en las fluctuaciones estadísticas en el caso de D_{YHY} afectará en mayor medida los resultados.

8. Apéndice A: Repaso de Teoría de Probabilidades

En este apéndice se darán conceptos básicos y complementarios en teoría de probabilidades. Su uso será frecuente a lo largo de la tesis.

Supongamos que estamos haciendo un determinado experimento (por ejemplo, lanzar tres monedas a la vez). Los resultados que se irán dando son las llamadas muestras de este experimento. El conjunto de todas las muestras se lo conoce como Espacio Muestral, Ω . Para nuestro ejemplo el espacio muestral es:

$$\{ xxx ; xxc ; xcc ; ccc ; cxx ; ccx ; cxc ; xcx \}.$$

Aunque sea un experimento simple los conceptos que hay detrás son difíciles. Es decir, si buscamos entender formalmente el experimento necesitamos mayor robustez en nuestro conocimiento en teoría de probabilidades. A continuación se darán conceptos básicos para poder formalizar nuestros experimentos estadísticos como el que vimos recién. Este marco teórico nos servirá como pie de apoyo para el desarrollo de la tesis, que constará con experimentos teóricos más complicados que el anterior. La meta a corto plazo será desarrollar los conceptos de Espacio de Probabilidad y Variable aleatoria.

Ahora bien, para poder definir adecuadamente nuestro espacio de probabilidades es necesario poner nuestra atención a los subconjuntos que se puedan formar de nuestro Espacio Muestral Ω . Pero no cualquier subconjunto nos interesa, solo la familia de subconjuntos σ - *algebra*.

Definición: (*σ - algebra*) Dado un espacio muestral Ω se llamará *σ - algebra* a la familia (\mathcal{B}) de subconjuntos de Ω que cumplan con:

1. Si $A \in \mathcal{B}$ entonces también lo está su complemento $A^c = \Omega - A$, es decir, si A es un subconjunto de Ω también lo es su complemento.
2. Sea A_n un colección contable de elementos de \mathcal{B} (osea subconjuntos de Ω), entonces $\bigcap A_n$ y $\bigcup A_n$ también existen en \mathcal{B} , y por ende son subconjuntos de Ω .

Nosotros queremos poder definir sobre estos subconjuntos (también se los puede tratar como eventos) del espacio muestral una medida de probabilidad. Es decir, que a los elementos de \mathcal{B} los llamaremos conjuntos medibles y al par (Ω, \mathcal{B}) le diremos espacio medible ya que sobre ellos se definirá la medida de probabilidad.

Las funciones medibles son aquellas que parten de un espacio medible y su imagen es otro espacio medible, esto es

$$f : (\Omega, \mathcal{B}) \rightarrow (\Gamma, \mathcal{B}') \tag{88}$$

Definición: (*Medida*)

Una medida μ es una clase especial de función medible que mapea un espacio medible a $[0, \infty]$ y que verifica que:

1. La medida del espacio vacío es cero, $\mu(\emptyset) = 0$
2. Sea A_n una sucesión de subconjuntos disjuntos, entonces

$$\mu\left(\bigcup A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$$

Dadas estas definiciones podemos ahora definir nuestra medida de probabilidad.

Definición: (*Medida de Probabilidad*)

Una medida P de probabilidad en un σ -*algebra* de subconjuntos de Ω , es una función que asigna un valor $P\{\beta\} \geq 0$, para cada elemento β de \mathcal{B} (o sea cada subconjunto de Ω), tal que $P\{\Omega\} = 1$. Y para cada colección enumerable de subconjuntos de Ω se tenga que;

$$P\left(\bigcup A_n\right) = \sum P(A_n) \quad (89)$$

Dada esta definición formal surge la pregunta de como calcularla, dado un cierto experimento (por ej. el que se dio al principio de las tres monedas). Esta pregunta se resolverá aproximandola con la frecuencia de aparición de cierto subconjunto de Ω (evento). La validez de esta aproximación se discutirá en los capítulos de la introducción. Este compendio de definiciones tiene como fin introducir dos conceptos importantes para el desarrollo de la tesis, el Espacio Probabilístico y la Variable Aleatoria (VA) . Teniendo estos dos conceptos podremos desarrollar gran parte de la teoría de Series Temporales.

Definición: (*Espacio Probabilístico*)

Un espacio probabilístico no es más que un Espacio Medible que además contiene todas las funciones de probabilidad. Esto es; el trio $(\Omega, \mathcal{B}, \mathbb{P})$, donde Ω es el Espacio Muestral, \mathcal{B} la σ -*algebra* de ese espacio y \mathbb{P} es la medida de probabilidad definida en ese σ -*algebra*.

Dado un espacio muestral podría surgir de nosotros una "pregunta" sobre los datos obtenidos, en nuestro ejemplo anterior podría surgir asignarle a cada "tirada de las tres monedas" (o sea un punto de Ω) el valor de caras que tiene ese experimento, en este caso sería:

$$\begin{aligned} xxc &\rightarrow 1 \\ ccc &\rightarrow 3 \\ cxx &\rightarrow 1 \\ &\dots \end{aligned}$$

Cada asignación del número responde a la pregunta hecha. Pero no es la única "pregunta" que uno puede hacerle, es decir, para un espacio muestral pueden haber muchas

”preguntas”. La asignación del número 3 al evento $\zeta\zeta\zeta$ se da mediante una función que llamaremos Variable Aleatoria.

Definición: (*Variable Aleatoria*)

Una variable aleatoria X es una función que asigna a cada punto del Espacio Muestral Ω un valor real. Esto no es más que:

$$X : \Omega \rightarrow \mathfrak{R} \quad (90)$$

El conjunto de valores que puede tomar X es $\{x_1; x_2 \dots\}$, y en esta tesis se lo llamará Rango de la Función (o variable) Aleatoria X .

La Variable Aleatoria al estar definida sobre cada punto de Ω es mensurable, es decir, tiene asignada una distribución de probabilidad. Esto significa que existe

$$P\{X = x_j\} = f(x_j) ; \forall j \text{ con } j : 1; 2; \dots \quad (91)$$

Es imprescindible poder definir el concepto de momento de una variable aleatoria;

Definición: (*Momento de una Variable Aleatoria*)

Sea X definida sobre un espacio probabilístico $(\Omega, \mathcal{B}, \mathbb{P})$ y $\omega \in \Omega$ tenemos que;

Se define Varianza de una Variable aleatoria X como;

$$Var(X) \doteq Cov(X, X) = E[(X - E(X))^2]. \quad (92)$$

$$E(X) \doteq \sum_{\omega \in \Omega} X(\omega)P(X(\omega)) \quad (93)$$

Definición: (*Distribución finito-dimencional*)

Si $\{X_t\}$ es un proceso estocástico se llamará distribuciones finito-dimensionales a las districiones de probabilidad conjunta de las subfamilias de $(X_t)_{t \in T}$. Una subfamilia sería el conjunto que comprende $(t_1 = 3, \dots, t_n = n + 3) \in \mathbb{T}$, o sea (X_3, \dots, X_{n+3}) , entonces la distribución de probabilidad

$$P_{t_1, \dots, t_n}(x_3, \dots, x_{n+3}) = P(X_3 = x_3, \dots, X_{n+3} = x_{n+3}) \quad (94)$$

es una distribución finito-dimensional, y donde los valores $(x_3, \dots, x_{n+3}) \in \mathfrak{R}$. La distribución de un proceso está caracterizada por el conjunto de todas las distribuciones finito-dimensionales. En este caso se particularizó dando un valor específico a los parámetros $\{t_k\}$, pero obviamente se cumple para cualquier subconjunto de \mathbb{T} , pero el siguiente desarrollo exigiremos que $(t_1 <, \dots, < t_n)$.

Se puede ver que las distrubuciones finito-dimensionales del proceso estocástico $\{X_t\}$ cumplen dos propiedades muy importantes:

1. Si se tiene

$$P_{t_{k(1)}, \dots, t_{k(n)}}(x_{k(1)}, \dots, x_{k(n)}) = P_{t_1, \dots, t_n}(x_1, \dots, x_n); \text{ par cualquier permutacion } k(1), \dots, k(n).$$

2. Si se tiene

$$P_{t_1, \dots, t_n, t_{n+1}}(x_1, \dots, x_n, \mathfrak{R}) = P_{t_1, \dots, t_n}(x_1, \dots, x_n);$$

es decir, le arrego una coordenada sobre la cual no hay ninguna restricci3n real.

Teorema (Kolmogorov)

Supongase que tenemos un conjunto de distribuciones finito-dimensionales $P_{t_1, \dots, t_n}; t_i \in \mathbb{T}$, que satisface las condiciones enumeradas recientemente; entonces existe un espacio de probabilidad $(\Omega, \psi, \mathbb{P})$ y un proceso estocástico $X : T \times \Omega \rightarrow \mathfrak{R}$.

La familia de distribuciones finito-dimensionales es un concepto de suma importancia para entender en alg3n sentido el proceso, ya que en la practica con un n3mero considerable de observaciones un puede determinar las P_{t_1, \dots, t_n} , y en general nada m3s se puede obtener de las observaciones.

El hecho de no tener explícitamente el Proceso Estocástico, no obliga a pensar propiedades generales de los mismos para poder clasificarlos de alguna manera.

Definiciones Sea $\{X_t; t \in \mathbb{T}\}$ y $\{Y_t; t \in \mathbb{T}\}$ dos procesos estocásticos definidos sobre el mismo espacio probabilístico $(\Omega, \psi, \mathbb{P})$, decimos :

1. X_t y Y_t son iguales si $X_t(\omega) = Y_t(\omega)$ para todo $t \in \mathbb{T}$ y $\omega \in \Omega$.
2. X_t y Y_t son estocasticamente equivalentes si $P(X_t = Y_t) = 1$ para todo $t \in T$.

Variable Aleatoria Indicadora Es de mucha ayuda para las siguientes demostraciones poder definir la **VA** Indiadora, I_β . Sea α un evento cualquiera, es decir, $\beta \in \Omega$ es un elemento del σ -algebra, entonces se define

$$I_\beta = \begin{cases} 1 & ; \text{si y solo si ocurre } \beta \\ 0 & ; \text{si y solo si NO ocurre } \beta \end{cases} \quad (95)$$

Es decir; los eventos posible dados esta variable aleatoria son $I_\beta = 1$ y $I_\beta = 0$. El supraíndice β se refiere al evento ÉXITO, osea el que da 1. El evento ÉXITO, es en sí cualquier evento o cualquier subconjunto de Ω que deseemos diferenciar.

La variable aleatoria indicadora (también conocida como función indicadora) es de gran ayuda para entender los fenómenos de ÉXITO y FRACASO, como el caso del experimento de tirar n monedas todas independientes como se vio en el apartado **Experimento teórico de tirar monedas** de la sección **Introduccion**, como también para la ley débil de los grandes números. Y es por este último caso que se predende detallar esta variable aleatoria indicadora para un evento particular $\beta \equiv \{|X| \geq a\}$. Es decir, queremos

los $\omega \in \Omega$ tales que $|X| \geq a$, ese es nuestro evento (osea subconjunto de Ω). Entonces tenemos;

$$I_{(|X| \geq a)} = \begin{cases} 1 & ; \text{si y solo si para los } \omega \in \Omega \text{ tales que } |X| \geq a \\ 0 & ; \text{si y solo si para los } \omega \in \Omega \text{ tales que } |X| < a \end{cases} \quad (96)$$

Tendríamos entonces que $P(|X| \geq a) + P(|X| < a) = 1$ que es equivalente a $P(I_\beta = 1) + P(I_\beta = 0) = 1$; y si lo relacionamos con el experimento de tirar monedas tendríamos $P(|X| \geq a) = r$ y $P(|X| < a) = q = 1 - r$.

Veamos ahora una de las propiedades mas importantes de esta Variable aleatoria.

$$E(I_\beta) = 0 \cdot (1 - P(I_\beta = 1)) + 1 \cdot (P(I_\beta = 1)) = 0 \cdot P(I_\beta = 0) + 1 \cdot P(I_\beta = 1) = P(I_\beta = 1). \quad (97)$$

Por definición de $E(\cdot)$, barre por todo el espacio muestral, pero en nuestro caso “dividimos” a nuestro espacio muestral de la siguiente manera $\Omega = (\text{que ocurra } \beta) \cup (\text{que no ocurra } \beta)$ y para el evento $\beta \equiv \{|X| \geq a\}$

$$E(I_\beta) = \sum_i I_{(|x_i| \geq a)} P(I_{(|x_i| \geq a)}) = 0 \cdot (1 - P(I_\beta = 1)) + 1 \cdot (P(I_\beta = 1)) = P(I_\beta = 1) \quad (98)$$

donde el índice i recorre todos los posibles valores de la variable aleatoria X , que obviamente estarán los que son mayorese iguales a a como los que no. Pero en este caso $P(I_\beta = 1) = P(|X| \geq a)$ ya que es nuestro evento Éxito, es decir, la probabilidad de que la variable aleatoria Indicadora I_β tome el valor uno es equivalente a la probabilidad de que se cumpla el evento $|X| \geq a$ porque lo definimos con un “si y solo si”. Por lo tanto tenemos que

$$E(I_{(|X| \geq a)}) = P(|X| \geq a); \quad (99)$$

Supongamos ahora que tenemos una función creciente $g : \mathfrak{R}_+ \rightarrow \mathfrak{R}_+$ de una variable aleatoria X no negativa, tal que $E(g(X)) < \infty_+$, y $\forall a \in \mathfrak{R}$. Con prosedimiento parecidos a los anteriores podemos ver que

$$E(g(a)I_{(X \geq a)}) = 0 \cdot g(a) \cdot (1 - P(g(a)(I_{(X \geq a)} = 0))) + 1 \cdot g(a) \cdot P(g(a)(I_{X \geq a} = 1)) \quad (100)$$

Pero como $g(a)$ es una constante dado un a fijo, tenemos que $P(g(a)I_{X \geq a}) = g(a)P(X \geq a)$. Entonces la ecuación (100) nos queda de la siguiente forma;

$$E(g(a)I_{X \geq a}) = g(a) \cdot P(I_{X \geq a} = 1) = g(a) \cdot P(X \geq a) \quad (101)$$

Esta igualdad será de utilidad para demostrar la desigualdad de Chevychev.

9. Apéndice B: Convergencias y Desigualdades

Consideremos una sucesión de variables aleatorias $\{X_n; n \in \mathbb{N}\}$ y una variable aleatoria X definidas sobre el mismo espacio muestral $(\Omega, \psi, \mathbb{P})$. Dado que X_n y X son funciones que van de Ω a \mathfrak{R} sale cuestionarse que tipo de convergencia se pueden dar de funciones a funciones, tales como la convergencia punto a punto, uniforme o cuadrática. Pero el hecho de que estas funciones ($\{X_n\}$ y X) sean aleatorias nos exige redefinir las condiciones de convergencia. Definiremos tres conceptos de convergencias que son muy importantes en la teoría de probabilidad, como son la “convergencia en probabilidad”, la “convergencia casi segura” y la “convergencia en media r^{th} ”. Como nuestro objetivo en la sección “Introducción” es desarrollar la Ley débil de los grandes números le daremos mas importancia a la convergencia en probabilidad.

Definición: Convergencia en Probabilidad Dadas la sucesión de variables aleatorias $\{X_n; n \in \mathbb{N}\}$ y la variable aleatoria X definidas sobre el mismo espacio muestral $(\Omega, \psi, \mathbb{P})$; decimos que la sucesión $\{X_n\}_{n \in \mathbb{N}}$ converge en probabilidad a la variable aleatoria X si ocurre que para todo $\varepsilon > 0$;

$$\lim_{n \rightarrow +\infty} P(|X_n - X| \geq \varepsilon) = 0$$

Informalmente podemos interpretar como que si tomamos un valor de $\varepsilon > 0$ pero fijo la probabilidad de que X_n este en el intervalo $(X - \varepsilon; X + \varepsilon)$ tiende a 1 cuando n tiende a infinito.

Notación: $X_n \xrightarrow{P} X$

Definición: Convergencia Casi Segura Dadas la sucesión de variables aleatorias $\{X_n; n \in \mathbb{N}\}$ y la variable aleatoria X definidas sobre el mismo espacio muestral $(\Omega, \psi, \mathbb{P})$; decimos que la sucesión $\{X_n\}_{n \in \mathbb{N}}$ converge de forma Casi Segura a X si para todo $\varepsilon > 0$ ocurre;

$$P\left(\lim_{n \rightarrow +\infty} X_n = X\right) = 1$$

Notación: $X_n \xrightarrow{c.s.} X$

Se puede demostrar, nosotros no lo haremos que $(X_n \xrightarrow{c.s.} X) \implies (X_n \xrightarrow{P} X)$.

Definición: Convergencia en media r^{th} ; $r > 0$ Dadas la sucesión de variables aleatorias $\{X_n; n \in \mathbb{N}\}$ y la variable aleatoria X definidas sobre el mismo espacio muestral $(\Omega, \psi, \mathbb{P})$; decimos que la sucesión $\{X_n\}_{n \in \mathbb{N}}$ converge en media r^{th} a X si ocurre;

$$E|X_n - X|^r \rightarrow 0; \text{ cuando } n \rightarrow \infty$$

Notación: $X_n \xrightarrow{r} X$

Desigualdad de Chebychev Sea X una variable aleatoria no negativa (tambien se puede decir que es para cualquier variable aleatoria pero usamos $|X|$ en ves de X) y sea

$g : \mathfrak{R}_+ \rightarrow \mathfrak{R}_+$ una función creciente tal que $E[g(X)] < \infty$. Entonces $\forall a \in \mathfrak{R}$ tenemos que

$$g(a)P(X \geq a) \leq E[g(X)]; \quad (102)$$

Demostración:

Por hipótesis nuestro evento son todos los valores de $X \geq a$; al ser X no negativa y $g(X)$ creciente entonces la desigualdad anterior no se ve afectada si le aplico $g(\cdot)$ en ambos lados; esto nos da

$$g(a) \cdot I_{X \geq a} \leq g(X) \quad (103)$$

donde

$$g(a) \cdot I_{(|X| \geq a)} = \begin{cases} g(a) & ; \text{si y solo si para los } \omega \in \Omega \text{ tales que } |X| \geq a \\ 0 & ; \text{si y solo si para los } \omega \in \Omega \text{ tales que } |X| < a \end{cases} \quad (104)$$

Entonces si aplicamos la Esperanza en ambos lados de la desigualdad (103) tenemos;

$$E(g(a) \cdot I_{X \geq a}) \leq E(g(X)) \quad (105)$$

Pero como vimos en la ecuación 101 ($E(g(a)I_{X \geq a}) = g(a) \cdot P(X \geq a)$). Entonces tenemos finalmente que

$$g(a) \cdot P(X \geq a) \leq E(g(X)) \quad (106)$$

Desigualdad de Markov Un caso particular de la desigualdad de Chebychev es la desigualdad de Markov. Sea X una variable aleatoria no negativa; para cada $a > 0$, tenemos

$$P(X \geq a) \leq \frac{E(X)}{a} \quad (107)$$

Demostración: Es cuestión de reemplazar $g(X) \rightarrow X$ lo que lleva a que $g(a) = a$

Pequeña desigualdad de Chebychev Sea X una variable aleatoria de varianza finita. Para cada $a > 0$ se cumple que ;

$$P(|X - E(X)| \geq a) \leq \frac{V(x)}{a^2} \quad (108)$$

Demostración: Como $(X - E(X))^2$ es una variable aleatoria podemos aplicar la desigualdad de markov (poniendo a^2 en vez de a)tenemos;

$$P[(X - E(X))^2 \geq a^2] \leq \frac{E[(X - E(X))^2]}{a^2} = \frac{V(X)}{a^2} \quad (109)$$

Pero como la desigualdad $(X - E(X))^2 \geq a^2$ es equivalente a $|X - E(X)| \geq a$ (vemos que pusimos el valor absoluto de $X - E(X)$ ya que tiene que ser no negativa) entonces el teorema queda demostrado.

10. Apendice C: Algunas cuestiones sobre Teoría de la Información

Intentaremos describir ciertos conceptos importantes que derivan de la teoría de la información. Como ejes del apéndice desarrollaremos el concepto de Entropía condicional, conjunta, y definiremos la Información Mutua. Iremos construyendo el conocimiento de una forma lenta y concreta para poder llegar al final sin huecos lógicos. Empezaremos definiendo la probabilidad conjunta; parecerá recurrente ya que la hemos utilizado en otros apartados pero será útil revisarla de una forma más detallada.

Definición: Probabilidad conjunta Se define a la probabilidad conjunta como

$$P(X, Y) = P(X = x \cap Y = y); \forall x, y \text{ existentes en el rango de las variables aleatorias.} \quad (110)$$

Esta igualdad nos dice que la probabilidad conjunta es una medida de probabilidad asignada al conjunto intersección de los dos conjuntos. Esto es, la probabilidad de que ocurra $X = x$ y $Y = y$ a la vez. Si esto no fuera posible, es decir, si de las muestras obtenidas de un experimento no se puede dar la situación de que la variable $X = x$ y que la variable $Y = y$ entonces $P(X, Y)$ no existe. Entendamos ahora el porqué del símbolo “intersección”. Cuando dos eventos tienen intersección no nula, $A \cap B \neq \emptyset$, estamos diciendo que de una medición el resultado pertenece a los dos grupos. Con el mismo pensamiento podemos decir que si medimos la posición de una partícula, los resultados pertenecerán a los grupos “ XeY ”. Entonces la probabilidad de que el evento $X = x$ se mida a la vez que el evento $Y = y$ es $P(X, Y) = P(X = x \cap Y = y)$. Esto es la probabilidad conjunta, veremos que la Entropía conjunta tiene (mas alla que es una entropía y no un probabilidad lo que está en juego) una relación distinta con los conjuntos intersección de los eventos.

Definición: Probabilidad Condicional Buscamos la probabilidad de que un cierto evento Y , ocurra si tenemos la certeza de que ocurrió el evento $X = x_i$, y la llamaremos $P(Y|X = x_i)$. Se demuestra fácilmente (con una visión frecuentística) que

$$P(Y|X = x_i) = \frac{P(X = x_i, Y)}{P(X = x_i)} \quad (111)$$

donde

$$P(X = x_i) = \sum_{y_j \in \mathcal{Y}} P(x_i, y_j) \quad (112)$$

es la probabilidad marginal, y $P(X = x_i, y_j)$ es la probabilidad conjunta. Lo que estamos haciendo al calcular la probabilidad marginal es asignarle una medida de probabilidad a eventos que no dependen de la variable y por eso es la suma en la ecuación 112. Es bueno

diferenciar la probabilidad condicional de la conjunta, aunque las dos están basadas sobre el conjunto muestral intersección su significado es muy distinto. Para poder entenderlo usaremos una técnica frecuentista, aunque este método no está cerca de ser formal. Sea $N = \text{rank}(\Omega)$, es decir, el número de elementos del espacio muestral total. Por teoría de conjuntos podemos deducir que $\text{rank}(\Omega) = \text{rank}(A) + \text{rank}(B) - \text{rank}(A \cap B)$. Esto nos dice que el valor N incluye muestras de los eventos intersección, es decir, que la medida de probabilidad asignada a ese subconjunto de Ω será la probabilidad conjunta. La probabilidad condicional es una medida de probabilidad asignada al conjunto intersección de los eventos pero donde el espacio que lo contiene ya no es Ω sino B o A depende como se plantee. Esto es porque

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\text{rank}(A \cap B)}{\text{rank}(\Omega)} \cdot \frac{\text{rank}(\Omega)}{\text{rank}(B)} \quad (113)$$

Que si lo llevamos a nuestro caso tenemos

$$P(Y|X = x_i) = \frac{\text{rank}(Y \cap X = x_i)}{\text{rank}(\Omega)} \cdot \frac{\text{rank}(\Omega)}{\text{rank}(X = x_i)} \quad (114)$$

Como vemos hemos reducido nuestro espacio muestral al subconjunto de $X = x_i$, y calculamos la probabilidad de que ocurra el evento Y pero ahora no sobre el espacio muestral Ω sino sobre el subconjunto $\{X = x_i\}$. Vemos que si los eventos son independientes entonces la probabilidad conjunta será el producto de las probabilidades marginales, por lo tanto la condicional quedará.

$$P(Y|X = x_i) = P(Y) \quad (115)$$

Busquemos entender ahora la Entropía condicional; $H(Y|X)$.

10.1. Entropía Condicional

Sea X e Y dos variables aleatorias, definidas sobre el mismo espacio probabilístico, entonces

$$H(Y|X) = \sum_{x_i \in X} P(X = x_i) H(Y|X = x_i) \quad (116)$$

Ahora nos ocuparemos de entender $H(Y|X =_i)$. Por definición tenemos que

$$\begin{aligned}
H(Y|x_i) &\equiv H(Y|X = x_i) \\
&\doteq \sum_{y_j \in \mathcal{Y}} P(y_j|x_i) \cdot \log(P(y_j|x_i)) \\
&= (-1) \sum_{y_j \in \mathcal{Y}} \frac{P(x_i, y_j)}{P(x_i)} \cdot \log\left(\frac{P(x_i, y_j)}{P(x_i)}\right) \\
&= (-1) \sum_{y_j \in \mathcal{Y}} \frac{P(x_i, y_j)}{P(x_i)} \cdot \left(\log(P(x_i, y_j)) - \log(P(x_i))\right) \\
&= (-1) \frac{1}{P(x_i)} \left(\sum_{y_j \in \mathcal{Y}} [P(x_i, y_j) \cdot \log(P(x_i, y_j))] - \sum_{y_j \in \mathcal{Y}} [P(x_i, y_j) \cdot \log(P(x_i))] \right) \\
&= \frac{-1}{P(x_i)} \left(-H(X = x_i, Y) - \log(P(x_i)) \cdot \sum_{y_j \in \mathcal{Y}} P(x_i, y_j) \right) \\
&= \frac{1}{P(x_i)} (H(X = x_i, Y) + \log(P(x_i)) \cdot P(x_i))
\end{aligned} \tag{117}$$

Teniendo esto podemos volver a la ecuación (116) reemplazándolo;

$$\begin{aligned}
H(Y|X) &= \sum_{x_i \in \mathcal{X}} P(x_i) \cdot H(Y|x_i) \\
&= \sum_{x_i \in \mathcal{X}} P(x_i) \left[\frac{1}{P(x_i)} \left(H(x_i, Y) + \log(P(x_i)) \cdot P(x_i) \right) \right] \\
&= \sum_{x_i \in \mathcal{X}} H(x_i, Y) + P(x_i) \cdot \log(P(x_i)) \\
&= H(X, Y) - H(X)
\end{aligned} \tag{118}$$

El primer sumando de la última igualdad de la ecuación (118) es la incertidumbre producida por las dos variables en conjunto, o sea, $\{XeY\}$ llamada Entropía Conjunta, y definida así

$$H(X, Y) \doteq \sum_{x_i \in \mathcal{X}; y_j \in \mathcal{Y}} P(x_i, y_j) \cdot \log(P(x_i, y_j)) \tag{119}$$

Tratemos de entender la ecuación (118); $H(Y|X)$. Empecemos por ver que, como lo dice explícitamente la ecuación, $H(Y|X)$ es la resta de la información total menos la información aportada por a variable X . Se puede interpretar entonces que es la información que aporta solo la variable Y . Es decir, si tomásemos una medición, la información

contenida en esa información pertenece solamente a la variable y . Esto nos dice que la entropía condicional actúa de forma muy distinta respecto de los conjuntos muestrales como lo hace la probabilidad condicional. Esta última toma al conjunto intersección de los eventos del espacio muestral (sabiendo que se reduce el espacio muestral al del otro evento) y les asigna una medida de probabilidad que relaciona esos dos subconjuntos. Pero la entropía no, ya que $H(Y|X)$ es algo (información) que solo le pertenece a la variable (o evento) Y y no al evento $(Y \cap X)$.

10.2. Desigualdad de Jensen “finita”

Esta desigualdad tiene gran importancia en muchos ámbitos de la ciencia nosotros intentaremos darle un enfoque estadístico así podremos luego relacionarla con la TI. Para empezar tendremos que definir en sentido estricto el significado de una función convexa.

Definición: Función Convexa Sea I un intervalo en \mathbb{R} ; y $f : I \rightarrow \mathbb{R}$. La función f se llama Convexa si para todos los $x, y \in I$ y para todos $a, b \geq 0$, con $a + b = 1$, se cumple la siguiente desigualdad,

$$f(ax + by) \leq af(x) + bf(y); \quad (120)$$

Uno puede darle a esta definición una interpretación geométrica, viendo que los puntos $(ax + by, af(x) + bf(y))$ están sobre la recta que une los puntos $(x, f(x))$ y $(y, f(y))$. Esto se ve ya que si tomamos $y > x$, el punto $ax + by > ax + bx = x(a + b) = x$, dado que por hipótesis $a + b = 1$.

se puede demostrar por inducción que si $f : I \rightarrow \mathbb{R}$ es convexa; y $x_1, \dots, x_n \in I$ con $\gamma_1, \dots, \gamma_n \geq 0$ y $\gamma_1 + \dots + \gamma_n = 1$, entonces

$$f\left(\sum_{k=1}^n \gamma_k x_k\right) \leq \sum_{k=1}^n \gamma_k f(x_k) \quad (121)$$

Con la visión puesta en la Sección (3. 2. 1) referida a la Divergencia de Kullback-Leibler, desarrollaremos algunos ejemplos referidos a esta desigualdad. Se puede ver que si f es convexa y X es una variable aleatoria entonces

$$f(E\{X\}) \leq E\{f(X)\} \quad (122)$$

11. Apendice D: Proyecciones entre Procesos Estocásticos

Con la mirada puesta en la sección 3,2 y 5,2 daremos algunas herramientas que luego serán muy útiles en esas secciones. En la sección 3,2 hablaremos sobre distancias entre distribuciones de probabilidad, estas nos servirán para poder distinguir una cadena simbólica de otra. Como hemos planteado en las secciones anteriores a la 3,2 el porqué usar las distribuciones de probabilidad y no el proceso estocástico subyacente a ellas también se explicará en esta sección. Es decir, diremos porque buscamos distancias entre distribuciones de probabilidades y no entre procesos estocásticos. Pero para poder entender este proceso lógico, primero tenemos que tener una base de conocimiento sobre las nociones de distancia entre Espacio de Hilbert y porqué no el Teorema de Proyección.

11.1. Nuestro espacio de Hilbert

A grandes rasgos un espacio de Hilbert es un espacio vectorial complejo (o no) dotado de algún producto interno y que además es completo en el sentido de norma. En los siguientes apartados nos ocuparemos en desarrollar formalmente esta idea como también en describir algunas conclusiones de las mismas (teorema de Proyección).

Espacios dotados con algún producto interno Sea Θ un espacio vectorial complejo (o no) es llamado espacio de producto interno si para cada par de elementos de Θ , μ y ν ; existe un número complejo (o no) $\langle \mu, \nu \rangle$; el cual llamaremos producto interno de μ y de ν tal que cumpla las siguientes propiedades

1. $\langle \mu, \nu \rangle = \overline{\langle \nu, \mu \rangle}$, conjugación compleja
2. $\langle \mu + \nu, \lambda \rangle = \langle \mu, \lambda \rangle + \langle \nu, \lambda \rangle$; para todo μ, ν y $\lambda \in \Theta$.
3. $\langle \alpha \mu, \nu \rangle = \alpha \langle \mu, \nu \rangle$, para todo μ y $\nu \in \Theta$ y $\alpha \in \mathbb{C}$.
4. $\langle \mu, \mu \rangle \geq 0$, para todo $\mu \in \Theta$
5. $\langle \mu, \mu \rangle = 0$ si y solo si $\mu = 0$

Vemos claramente que no definimos explícitamente el operador $\langle \cdot, \cdot \rangle$, pero si se describe su forma de relacionarse con los elementos de Θ . Esto traerá como consecuencia que se puedan formar distintos espacios con productos internos diferentes. Una definición que se desprende de la anterior es la de la norma

Definición : Norma Se define a la Norma de un vector como

$$\|\mu\| \doteq \sqrt{\langle \mu, \mu \rangle} \quad (123)$$

Una de las propiedades mas importantes de la norma es que cumple con la desigualdad triangular , es decir, $\|\mu + \nu\| \leq \|\mu\| + \|\nu\|$.

Sea $\{\nu_n; n = 1, 2, \dots\}$ una sucesión de elementos de Θ (espacio con producto interno). Se dirá que esta sucesión converge en norma a $\nu \in \Theta$, si $\|\nu_n - \nu\| \rightarrow 0$ cuando $n \rightarrow \infty$.

Definición: Sucesión de Cauchy Sea una sucesión $\{\nu_n; n = 1, 2, \dots\}$ de elementos de una espacio de producto interno Θ se dirá que es una sucesión de Cauchy si

$$\|\nu_n - \nu_m\| \rightarrow 0; \text{ cuando } m, n \rightarrow \infty \quad (124)$$

, también se lo puede interpretar como que para todo $\epsilon > 0$ existe un nmero $N(\epsilon)$ tal que

$$\|\nu_n - \nu_m\| < \epsilon; \text{ para todo } m, n > N(\epsilon) \quad (125)$$

Dado esto estamos en condiciones de definir un espacio de Hilbert.

Definición: Espacio de Hilbert Se dirá que Θ es un espacio de Hilbert si es un espacio de producto interno, y si toda sucesión de Cauchy $\{\nu_n; n = 1, 2, \dots\}$ converge a un elemento $\nu \in \Theta$.

Nuestro espacio de Hilbert Teniendo en nuestras manos todas estas definiciones podemos empezar que describir el espacio de Hilbert que es de nuestro interes. Consideremos el siguiente espacio Hilbert, al cual llamaremos $\mathcal{H} \equiv L^2(\Omega, \beta, \mathbb{P})$. Este espacio no es mas que la unión de un espacio probabilístico $(\Omega, \beta, \mathbb{P})$ con la colección \mathbb{G} de todas las variables aleatorias definidas en el espacio muestral Ω . Pero como lo definimos, nuestro espacio es L^2 , lo que implica que tiene que satisfacer

$$E(X^2) = \int_{\Omega} X(\omega)^2 P(d\omega) < \infty \quad (126)$$

Vemos claramente que la definición de nuestro producto interno es sobre el espacio muestral. Dejaremos para el lector interesado comprbar que precisamente el espacio vectorial \mathcal{H} es un espacio de Hilbert. Esto implica también comprobar que el producto interno intrinseco de nuestro espacio de Hilbert es presisamente el primer Momento $E(\cdot)$, es decir, ver que $\langle X, Y \rangle \equiv E(XY)$, donde X e Y son variables aleatorias definidas sobre el mismo espacio muestral. Un punto que también debe ser aclarado es que la igual anterior se cumple para todas las propiedades de producto interno excepto para la última, es decir, $\langle X, X \rangle = 0$ no implica que $X(\omega) = 0$ para todos los ω , si no solo nos dice que $P(X = 0) = 1$. Esto se lee como que X es equivalente en probabilidad a 0.

11.2. Teorema de Proyección

Ya debatido y descripto nuestro espacio de Hilbert $\mathcal{H} \equiv L^2(\Omega, \beta, \mathbb{P})$, podremos desarrollar el concepto de proyección de un subespacio en otro subespacio. Estas nociones nos ayudarán a entender luego a los procesos Deterministas y Aleatorios. Para interiorizarnos en el tema haremos un repaso de la “Aproximación Lineal” en \mathbb{R}^3 . Supongamos que tenemos tres vectores $\mathbf{y} = (1/4, 1/4, 1)$; $\mathbf{x}_1 = (1, 0, 1/4)$ y \mathbf{x}_2 escritos en la base canónica. Nuestro problema es encontrar la mejor “versión” de \mathbf{y} que se pueda construir como combinación lineal de \mathbf{x}_1 y \mathbf{x}_2 , es decir, buscamos el $\hat{\mathbf{y}} = \alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2$ tal que minimice el error cuadrático medio, que $S = \|\mathbf{y} - \hat{\mathbf{y}}\| = \|\mathbf{y} - \alpha_1 \mathbf{x}_1 - \alpha_2 \mathbf{x}_2\|$. Uno podría encarar este problema derivando respecto de α_1 y α_2 e igualando a cero tal que $S = (1/4 - \alpha_1)^2 + (1/4 - \alpha_2)^2 + (1 - \frac{1}{4}\alpha_1 - \frac{1}{4}\alpha_2)^2$ encuentre su mínimo. Una alternativa geométrica a este método es verificar para que α_1 y α_2 el vector $\mathbf{y} - \alpha_1 \mathbf{x}_1 - \alpha_2 \mathbf{x}_2$ sea ortogonal al subespacio generado por \mathbf{x}_1 y \mathbf{x}_2 , esto es

$$\langle \mathbf{y} - \alpha_1 \mathbf{x}_1 - \alpha_2 \mathbf{x}_2, \mathbf{x}_i \rangle = 0 ; i = 1, 2 \quad (127)$$

Uno puede resolver el sistema de ecuaciones y calcular los α_1 y α_2 que lo satisface.

El concepto de aproximación lineal en nuestro espacio de Hilbert Sea $\mathcal{H} \equiv L^2(\Omega, \beta, \mathbb{P})$ nuestro espacio de Hilbert mencionado anteriormente. Tenemos ahora tres variables aleatorias Y, X_1 y X_2 , que son obviamente elementos de $\mathcal{H} \equiv L^2(\Omega, \beta, \mathbb{P})$. Supongamos que solo X_1 y X_2 son observadas y queremos “estimar” la variable Y que no pudo ser observada. Lo que estamos buscando entonces es el $\hat{Y} = \alpha_1 X_1 + \alpha_2 X_2$ que minimice el error cuadrático medio, o sea

$$S = E|Y - \alpha_1 X_1 - \alpha_2 X_2|^2 = \|Y - \alpha_1 X_1 - \alpha_2 X_2\|^2 \quad (128)$$

encuentre su mínimo. Como aclaramos en los párrafos anteriores, nosotros elegiremos la versión geométrica. Nuestro objetivo es encontrar la variable aleatoria \hat{Y} que vive en el subespacio

$$\mathcal{M} = \{X \in \mathcal{H} : X = a_1 X_1 + a_2 X_2; a_1, a_2 \in \mathbb{R}\} \quad (129)$$

cuya distancia cuadrática a Y , $\|Y - \hat{Y}\|^2$ sea lo más chica posible. Como hicimos en el apartado anterior buscamos el \hat{Y} que viva en \mathcal{M} tal que $Y - \hat{Y}$ sea ortogonal a \mathcal{M} , o sea

$$\langle Y - a_1 X_1 - a_2 X_2, X \rangle = 0; \text{ para todo } X \in \mathcal{M} \quad (130)$$

o equivalentemente

$$\langle Y - a_1 X_1 - a_2 X_2, X_i \rangle = 0; i = 1, 2 \quad (131)$$

Teniendo esto podemos armar un sistema de ecuaciones para encontrar los valores de a_1 y a_2

$$a_1 E(X_1^2) + a_2 E(X_2 X_1) = E(Y X_1) \quad (132)$$

$$a_1 E(X_1 X_2) + a_2 E(X_2^2) = E(Y X_2) \quad (133)$$

Teorema de Proyección Si \mathcal{M} es un subespacio cerrado de nuestro espacio de Hilbert $\mathcal{H} \equiv L^2(\Omega, \beta, \mathbb{P})$ y $X \in \mathcal{H}$ luego

1. existe un unico elemento $\hat{X} \in \mathcal{M}$ tal que;

$$\|X - \hat{X}\| = \inf_{Y \in \mathcal{M}} \|X - Y\| \quad (134)$$

2. si $\hat{X} \in \mathcal{M}$ y $\|X - \hat{X}\| = \inf_{Y \in \mathcal{M}} \|X - Y\|$ si y solo si $\hat{X} \in \mathcal{M}$ y $(X - \hat{X}) \in \mathcal{M}^\perp$

Luego al elemento \hat{X} es llamado la proyección de X en \mathcal{M} . Dado esto podemos seguir leyendo los la sección 3

12. Apéndice E: Representación Espectral

En este apéndice se intentará resumir varios aspectos de la representación espectral de los procesos estocásticos. Se prefirió reservarlo para un apéndice para no desviar la concentración del lector en el capítulo referido al Ruido Rosa, y que cada uno elija informarse más sobre el tema en cuestión.

Se dice que un proceso estocástico $\{X_t\}$ es un proceso estacionario a valores complejos si $E|X_t|^2 < \infty$, $E(X_t)$ es independiente de t y $E(X_{t+h}, \overline{X_t})$.

Vale remarcar que las variables aleatorias complejas X sobre el espacio probabilístico (Ω, \mathcal{B}, P) , que satisfacen $E|X|^2 < \infty$, forman un espacio de Hilbert complejo con producto interno:

$$\langle X, Y \rangle = E(X, \overline{Y}) \quad (135)$$

Y esto nos permite redefinir la autocovarianza para procesos estacionarios complejos $\{X_t\}$ escrita en la ecuación (6.),

$$\gamma(h) = E(X_{t+h}, \overline{X_t}) - E(X_{t+h})E(\overline{X_t}) \quad (136)$$

Se puede notar que si restringimos al proceso $\{X_t\}$ a valores reales entonces las definiciones de producto interno y consecuentemente la autocovarianza serán de valores reales como las definimos en capítulos anteriores. Vale aclarar ciertas propiedades de la función autocovarianza

1. $\gamma(0) \geq 0$
2. $|\gamma(h)| \leq \gamma(0)$ para todo entero h
3. $\gamma(\cdot)$ es hermitiana, es decir $\gamma(h) = \overline{\gamma(-h)}$

Hay veces que las señales que trabajamos presentan diferencias que solo se pueden ver si las escribimos de una manera distinta. Supongamos que tenemos un proceso estocástico y lo escribimos de la siguiente manera

$$X_t = \sum_{j=1}^n A(\nu_j) e^{it\nu_j} \quad (137)$$

donde $-\pi < \nu_1 < \nu_2 < \dots < \nu_n = \pi$ y $A(\nu_1), \dots, A(\nu_n)$ son coeficientes no-correlacionados ($E(A(\nu_j)A(\nu_i)) = 0$) de valor complejo, tal que

$$E(A(\nu_j)) = 0, j = 1, \dots, n \quad (138)$$

y

$$E(A_{\nu_j}, \overline{A(\nu_j)}) = \sigma_j^2 \quad (139)$$

Se puede ver que la ecuación (137) es la transformada de Fourier discreta. De las ecuaciones (138. y 139) podemos ver que para un proceso estacionario se tiene que

$$E(X_t) = 0 \quad (140)$$

y

$$E(X_{t+h}X_t) = \sum_{j=1}^n \sigma_j^2 e^{ih\nu_j} \quad (141)$$

Lo que se puede ver nuevo que el proceso es estacionario ya que el momento y la autocovarianza no dependen de t. Si escribimos la anterior expresión como una integral de Riemann podemos ver que el proceso estacionario $\{X_t\}$ como autocovarianza a la función

$$\gamma(h) = \int_{(-\pi, \pi]} e^{ih\nu} dF(\nu) = \int_{(-\pi, \pi]} e^{ih\nu} f(\nu) d\nu \quad (142)$$

La función F se la conoce como distribución espectral del proceso $\{X_t\}$ y le asigna un peso a cada frecuencia del intervalo $(-\pi, \pi]$ y $f(\nu)$ se la conoce como Densidad Espectral. Podemos ver que si hacemos la transformada de la autocovarianza tenemos

$$f(\nu) = \int_{-\infty}^{\infty} \gamma(\tau) e^{i\tau\nu} d\tau \quad (143)$$

que es la Densidad Espectral. Como se ve la densidad espectral depende de la frecuencia y la autocovarianza del tiempo. Las utilidades de esta perspectiva pueden verse en la caracterización de ruidos como también de otras cadenas deterministas.

Si recordamos la definición de Ruido Blanco :

$$E(X_t) = 0 \quad (144)$$

y que cumple

$$\gamma(h) = \sigma^2 \delta(h - 0) \quad (145)$$

Teniendo esto podemos ver que la densidad Espectral del Ruido Blanco no es mas que una constante, ya que

$$f(\nu) = \int \gamma(h) e^{i\omega h} dh = \int \sigma^2 \delta(h - 0) e^{i\omega h} = \sigma^2 \quad (146)$$

Como se ve la densidad espectral no depende de la frecuencia.

13. Apéndice F: Código fuente del algoritmo de la “Ventana Móvil”

```

#include <stdio.h> // libreria basica de entrada y salida
#include <stdlib.h> // libreria para el uso del rand()
#include <time.h> // libreria para el uso del time()
#include <math.h>

void main ()
{
//declaraciones de variables
const int N=25000, c=8, pal=256/*256+1*/ /*2{C}*/ ,n=10000/*media ventana*/
;
//size_t fwrite(void *cadena_final, size_t(int), size_t(N-1),FILE *
cadena_final) // falta declaracion
int i,k,j,l,h,t,q,d,m,e,p,z;
//int printf () ,scanf("%d",&N) //falta
declaracion
int B[2*N-1 /*2*N-2-1*/];
int cadena_final[(2*N-1)-(c-1)/*largo-(c-1)*/];
int u[256/*(pow(2,c)-1),,y el cero*/];
int g[256/*(pow(2,c)-1),,y el cero*/];
int cantidad[256/*(pow(2,c)-1)*/];
const double pi=(3.14159265), peso_I=1/2., peso_D=1/2.;
double a,b,hora=time(0);
/* ver largo de la cadena final(son dos lugares menos que la original
larga)*/
double *prob_I=calloc((2*N)*(pal),sizeof(double));/*((2*N-1)-(c-1)-2*n)*
(pal+1)*/
double *prob_D=calloc((2*N)*(pal),sizeof(double));
double Total[2*N];
//double prob_I[(2*N-1)-n*2-2/*dimension de Cadena_final menos lo lugares
inabilitados del cursor por la ventana que topa*/][256+1/*(pow(2,c)-1)
,,y el cero,osea"d"*/];
//double prob_D[(2*N-1)-n*2-2/*igual que prob_izq-1*/][256+1/*(pow(2,c)-1)
,,y el cero,osea"d"*/];
double JSD[2*N/*igual que prob_I-1*/];
double suma_I[2*N/*igual que prob_I-1*/];
double suma_D[2*N/*igual que prob_I-1*/];
double vector_2 [N];
double vector_1 [N];
FILE* archivo_divergencia;
FILE*cadena_alfabetizada;
FILE*archivo_prob_I;
FILE*archivo_prob_D;
FILE*vector_total;
FILE*vector_total_binario;

```

```

FILE*suma_I_D;
FILE*archivo_1;
FILE*archivo_2;
FILE*jensen;

const char w,r;

srand48( time(NULL));

////////////////////////////////////

archivo_1=fopen("secuencia_1.txt","r");

for(i=0;i<N;i++){
    fscanf(archivo_1,"%f",&vector_1[i]);
}

fclose(archivo_1);
//////////////////////////////////// armado del vector 2
archivo_2=fopen("secuencia_2.txt","r");

for(i=0;i<N;i++){
    fscanf(archivo_2,"%f",&vector_2[i]);
    printf("%f\n",vector_2[i]);
}

fclose(archivo_2);
//////////////////////////////////// armado del vector total
vector_total=fopen("vector_total_N.dat","w");
printf("\nvector_total:\n");
for(t=0;t<N/*N-1*/;t++){
    Total[t]=vector_1[t];
}
for(l=(N);l<(2*N)/*2*N-2*/;l++){
    Total[l]=vector_2[l-N+1]; // el (l-N+1) es porque quiero que
    Total empiece en (N-1) y caotic en 0=(N-1)-N+1
}
for(q=0;q<(2*N)/*2*N-2*/;q++){
printf("%f\n",Total[q]);
    fprintf(vector_total,"%d\ f_%f\n",q,Total[q]);
}
fclose(vector_total);

```

```

//////////////////////////////////// conversion a binario de la
cadena total
vector_total_binario=fopen("vector_total_binario_N.dat","w");
printf("\nvector_binario_total:\n");
for(h=0; h<(2*N)-1/*2N-2-1*/;h++){ //creo el vector en binario
    if(Total[h]<Total[h+1]){
        B[h]=0;
    }else{
        B[h]=1;
    }
    printf("%d",B[h]);
    fprintf(vector_total_binario,"%d\\f_\\d\\n",h,B[h]);
}
fclose(vector_total_binario);

////////////////////////////////////

//creacion de ventanas que se mueven de a un casillero
cadena_alfabetizada=fopen("cadena_alfabetizada_N.dat","w");
printf("\ncadena_alfabetizada:\n");
for(e=0;e<(2*N-1)-(c-1);e++){ /*la cantidad de "ventanitas" es el largo
de la cadena "vector_toral_binario" menos la ventanita menos uno ya
que se mueve de a un casillero*/

    cadena_final[e]=0;
    for(i=0;i<c;i++){

        cadena_final[e]=B[e+i]*pow(2,i)+cadena_final[e];

    }

    printf("%d", cadena_final[e]);
    fprintf(cadena_alfabetizada,"%d\\f_\\d\\n",e,cadena_final[e]);
}
fclose(cadena_alfabetizada);

//////////////////////////////////// cracion de la
ventana
archivo_prob_I=fopen("prob_I_ventana_palabras_N.dat","w");
archivo_prob_D=fopen("prob_D_ventana_palabras_N.dat","w");
suma_I_D=fopen("suma_I_D_ventan_palabras_N.dat","w");

for(p=n;p< 2*N-c-n/*(2*N-1-(c-1))-(n)*/;p++){ //esta bien
    queempiece en 250 y no 249 ya que se cuenta el cero y el puntero
    no lo tenemos en cuenta al contar la probabilidad
    suma_I[p]=0;
    suma_D[p]=0;

    for(d=0;d<pal/*(pow(2,c+1) y menor igual*/;d++){

```

```

        u[d]=0;
        g[d]=0;
        for (i=p-n; i<p; i++){
            if (cadena_final[i]==d){
                u[d]=u[d]+1;
            }
        }
        for (j=p+1; j<p+(n+1); j++){
            if (cadena_final[j]==d){
                g[d]=g[d]+1;
            }
        }

        prob_I[p*pal + d]=u[d]*1./n*1.;
        prob_D[p*pal + d]= g[d]*1./n*1.;

        suma_I[p]=prob_I[p*pal + d]+suma_I[p];
        suma_D[p]=prob_D[p*pal + d]+suma_D[p];
        fprintf(archivo_prob_I, "%d/%d\n", p, d, prob_I[p*pal + d]);
        fprintf(archivo_prob_D, "%d/%d\n", p, d, prob_D[p*pal + d]);

    }
    fprintf(archivo_prob_I, "\n");
    fprintf(archivo_prob_D, "\n");
    fprintf(suma_I_D, "%d\n", p, suma_I[p], suma_D[p]);

}

fclose(suma_I_D);
fclose(archivo_prob_I);
fclose(archivo_prob_D);
////////////////////////////////////jensen

    archivo_divergencia=fopen("JSD_ventana_palabras.dat", "w");

    for (p=n; p<(2*N-c-n); p++){

        JSD[p]=0;

        k=0;
        for (d=0; d<pal; d++){
            if (prob_I[p*pal+d]!=0 && prob_D[p*pal+d]!=0){
                k=k+1;
            }
        }

        for (d=0; d<pal; d++){
            if (prob_I[p*pal+d]!=0 && prob_D[p*pal+d]!=0){

```

```

JSD[p]=JSD[p] -(peso_I*prob_I[p*pal+d]+peso_D*prob_D[p*pal+
d])*log2(peso_I*prob_I[p*pal+d]+peso_D*prob_D[p*pal+d])
+ peso_I*(prob_I[p*pal+d]*log2(prob_I[p*pal+d])) +
peso_D*(prob_D[p*pal+d]*log2(prob_D[p*pal+d]));
}
if(prob_I[p*pal+d]==0 && prob_D[p*pal+d]!=0){
JSD[p]=JSD[p] -(peso_D*prob_D[p*pal+d])*log2(peso_D*prob_D
[p*pal+d]) + peso_D*(prob_D[p*pal+d]*log2(prob_D[p*pal
+d]));
}
if(prob_D[p*pal+d]==0 && prob_I[p*pal+d]!=0){
JSD[p]=JSD[p] -(peso_I*prob_I[p*pal+d])*log2(peso_I*prob_I[
p*pal+d]) + peso_I*(prob_I[p*pal+d]*log2(prob_I[p*pal+d
]));
}
if(prob_I[p*pal+d]==0 && prob_D[p*pal+d]==0){
JSD[p]=JSD[p];
}
}

fprintf(archivo_divergencia, "%d_\\f_-%f_\\n", p, JSD[p]);

}

fclose(archivo_divergencia);

free(prob_I);
free(prob_D);
}

```

14. Apéndice G: Código fuente del algoritmo de D_{YHY}

```

#include <stdio.h> // libreria basica de entrada y salida
#include <stdlib.h> // libreria para el uso del rand()
#include <time.h> // libreria para el uso del time()
#include <math.h>

#define cols 5
#define filas 32
void main ()
{
//declaraciones de variables
const int N=5000, c=5;
//size_t fwrite(void *cadena_final, size_t(int), size_t(N-1), FILE *
    cadena_final) // falta declaracion
int i,k,j,l,h,t,q,d,m,e,p,z,potencia,B_1[N /*2*N-2*/],B_2[N],cadena_final
    [(2*N-2)-(c-1)*largo-(c+1)*l],cantidad[256/*(pow(2,c)-1)*/],u_1[32/*(
    pow(2,c)-1),,y el cero*/],u_2[32],g[256/*(pow(2,c)-1),,y el cero*/],
    datos_arreglo[filas][cols],palabras_1[N-4][5],palabras_2[N-4][5],
    suma_pal_1,suma_pal_2,matriz_u_1[32][2],matriz_u_2[32][2],ordenado_1
    [32],ordenado_sin_el_ultimo_1[31],ordenado_2[32],
    ordenado_sin_el_ultimo_2[31],u_1_con_uno_mas[33],u_2_con_uno_mas[33],
    ordenado_u_1[32],ordenado_u_2[32],R_1[32][2],R_1_pal[32][2],
    matriz_u_1_rank[32][2],R_2[32][2],matriz_u_2_rank[32][2];
//int printf(),scanf("%d",&N) //falta
    declaracion

const double pi=(3.14159265), peso_I=1/2.,peso_D=1/2.;
double a,b,hora=time(0), puntos_datos[N/*N*/],Y[N-1/*N-1*/], Total[2*N-2/*
    dimension de Random+Y*/],prob_I[(2*N-2)-500/*dimension de Total menos
    dos lugares inhabilitados del cursor*/][32/*(pow(2,c)-1),,y el cero,osea
    "d"*/],prob_D[(2*N-2)-500/*igual que prob_izq-1*/][32/*(pow(2,c)-1),,y
    el cero,osea"d"*/],JSD[(2*N-2)-500/*igual que prob_I-1*/],suma_I[(2*N-2)
    -500/*igual que prob_I-1*/],suma_D[(2*N-2)-500/*igual que prob_I-1*/],
    vector_caotico[N-1],vector_2[5000],vector_1[5000],prob_pal_1[32],
    prob_pal_2[32],suma_prob_1,suma_prob_2,JSD_sola,suma_con_abs,
    suma_sin_abs,D;
//double *puntos_datos,*ordenado,*Y,*V,*cadena_final; /* ver largo de la
    cadena final(son dos lugares menos que la original larga)*/
FILE* archivo_divergencia,*archivo_pal_1,*archivo_pal_2,*
    archivo_prob_pal_1,*archivo_prob_pal_2,*vector_total,*vector_1_binario
    ,*vector_2_binario,*suma_I,D,*archivo_1,*archivo_2,*conteo_pal_1,*
    conteo_pal_2,*archivo_datos,*arreglo_D,*archivo_msj_igualdad_1,*
    archivo_msj_igualdad_2,*archivo_rank_1,*archivo_rank_2,*
    archivo_distancia,*archivo_suma_prob_1,*archivo_suma_prob_2;
//FILE* probabilidades-I-1-0-D-1-0;
const char w,r;

```

```

srand48( time(NULL) );

////////////////////////////////////

archivo_1=fopen("logistica.txt","r");
    // importante el "%lf" en vez de "%f"
for(i=0;i<N;i++){
    fscanf(archivo_1,"%lf",&vector_1[i]); // acordarse del puntero &
    !!!!!
    // printf("%f \n",vector_1[i]);

    }

fclose(archivo_1);
//////////////////////////////////// armado del vector
caotico
archivo_2=fopen("pink_noise.txt","r");
    // importante el "%lf" en vez de "%f"
for(i=0;i<N;i++){
    fscanf(archivo_2,"%lf",&vector_2[i]);
    // printf("%f \n",vector_2[i]);

    }

fclose(archivo_2);

//////////////////////////////////// conversion a binario de la
cadena 1
vector_1_binario=fopen("vector_uno_binario.dat","w");
printf("\nvector_1_binario::\n");
for(h=0; h<(N)/2;h++){ //creo el vector en binario
    if(vector_1[h]<vector_1[h+1]){
        B_1[h]=0;
    }else{
        B_1[h]=1;
    }
    printf("%d",B_1[h]);
    fprintf(vector_1_binario,"%d\ f_%d\n",h,B_1[h]);
}
fclose(vector_1_binario);
////////////////////////////////////conversion de la cadena dos

vector_2_binario=fopen("vector_dos_binario.dat","w");
printf("\nvector_2_binario::\n");
for(h=0; h<(N)/2;h++){ //creo el vector en binario
    if(vector_2[h]<vector_2[h+1]){

```

```

        B_2[h]=0;
    }else{
        B_2[h]=1;
    }
    printf(" %d",B_2[h]);
    fprintf(vector_2_binario , " %d\ f_ %d\n",h,B_2[h]);
}
fclose(vector_2_binario);

////////////////////////////////////// creacion de
    todas las posibles cadenas binarias de largo m
    archivo_datos=fopen("cadenas_binarias_5.dat","r");
    arreglo_D=fopen("arreglo_datos_cadenas","w");

        // importante el "%lf" en vez de "%f"
    for(i=0;i<filas;i++){

        for(j=0;j<cols;j++){
            fscanf(archivo_datos , " %d",&datos_arreglo[i][j]);
            printf(" %d",datos_arreglo[i][j]);
            fprintf(arreglo_D , " %d",datos_arreglo[i][j]);
        }
        printf("\n");
        fprintf(arreglo_D , "\n");
    }
    fclose(archivo_datos);
    fclose(arreglo_D);

////////////////////////////////////// comparacion para ver que son todas distintas

    for(i=0;i<filas;i++){

        for(h=0;h<filas;h++){

            if (i!=h){
                k=0;
                for(j=0;j<cols;j++){

                    if(datos_arreglo[i][j]==datos_arreglo[h][j]){
                        k=k+1;
                    }
                }
            }
        }
    }

```

```

        }
        if(k==5){

            }
        }
    }

}

////////////////////////////////////// creacion del arreglo con las palabras
vector binario uno

archivo_pal_1=fopen("palabras_uno.dat","w");
printf("\npaleabras del vector uno:\n");

    for(h=0;h<(N-4);h++){

        for(e=h;e<(h+5);e++){
            palabras_1[h][e-h]= B_1[e];

            printf(" %d", palabras_1[h][e-h]);
            fprintf(archivo_pal_1,"%d\f", palabras_1[h][e-h]);
        }
        printf("\n");
        fprintf(archivo_pal_1,"\n");

    }
fclose(archivo_pal_1);

////////////////////////////////////// creacion del arreglo con
las palabras vector binario dos

archivo_pal_2=fopen("palabras_dos.dat","w");
printf("\npaleabras del vector dos:\n");

    for(h=0;h<(N-4);h++){

        for(e=h;e<(h+5);e++){
            palabras_2[h][e-h]= B_2[e];

```



```

conteo_pal_2=fopen("conteo_palabras_2.dat","w");
printf("\nconteo de palabras_2\n");
suma_pal_2=0;

for(i=0;i<32;i++){
    u_2[i]=0;
    for(h=0;h<N-4;h++){
        k=0;
        for(e=0;e<5;e++){

            if(palabras_2[h][e]==datos_arreglo[i][e]){

                k=k+1;

            }

        }
        if(k==5){

            u_2[i]=u_2[i] +1;

        }
        if(k!=5){
            u_2[i]=u_2[i];
        }

    }
    suma_pal_2=suma_pal_2 + u_2[i];
    printf(" ::: %d ::: %d\n", i, u_2[i]);
    fprintf(conteo_pal_2, "%d \f %d\n", i, u_2[i]);

}

printf("suma_pal_dos: %d\n", suma_pal_2);

fclose(conteo_pal_2);

//////////////////////////////////// calculo de
probabilidad de palabras de cadena uno
archivo_suma_prob_1=fopen("suma_prob_1.dat","w");
archivo_prob_pal_1=fopen("prob_pal_uno.dat","w");
printf("\nprobabilidad de palabras de cadena uno:\n");
suma_prob_1=0.;
for(i=0;i<32;i++){
    prob_pal_1[i]=u_1[i]*1./(N-4)*1.;
    suma_prob_1=suma_prob_1 + prob_pal_1[i];
    printf("%d ::: %f\n", i, prob_pal_1[i]);
}

```

```

        fprintf(archivo_prob_pal_1, "%d\ f_%f\n", i, prob_pal_1[i]) ;
    }

    printf("suma_de_prob:_%f\n", suma_prob_1);
    fprintf(archivo_suma_prob_1, "suma_probabilidad_1::%f", suma_prob_1);

fclose(archivo_suma_prob_1);

//////////////////////////////////// calculo de probabilidades
de oalabras pra la cadena dos

archivo_suma_prob_2=fopen("suma_prob_2.dat", "w");
archivo_prob_pal_2=fopen("prob_pal_dos.dat", "w");
printf("\n_probabilidad_de_palabras_de_cadena_dos:\n");
suma_prob_2=0.;
    for(i=0; i<32; i++){
        prob_pal_2[i]=u_2[i]*1./(N-4)*1.;
        suma_prob_2=suma_prob_2 + prob_pal_2[i];
        printf(" %d::_%f\n", i, prob_pal_2[i]);
        fprintf(archivo_prob_pal_2, "%d\ f_%f\n", i, prob_pal_2[i]) ;
    }

    printf("suma_de_prob:_%f\n", suma_prob_2);
    fprintf(archivo_suma_prob_2, "suma_probabilidad_2::%f", suma_prob_2)
;
fclose(archivo_suma_prob_2);

//////////////////////////////////// Ranquenoado cadena uno
printf("\n_matriz_u_uno:\n");
    for(i=0; i<32; i++){
        matriz_u_1[i][0]=i;
        matriz_u_1[i][1]=u_1[i];
        printf(" %d::%d\n", matriz_u_1[i][0], matriz_u_1
[i][1]);
    }

////////////////////////////////////
printf("\nu_1_con_uno_mas:\n");
u_1_con_uno_mas[32]=4996;
    for(i=0; i<32; i++){
        u_1_con_uno_mas[i]=u_1[i];
    }
    for(i=0; i<33; i++){
printf("_%d::%d\n", i, u_1_con_uno_mas[i]);
    }

```

```

////////////////////////////////////
do{                                     //ordeno los puntos
    for(i=1; i<32; i++){                //intercambia dos elemntos del
        vector que estan en mal orden
            b=u_1_con_uno_mas[i];
            a=u_1_con_uno_mas[i-1];
            if (u_1_con_uno_mas[i]<u_1_con_uno_mas[i-1]){
                u_1_con_uno_mas[i]=a;
                u_1_con_uno_mas[i-1]=b;
            }else{
                u_1_con_uno_mas[i-1]=a;
                u_1_con_uno_mas[i]=b;
            }
        }
    k=0;
    for(j=0;j<32;j++){                  /*verifico si el vector
        intercambiado ya estaordenado,,,sino lo esta k sera menor a N+1
        */
        if(u_1_con_uno_mas[j]<=u_1_con_uno_mas[j+1]){
            k=k+1 ;
            ordenado_1[j]=u_1_con_uno_mas[j];
        }
    }
}while (k<32);

printf("\nordenado:\n");
for(t=0;t<33;t++){
    printf("%d\n",ordenado_1[t]);
}
printf("\nordenado_sin_el_ultimo:\n");
for(i=0;i<33-1/**/;i++){
    ordenado_sin_el_ultimo_1[i]=ordenado_1[i];
    printf("%d\n",ordenado_sin_el_ultimo_1[i]);
}

////////////////////////////////////ahora lo hago de mayor a
menor

    printf("\nordenado_me_mayor_a_menor\n");
    for(i=0;i<32;i++){
        ordenado_u_1[i]=ordenado_sin_el_ultimo_1[31-i];
        printf("%d::%d\n",i,ordenado_u_1[i]);
    }

//////////////////////////////////// ranqueo

```

```

R_1[0][0]=1;
  i=1;
  k=1;
  do{

      if(ordenado_u_1[i]<ordenado_u_1[i-1]){
          R_1[i][0]=k+1;
          k=k+1;
      }
      if(ordenado_u_1[i]==ordenado_u_1[i-1]){

          R_1[i][0]=k;
          k=k;
      }

      i=i+1;

  }while(i<32);
for(t=0;t<32;t++){
  R_1[t][1]=ordenado_u_1[t];//R_1[][] tiene en la columna "0" el
  ranking y en la columna"1" la cantidad de veces que aparece
  la palabra en esa posicion del ranking
  }

  for(j=0;j<32;j++){

  }

////////////////////////////////////

archivo_rank_1=fopen("pal_rank_cantdeveces_1.dat","w");
printf("\nranking_con_palabras_uno\n");
  for(i=0;i<32;i++){
      matriz_u_1_rank[i][0]=matriz_u_1[i][0]; ////la matriz "
      matriz_u_1_rank[][]" tiene en la primera columna la
      palabra (identificada con un numero del 0 al 31) y en
      la segunda columna el ranking de la palabra
  }

  for(i=0;i<32;i++){
      for(h=0;h<32;h++){
          if(matriz_u_1[i][1]==R_1[h][1]){
              matriz_u_1_rank[i][1]=R_1[h][0];
          }
      }
  }
}

```

```

    for (i=0;i<32;i++){
    printf(" %d::%d::%d\n" ,matriz_u_1_rank[i][0] ,matriz_u_1_rank[i][1] ,
        matriz_u_1[i][1]);
    fprintf(archivo_rank_1," %d_\f_%d_\f_%d_\n" ,matriz_u_1_rank[i][0] ,
        matriz_u_1_rank[i][1] ,matriz_u_1[i][1]);
        }

fclose(archivo_rank_1);
//
////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////
//////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////// Ranquenedo cadena uno
printf("\n\nmatriz_u_dos:\n");
        for (i=0;i<32;i++){
            matriz_u_2[i][0]=i;
            matriz_u_2[i][1]=u_2[i];
            printf(" %d::%d\n" ,matriz_u_2[i][0] ,matriz_u_2
                [i][1]);
        }

////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////
printf("\nu_1_con_uno_mas:\n");
    u_2_con_uno_mas[32]=4996;
        for (i=0;i<32;i++){
            u_2_con_uno_mas[i]=u_2[i];
        }
        for (i=0;i<33;i++){
printf(" %d::%d\n" ,i ,u_2_con_uno_mas[i]);
        }

////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////

do{
                                                    //ordeno los puntos

    for(i=1; i<32; i++){
                                                    //intercambia dos elemntos del
        vector que estan en mal orden
            b=u_2_con_uno_mas[i];
            a=u_2_con_uno_mas[i-1];
            if (u_2_con_uno_mas[i]<u_2_con_uno_mas[i-1]){
                u_2_con_uno_mas[i]=a;
                u_2_con_uno_mas[i-1]=b;
            }else{

```

```

        u_2_con_uno_mas [ i-1]=a;
        u_2_con_uno_mas [ i]=b;
    }
}
k=0;
for (j=0;j <32;j++){
    /*verifico si el vector
    intercambiado ya estaordenado,,, sino lo esta k sera menor a N+1
    */
    if (u_2_con_uno_mas [ j]<=u_2_con_uno_mas [ j+1]){
        k=k+1 ;
        ordenado_2 [ j]=u_2_con_uno_mas [ j];
    }
}
}while (k<32);

printf("\nordenado_dos:\n");
for (t=0;t<33;t++){
    printf("%d\n",ordenado_2 [ t] );
}
printf("\nordenado_dos_sin_el_ultimo:\n");
for (i=0;i<33-1/**/;i++){
    ordenado_sin_el_ultimo_2 [ i]=ordenado_2 [ i];
    printf("%d\n",ordenado_sin_el_ultimo_2 [ i] );
}

////////////////////////////////////////ahora lo hago de mayor a
menor

    printf("\nordenado_mayor_a_menor\n");
    for (i=0;i <32;i++){
        ordenado_u_2 [ i]=ordenado_sin_el_ultimo_2 [31-i];
        printf("%d::%d\n",i,ordenado_u_2 [ i] );
    }

//////////////////////////////////////// ranqueo
//printf("\n ranking dos \n");
R_2 [0][0]=1;
i=1;
k=1;
do{

    if (ordenado_u_2 [ i]<ordenado_u_2 [ i-1]){
        R_2 [ i][0]=k+1;
        k=k+1;
    }
    if (ordenado_u_2 [ i]==ordenado_u_2 [ i-1]){

        R_2 [ i][0]=k;
        k=k;
    }
}

```

```

}
i=i+1;
}while(i<32);
for(t=0;t<32;t++){
R_2[t][1]=ordenado_u_2[t];//R_1[ ][ ] tiene en la columna "0" el
ranking y en la columna"1" la cantidad de veces que aparece
la palabra en esa posicion del ranking
}

for(j=0;j<32;j++){
// printf("%d:: %d:: %d \n",j,R_2[j][0],R_2[j][1]);
}

////////////////////////////////////

archivo_rank_2=fopen("pal_rank_cantdeveces_2.dat","w");
printf("\nranking_con_palabras_2\n");
for(i=0;i<32;i++){
matriz_u_2_rank[i][0]=matriz_u_2[i][0]; ////la matriz "
matriz_u_1_rank[ ][ ]" tiene en la primera columna la
palabra (identificada con un numero del 0 al 31) y en
la segunda columna el ranking de la palabra
}

for(i=0;i<32;i++){
for(h=0;h<32;h++){
if(matriz_u_2[i][1]==R_2[h][1]){
matriz_u_2_rank[i][1]=R_2[h][0];
}
}
}

for(i=0;i<32;i++){
printf("%d::%d::%d\n",matriz_u_2_rank[i][0],matriz_u_2_rank[i][1],
matriz_u_2[i][1]);
fprintf(archivo_rank_2,"%d\ f_ %d\ f_ %d\n",matriz_u_2_rank[i][0],
matriz_u_2_rank[i][1],matriz_u_2[i][1]);
}
fclose(archivo_rank_2);
////////////////////////////////////calculo de la distancia
,,,primero calculo la suma con los ranking
suma_con_abs=0;

```

```

    for (i=0;i<32;i++){
        suma_con_abs= suma_con_abs + abs(matriz_u_1_rank[i][1]-
            matriz_u_2_rank[i][1])*prob_pal_1[i]*prob_pal_2[i];
    }
    printf("\nsuma_con_abs:%f\n", suma_con_abs);
suma_sin_abs=0;

    for (i=0;i<32;i++){
        suma_sin_abs= suma_sin_abs + prob_pal_1[i]*prob_pal_2[i];
    }
    printf("\nsumasin_abs:%f\n", suma_sin_abs);
//////////7 distancia
    archivo_distancia=fopen("distancia_log_pink_N.dat","w");
    D = suma_con_abs*1. / 31*suma_sin_abs*1.;

    printf("\ndistancia_total_log_pink:_%f\n",D);
    fprintf(archivo_distancia, "%f", D);

}

```

15. Apéndice H: Código fuente para el cálculo del ruido Rosa

```

%=====
% GenRuidosfkseriesLargas.m
% este programa utiliza el generador randn para generar ruido 1/f^k
% de cada generador hace 10 versiones
%=====

%clear;clc;
format long;

Tsam=1, k=input('k? ')
for igen=1:100

igen
    gen1=rand(100000,1);% ojo que por ahi tarda mucho
    gen=gen1-mean(gen1);% este generador es ruido blanco sin componente de
        continua
    N=length(gen);% se mide el largo del vector
    %
    %
    fsam=1/Tsam;% esta es la frecuencia de muestreo
    f = fsam*(1:(N/2))/(N);f=f'; %La frecuencia en Hz en funcion de la
        frecuencia de muestreo.
    flog=log(f);flog=flog';% eje logaritmico para los graficos logaritmicos en
        frecuencia
    Nfreq=length(f);% el eje de frecuencias se representa hasta la mitad de los
        datos
    %=====

    %FFT del ruido original (blanco)
    %=====

    y1=fft(gen,N); %fft del ruido blanco(igual longitud que el archivo gen)
    y1=y1/N;% se normaliza el vector
    Py1=y1.*conj(y1);% se calcula el espectro de potencia
    Py1log=log(Py1);% se calcula en escala logaritmica. No usar el log log de
        la figura porque asi
    %
    %=====

%AHORA VAMOS A GENERAR RUIDO 1/f^k
%=====

```

```

auxi=sqrt(f);% voy a multiplicar por 1/sqrt(f) el vector obtenido en la FFT
auxi=auxi.^k;
auxi=1./auxi;% aqui invert?
y2=y1(1:Nfreq).*auxi;% este era la FFT del generador. La multiplico por
    auxi
Py2=y2.*conj(y2);% se calcula el espectro de potencia
Py2log=log(Py2);% se calcula en escala logaritmica. No usar el log log de
    la figura porque asi
%no se puede luego fitear
for j=1:Nfreq
    y3(j)=y2(j);
    y3(Nfreq+j)=real(y2(Nfreq-j+1))-i*imag(y2(Nfreq-j+1));%para funciones
        reales F(w)=F*(-w)
end
y3=y3';
Py3=y3.*conj(y3);% se calcula el espectro de potencia. Debiera ser igual al
    de y3.
Py3log=log(Py3);% se calcula en escala logaritmica. No usar el log log de
    la figura porque asi
%no se puede luego fitear
%
genk1=ifft(y3,N)*Nfreq;%la ifft puede tener peque?os valores imaginarios.
regen=real(genk1);
ingen=imag(genk1);
%ave(['gen',num2str(igen),'f',num2str(k)'.txt'],'regen','-ascii','-
    double')

for j=1:length(regen)
M_ruidos(:,igen)=regen;
end
end

```

Referencias

1. ALBERT C.-C. YANG:, SHU-SHYA HSEU y HUEY-WEN YIEN . “Linguistic Analysis of the Human Heartbeat Using Frequency and Rank Order Statistics” ; VOLUME 90, NUMBER 10; 14 MARCH 2003, PHYSICAL REVIEW LETTERS.
2. PETER J. BROCKWELL: “Time Series: Theory and Methods” Second edition
3. FRANCISCO TORRE: “Determinismo, Caos, Azar e Incertidumbre“
4. DAVID LUENGO GARCIA: “Estimación Óptima de Secuencias Caóticas con Aplicaciones en Comunicaciones”
5. HALLEY M. JOHN: “Extinction Risk and the $\frac{1}{\nu}$ Family of Noise Models”
6. IVO GROSSE, PEDRO GALVAN, RAMÓN ROLDAN, JOSÉ OLIVER: “Analysis of symbolic sequences using the Jensen-Shannon divergence”
7. JAVIER SPARACINO: “Cuantificadores de Información en Física” Tesis Doctoral en Física, FaMAF
8. ANA MAJTEY: “Sobre la Noción de Distancia en Física Estadística y Mecánica Cuántica” Tesis Doctoral en Física, FaMAF
9. CHRISTOPH BANDT, BERND POMPE: “Permutation entropy — a natural complexity measure for time series”
10. GUILLERMO ABRAMSON: “Sistemas Dinámicos y Caos”
11. OSBORNE A.R , PROVENZALE A.:(1989) Physics D 35: 357-381
12. SUGIHARA G. , MAY R.M.:(1983) Nature 344: 734-741
13. KAPLAND.T , GLASS L.:(1992) Physical Rev. Lett. 68: 427-430
14. KANTZ H. , OLBRICH E.:(2000) Physics A 280: 34-48
15. KOLMOGOROV A.N.: (1959) Dokl. Akad. Nauk USSR 119: 861-864
16. SINAI Y.G.: (1959) Dokl. Akad. Nauk. USSR 124: 768-771
17. ROSSO O. , LARRONDO H.A.(2007) Physical Rev. Lett.: Distinguishing Noise from Chaos
18. LOPEZ-RUIZ R. , MANCINI H., CALBET, X.: (2010) A Statistical Measure of Complexity
19. COVER THOMAS: “Elements of information theory”

