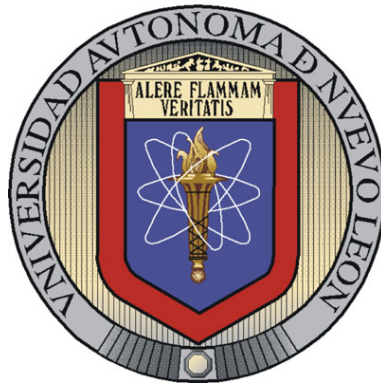


UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE INGENIERÍA MECÁNICA Y ELÉCTRICA

DIVISIÓN DE ESTUDIOS DE POSGRADO



NUEVAS TÉCNICAS DE PROCESOS ESTOCÁSTICOS Y APRENDIZAJE  
ESTADÍSTICO PARA SERIES DE TIEMPO

POR

DEXMONT ALEJANDRO PEÑA CARRILLO

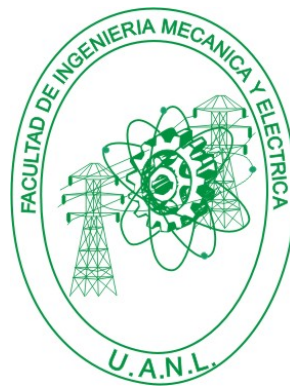
TESIS

EN OPCIÓN AL GRADO DE  
MAESTRO EN CIENCIAS EN INGENIERÍA DE SISTEMAS

JUNIO 2008

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE INGENIERÍA MECÁNICA Y ELÉCTRICA  
DIVISIÓN DE ESTUDIOS DE POSGRADO



NUEVAS TÉCNICAS DE PROCESOS ESTOCÁSTICOS Y APRENDIZAJE  
ESTADÍSTICO PARA SERIES DE TIEMPO

POR

DEXMONT ALEJANDRO PEÑA CARRILLO

TESIS

EN OPCIÓN AL GRADO DE  
MAESTRO EN CIENCIAS EN INGENIERÍA EN SISTEMAS

JUNIO 2008

**Universidad Autónoma de Nuevo León**  
**Facultad de Ingeniería Mecánica y Eléctrica**  
**División de Estudios de Posgrado**

Los miembros del Comité de Tesis recomendamos que la Tesis “NUEVAS TÉCNICAS DE PROCESOS ESTOCÁSTICOS Y APRENDIZAJE ESTADÍSTICO PARA SERIES DE TIEMPO”, realizada por el alumno DEXMONT ALEJANDRO PEÑA CARRILLO, matrícula 1051394, sea aceptada para su defensa como opción al grado de Maestro en Ciencias en Ingeniería de Sistemas.

El Comité de tesis

---

Dr. J. Arturo Berrones Santos  
Asesor

---

Dra. Satu Elisa Schaeffer  
Revisor

---

Dr. Igor Litvinchev  
Revisor

Vo. Bo.

---

Dr. Guadalupe Alan Castillo Rodríguez  
Subdirector  
División de Estudios de Posgrado

---

# DEDICATORIA

---

A mis padres,

*Marcial Peña Ovalle,  
Ma. Concepción Carrillo Muñoz*

---

# AGRADECIMIENTOS

---

Agradezco a la coordinación del programa del PISIS por permitirme ser parte de él. A la Universidad Autónoma de Nuevo León y a su Facultad de Ingeniería Mecánica y Electrica por haberme otorgado apoyo en forma de becas de inscripción y colegiatura respectivamente. Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por haberme otorgado beca de manutención durante la duración de la maestría.

Agradezco profundamente al Dr. Arturo Berrones Santos quien como mi asesor de tesis me apoyó y brindo su tiempo constantemente para resolver las dudas que se presentaron durante el desarrollo de la tesis, además de darme comentarios de aliento.

A los miembros del comité de tesis Dra. Satu Elisa Schaeffer y Dr. Igor Litvinchev por sus valiosas recomendaciones para mejorar la calidad de la tesis.

A mi familia, profesores, amigos y compañeros por su apoyo durante mi estancia en el programa.

---

# PREFACIO

---

El trabajo desarrollado en esta tesis trata el área de pronóstico en series de tiempo y aprendizaje estadístico. Se presenta un algoritmo para pronóstico en series de tiempo con relaciones lineales que solamente requiere la estimación de un parámetro, además de que solo utiliza operaciones lineales para su funcionamiento, permitiendo así un computo eficiente. Para mostrar la eficiencia de este algoritmo se realiza una comparación con un modelo clásico de pronóstico, obteniendo resultados comparables a éste y en algunos casos mejores.

Para el caso de aprendizaje estadístico, se presenta un algoritmo para la obtención de la función de densidad de probabilidad asociada un problema de optimización no lineal continuo. Este algoritmo utiliza solamente operaciones lineales para su funcionamiento. Es probado con distintos problemas de referencia obteniendo muy buenos resultados en la localización del punto óptimo. Es probado además para entrenar una red neuronal que pronostica una serie de tiempo con relaciones no lineales, presentando un excelente desempeño. Además del algoritmo utilizado, se presenta una metodología para analisis de series de tiempo no lineales.

Con la introducción de estos algoritmos se cubren los capos de series de tiempo lineales y no lineales, utilizando algoritmos que requieren solamente operaciones lineales obteniendo así un conjunto de herramientas para pronóstico que presentan un bajo costo computacional.

---

# ÍNDICE GENERAL

---

<b>Dedicatoria</b>	<b>III</b>
<b>Agradecimientos</b>	<b>IV</b>
<b>Prefacio</b>	<b>V</b>
<b>Índice general</b>	<b>V</b>
<b>Resumen</b>	<b>1</b>
<b>1. Introducción</b>	<b>3</b>
1.1. Descripción del problema . . . . .	4
1.2. Objetivo . . . . .	5
1.3. Justificación . . . . .	6
1.4. Contribuciones . . . . .	7
<b>2. Marco teórico</b>	<b>9</b>
2.1. Importancia de un pronóstico . . . . .	9
2.2. Pronóstico de series de tiempo . . . . .	10
2.2.1. Resúmenes gráficos . . . . .	11
2.2.2. Resúmenes numéricos . . . . .	14
2.2.3. Medidas de error de pronóstico . . . . .	17
2.2.4. Operaciones con series de tiempo . . . . .	21
2.3. Modelos matemáticos estocásticos y determinísticos . . . . .	24
2.4. Procesos estocásticos estacionarios . . . . .	25

2.4.1.	Media y varianza de un proceso estacionario . . . . .	25
2.4.2.	Autocovarianza y coeficientes de autocorrelación . . . . .	26
2.4.3.	Matriz de autocovarianza y definida positiva . . . . .	27
2.4.4.	Condiciones satisfechas por las autocorrelaciones de un pro- ceso estacionario . . . . .	28
2.4.5.	Estacionariedad de funciones lineales . . . . .	29
2.4.6.	Proceso Gaussiano . . . . .	29
2.4.7.	Estacionariedad débil . . . . .	30
2.5.	Funciones de Autocovarianza y Autocorrelación . . . . .	30
2.5.1.	Estimación de las funciones de Autocovarianza y autocorrela- ción . . . . .	31
2.6.	Representaciones de Fourier para series de tiempo . . . . .	32
2.7.	La transformada de Fourier para muestras tomadas de datos discretos	35
2.8.	La transformada de Fourier en tiempo discreto . . . . .	36
2.8.1.	Propiedades espectrales de modelos estacionarios . . . . .	37
2.9.	Modelos no lineales para series de tiempo . . . . .	40
2.9.1.	Resúmenes gráficos . . . . .	41
2.10.	Estacionariedad, pruebas para determinarla . . . . .	42
2.11.	Metodos del espacio de Fases . . . . .	43
2.11.1.	Determinismo, unicidad en el espacio de fases . . . . .	43
2.11.2.	Reconstrucción de retardos . . . . .	45
2.11.3.	Obtención de dimensiones de recurrencia: Vecinos falsos . . .	46
2.12.	Inspección visual de datos . . . . .	48
2.13.	Gráficas de recurrencia . . . . .	49
2.14.	Redes neuronales . . . . .	51
2.14.1.	El problema de filtrado adaptativo . . . . .	54
<b>3.</b>	<b>Modelo</b>	<b>57</b>
3.1.	Modelos lineales para series de tiempo . . . . .	57



3.2. El modelo AR . . . . .	57
3.2.1. Razón de la utilización del modelo AR . . . . .	59
3.3. El modelo KNNR . . . . .	60
3.3.1. Redes de conocimiento . . . . .	61
3.4. Redes de conocimiento y análisis de series de tiempo . . . . .	64
3.5. Obtención de funciones de densidad para problemas de optimización no lineales . . . . .	70
3.6. Algoritmo Fokker–Planck Estacionario . . . . .	71
<b>4. Descripción del problema</b>	<b>74</b>
4.1. Series de tiempo lineales, el modelo KNNR . . . . .	74
4.2. Series de tiempo no lineales, obtención de funciones de densidad . .	75
<b>5. Experimentación</b>	<b>77</b>
5.1. Comparación de los modelos AR y KNNR . . . . .	77
5.1.1. Combinaciones lineales de funciones coseno con ruido aditivo	78
5.1.2. Mapas no lineales . . . . .	79
5.1.3. Procesos estocásticos lineales . . . . .	82
5.1.4. Datos experimentales . . . . .	84
5.2. Obtención de funciones de densidad para problemas de optimización no lineales . . . . .	86
5.2.1. Función de Levy No. 5 . . . . .	87
5.2.2. Problema XOR . . . . .	90
5.2.3. Red neuronal para aprender el mapa logístico . . . . .	92
<b>6. Conclusiones</b>	<b>95</b>
6.1. Trabajo futuro . . . . .	96
<b>Índice de figuras</b>	<b>99</b>
<b>Índice de cuadros</b>	<b>102</b>

---

<b>Lista de símbolos</b>	<b>104</b>
<b>Lista de abreviaciones</b>	<b>105</b>

---

# RESUMEN

---

Dexmont Alejandro Peña Carrillo

Candidato para el grado de Maestro en Ciencias en Ingeniería de Sistemas

Universidad Autónoma de Nuevo León

Facultad de Ingeniería Mecánica y Eléctrica

Titulo del trabajo: Nuevas técnicas de procesos estocásticos y aprendizaje estadístico para series de tiempo

El presente trabajo trata el área de pronóstico en series de tiempo con relaciones lineales y aprendizaje estadístico. En el caso de series de tiempo lineales se presenta un algoritmo para pronóstico que requiere la estimación de solamente un parámetro para su funcionamiento. Para el caso del aprendizaje estadístico se presenta un algoritmo capaz de estimar la función de densidad asociada a un problema de optimización no lineal sin restricciones y con variables continuas.

Entre los principales aportes de este trabajo se encuentra una comparación del modelo presentado con un modelo clásico de pronostico en series de tiempo lineales, mostrando su efectividad bajo distintos escenarios de prueba, tanto reales como simulados. Se resuelve la problemática de una empresa de la localidad, la cual

consistía en determinar la posibilidad de realizar un pronóstico para una familia de productos de la cual se contaba con una cantidad de datos considerablemente pequeña, sin embargo, se obtuvo un pronóstico cumpliendo holgadamente las expectativas de la empresa.

En el caso de aprendizaje estadístico se presenta un algoritmo con el que es posible obtener una expresión para la función de densidad de probabilidad asociada al problema de optimización. Una de las ventajas de que presenta este algoritmo sobre otros métodos radica en que utiliza solamente operaciones lineales para realizar la estimación de la función de densidad, además de que es posible especificar la complejidad de computo que se desea realizar. Este método es probado con distintos problemas de prueba

Entre otros resultados se presenta la metodología utilizada para el tratamiento de series de tiempo con relaciones no lineales, ya que estas fueron utilizadas para probar el éxito obtenido al tratar el problema de aprendizaje de una red neuronal mediante el algoritmo de aprendizaje estadístico propuesto.

---

Dr. Arturo Berrones Santos

Asesor

# INTRODUCCIÓN

---

Al momento de analizar el funcionamiento de una organización eficiente, es frecuente encontrar que estas invierten una cantidad considerable de tiempo, esfuerzo y personal en realizar una planeación de lo que se realizará a distintos intervalos de tiempo. Esta necesidad surge debido a que es indispensable prever las situaciones a las que se debe enfrentar, tales como, variación en la demanda de algún producto, precio de los materiales, etc. Sin embargo el tiempo de espera entre las planeación y la ejecución de las acciones puede variar desde unos cuantos días u horas como en el caso de planes de producción o transportación, hasta varios años, como es el caso de las inversiones de capital. Existen diversas herramientas utilizadas para realizar una buena planeación, sin embargo, en el juego del día a día, el pronóstico ha obtenido papel cada vez más importante en la toma de decisiones.

El pronóstico es utilizado en diversos campos de toda organización, los cuales van del tipo cuantitativos como en la estimación de la manera en que ocurre un fenómeno de interés, el cual puede ser las ventas realizadas durante el año en un supermercado o la intensidad de una señal en circuitos electrónicos; hasta del tipo cualitativo, el cual ocurre por ejemplo, cuando un médico es capaz de diagnosticar alguna enfermedad analizando los latidos del corazón y la presión sanguínea del paciente por un determinado lapso de tiempo.

Un caso de utilización de pronósticos muy conocido por todos está en el análisis climatológico. En esta situación se usa como referencia la variación de temperatura, humedad, velocidad y dirección del viento prevaleciente para pronosticar el estado del clima los siguientes días.

## 1.1. DESCRIPCIÓN DEL PROBLEMA

En el campo de pronósticos cuantitativos se han desarrollado diferentes métodos para predecir el futuro. Sin embargo un problema encontrado frecuentemente es que los modelos obtenidos para pronosticar el fenómeno deseado contienen una gran cantidad de parámetros, los cuales deben ser ajustados a los datos obtenidos para lograr el mejor desempeño posible. Este procedimiento de ajuste es costoso desde el punto de vista computacional. Debido a esto, un método capaz de calcular los valores de los parámetros de una manera eficiente es de gran importancia.

Para clasificar los tipos de métodos existentes para optimizar los parámetros de los modelos, es necesario primero clasificar los modelos en sí; estos pueden dividirse en *modelos lineales*, los cuales buscan una relación lineal entre los datos de entrada y salida para realizar su pronóstico; y *modelos no lineales*, los cuales presentan una relación no lineal entre los datos de entrada y la salida del sistema.

En el contexto de modelos lineales, es común encontrarse con el problema de ajustar el modelo a series de tiempo que presentan múltiples periodos combinados, es decir, contienen más de una componente periódica y se encuentran mezcladas. Esto es de particular importancia ya que ocurre al tratar de obtener información en distintas escalas de tiempo, como puede ser a nivel mes, semanas, días e incluso horas, en las cuales se mezclan ciclos a nivel mensual con ciclos semanales y diarios, por ejemplo.

Hay ocasiones en las cuales una relación lineal no es suficiente para representar los datos de la vida real, en estos casos se utilizan modelos no lineales, los cuales además de la problemática que presentan los problemas lineales, sufren del hecho que evaluar una función no lineal es computacionalmente más costoso que la evaluación de una función lineal.

Debido a esto es necesario aprovechar de una mejor manera la linealidad de los datos cuando esto existe, utilizando modelos que utilicen una menor cantidad de parámetros además de tener un número fijo de los mismos.

Para el caso de los métodos no lineales, es necesario obtener algoritmos que permitan un ajuste de parámetros eficiente y de ser posible obtener información estadística de los mismos.

## 1.2. OBJETIVO

Debido a que en la vida real es común encontrar la necesidad de obtener un pronóstico en intervalos de tiempo muy cortos (del orden de segundos o minutos), ya que de esto dependen tareas críticas como el cálculo de cantidad de combustible que se debe proporcionar a un avión, o el diagnóstico de un paciente con alguna deficiencia nerviosa. El objetivo de este trabajo es presentar herramientas con un bajo costo computacional, y que son capaces de obtener resultados comparables a los obtenidos mediante métodos más complejos y en algunos casos mejores. Con el fin de alcanzar este objetivos se deben cumplir las siguientes metas a corto plazo:

- Presentar de un algoritmo lineal para pronóstico capaz de obtener un buen pronóstico en series de tiempo que presentan múltiples periodos utilizando solamente un parámetro, así como una comparación con otros métodos exis-

tentes. En este caso el modelo AR bajo distintos escenarios, uno de los cuales son datos de ventas obtenidos de una empresa del ramo alimenticio de la localidad.

- Obtener un método capaz de obtener la función de densidad de probabilidad asociada con el problema de ajuste de parámetros en métodos de pronóstico no lineales. Además de evaluar este método bajo distintos escenarios.

### 1.3. JUSTIFICACIÓN

Debido a la gran importancia de un pronóstico en las organizaciones es necesario contar con herramientas capaces de obtener resultados de calidad bajo distintos escenarios. El algoritmo de pronóstico lineal es de reciente aparición, por lo que no ha sido probado extensivamente. Debido a esto, en este trabajo se realizan distintas pruebas en diversos escenarios, además de comparar sus resultados contra un algoritmo clásico de pronóstico con el fin de comparar su desempeño.

Para el caso de problemas no lineales, debido a que la evaluación de función de costo es costosa desde el punto de vista computacional, es necesario conocer información acerca de los parámetros del problema. En este caso, una expresión para la función de densidad de probabilidad permite realizar un análisis estadístico de como se comportan estos parámetros, permitiendo así la mejor elección de los mismos.



## 1.4. CONTRIBUCIONES

En este trabajo se presenta un algoritmo que aprovecha relaciones lineales en los datos de entrada para realizar su pronóstico, además presenta la ventaja de requerir la evaluación de solamente un parámetro, el cual es calculado a partir de los datos de entrada, con esto se logra reducir el tiempo de computo además de obtener resultados comparables y en algunos casos mejores que los obtenidos mediante métodos más complejos. Se realiza una comparación contra un modelo clásico de pronóstico y se presentan los resultados. Este modelo es probado bajo diversos escenarios, uno de los cuales es un problema del mundo real, en donde se pronostican las ventas a nivel nacional de una empresa del ramo alimenticio de la localidad.

Para el caso en que se presentan relaciones no lineales se presenta un método capaz de obtener la función de densidad de probabilidad asociadas al problema de ajuste de parámetros. Éste método es capaz de obtener una expresión matemática para la función de densidad con muy pocas evaluaciones de la función objetivo, permitiendo además utilizar la expresión obtenida para obtener estadísticas del comportamiento del problema tratado.

Como resultado del presente trabajo se realizó la escritura de tres artículos, dos de ellos en publicaciones internacionales y una más se encuentra en preparación. Las publicaciones son las siguientes:

- Dexmont Peña, Ricardo Sánchez, Arturo Berrones: Stationary Fokker - Planck Learning for the Optimization of Parameters in Nonlinear Models. MICAI 2007: 94-104
- Arturo Berrones, Dexmont Peña, Ricardo Sanchez, Stationary Density of Stochastic Search Process. Encyclopedia of Artificial Intelligence 2008.
- Arturo Berrones, Dexmont Peña. Identification of Relevant Frequencies and

---

Forecasting of Short Time Series Under Strong Additive Noise. En preparación.

## MARCO TEÓRICO

---

En este capítulo se presenta la documentación necesaria para entender el contexto en el que figura el aporte de este trabajo, por lo que se hace referencia a teoría tomada de distintas fuentes, las cuales se enumeran a continuación: importancia de un pronóstico [9, 14], pronóstico en series de tiempo [5, 9, 14], modelos estocásticos [5], procesos estocásticos estacionarios [5], Funciones de autocovarianza [5, 14], representaciones de fourier para series de tiempo [9], la transformada de fourier [9, 28], modelos no lineales para series de tiempo [11], pruebas para determinar estacionalidad [11], metodos del espacio de fases [11], inspección visual de datos [9, 11], graficas de recurrencia [11], redes neuronales [8].

### 2.1. IMPORTANCIA DE UN PRONÓSTICO

Toda organización que desea ser eficiente, se encuentra ante el problema de realizar una buena planeación de sus actividades a todos los niveles de la misma. Sin embargo el tiempo de espera entre esta planeación y la obtención de los resultados varia enormemente, puede ser desde años, como en el caso de inversiones de capital; unos cuantos días u horas, como en el diseño de planes de producción o transportación; hasta segundos como es el cálculo de la respuesta del impulso de un

motor en un automóvil o un avión. Debido a esto es indispensable contar con herramientas de pronóstico capaces de proporcionar resultados de una manera efectiva y eficiente.

Una manera de definir la palabra *pronóstico* es predecir los eventos futuros en base a la información con que se cuenta y que esta en el presente o el pasado. El pronóstico es utilizado para predecir el valor de *señales* en el futuro, las cuales se definen formalmente como la función de una o más variables que transportan información acerca de la naturaleza de un fenómeno físico. Ejemplo de señal puede ser la voz en una conversación, ya sea frente a frente o a través de un canal telefónico. Otro ejemplo que presenta señales de interés, ocurre cuando al escuchar los latidos del corazón de un paciente y el monitoreo de su presión sanguínea, un médico es capaz de diagnosticar alguna lesión o enfermedad.

Debido a la necesidad de obtener pronósticos se han desarrollado diversos modelos. Todos ellos trabajan sobre un conjunto de datos que provienen de una secuencia de se observaciones en el tiempo, al cual se le llama *serie de tiempo*. Estas observaciones provienen de una función continua  $y(t)$  (la señal) muestreada a una tasa de tiempo uniforme.

## 2.2. PRONÓSTICO DE SERIES DE TIEMPO

Dada la definición de una serie de tiempo, se tiene que sus valores provienen de una función continua  $y(t)$ , por lo que el *pronóstico de la serie de tiempo* consiste en obtener el valor de la función  $y(t+k)$  en donde  $t$  es el tiempo actual y  $k$  es el periodo de tiempo que se desea pronosticar.

### 2.2.1. RESÚMENES GRÁFICOS

La manera más obvia de representar una serie de tiempo es una *gráfica de tiempo*, en la cual los datos se grafican sobre tiempo. Esta representación inmediatamente revela si existe alguna tendencia, algún comportamiento estacional o alguna otra característica sistemática en los datos.

Existen distintas clasificaciones para una serie de tiempo en función de las características de la señal que la produce. Se dice que se tiene una *señal par* si satisface la condición  $y(-t) = y(t)$  para todo  $t$ . Una señal es *impar* si satisface la condición  $y(-t) = -y(t)$  para todo  $t$ . En otras palabras las señales pares son simétricas en torno al eje vertical u origen del tiempo, en tanto que las señales impares son anti-simétricas en torno al origen del tiempo.

También se tiene que una señal puede ser *periódica* si es una función que satisface  $y(t) = y(t + T)$  para todo  $T$ , donde  $T$  es una constante positiva. Cualquier señal que exista valor de  $T$  que cumpla la condición anterior recibe el nombre de *señal aperiódica o no periódica*.

Otra clasificación de una señal es en *señal determinista* la cual es aquella que no presenta incertidumbre con respecto a su valor en cualquier tiempo. En consecuencia, las señales deterministas pueden modelarse como funciones de tiempo completamente especificadas. Una señal que presenta incertidumbre en su ocurrencia real recibe el nombre de *señal aleatoria*.

Un paso importante en la selección del modelo de pronóstico apropiado es considerar los tipo de patrones en los datos, de tal manera que los métodos mas apropiados para estos patrones sean utilizados. Se pueden distinguir cuatro tipo de patrones comunes: *horizontal*, este existe cuando los datos fluctúan alrededor de una media (a esta serie se le llama *estacionaria* alrededor de la media); *estacional*, existe

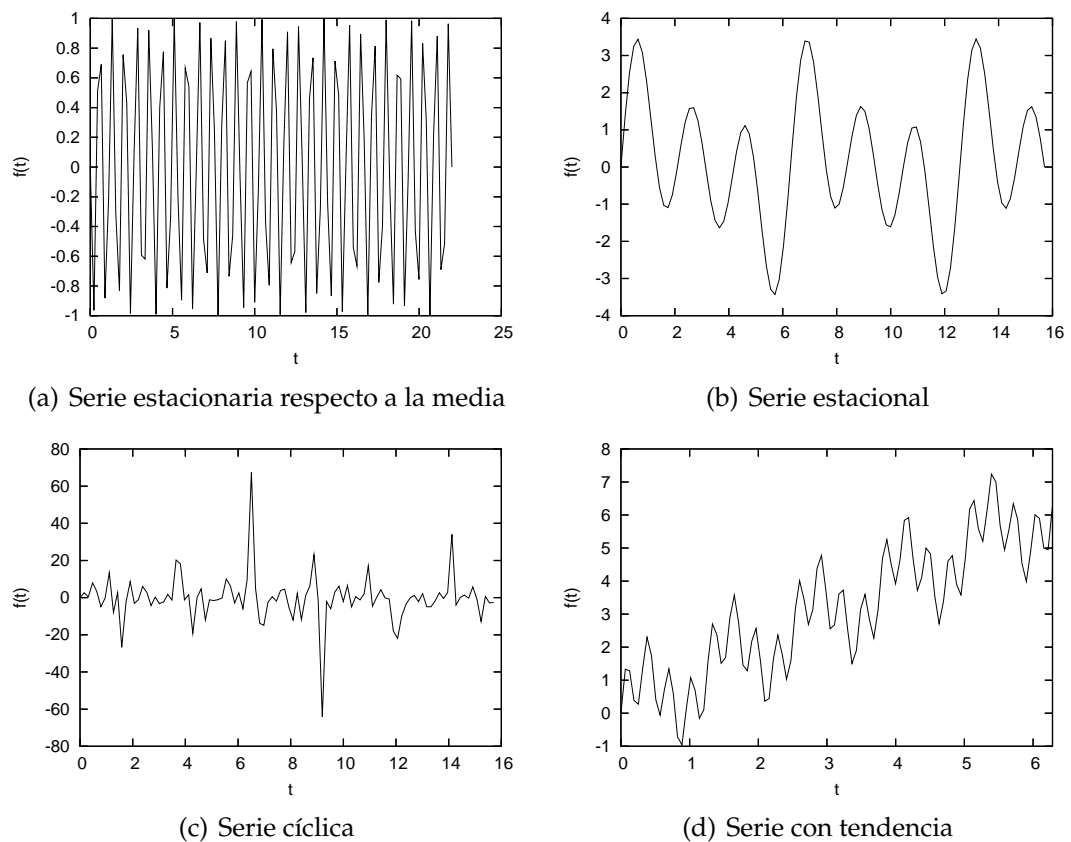


Figura 2.1: Distintos patrones encontrados en una serie de tiempo

cuando la serie es influenciada por un factor de estación, es decir, que ocurre cada determinado periodo de tiempo; se llama *periódicas* a una serie que repite sus valores cada cierto numero de intervalos de tiempo, algunas series también son llamadas de esta manera aunque no se repiten exactamente en cada periodo; patrón *cíclico* se le llama cuando los incrementos y disminución de los datos no son en un periodo fijo; *tendencia*, se le llama cuando hay un incremento o decremento a largo plazo en los datos. Un ejemplo de esto se puede apreciar en la figura 2.2.1. Es común encontrar combinación de los patrones mencionados anteriormente.

Para series de datos que son estacionales, frecuentemente se utiliza una *gráfica estacional*. Esta es similar a una gráfica de tiempo, sin embargo los datos de cada estación son traslapados. Una gráfica estacional permite que se visualice con mayor

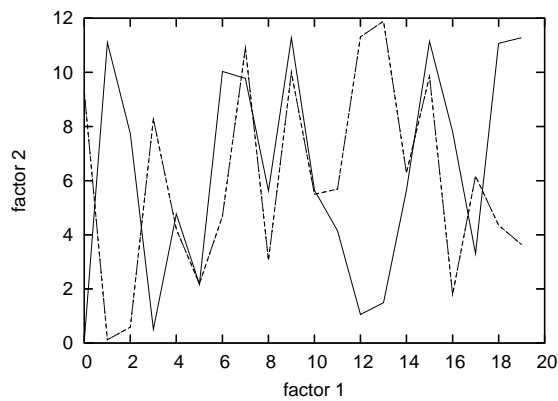


Figura 2.2: La gráfica estacional permite visualizar claramente los patrones estacionales, en este caso se ve como los periodos 9 y 15 presentan incrementos en ambas gráficas

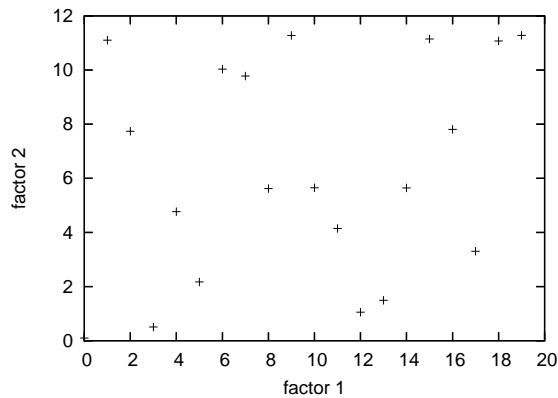


Figura 2.3: La gráfica de dispersión muestra la relación entre dos factores claridad el patrón estacional como se puede ver en la figura 2.2.

Otro tipo de gráfica es el de *dispersión* en la cual se grafican dos factores, uno contra el otro. Este tipo de gráfica muestra la relación entre estos dos factores, esto se puede apreciar en la figura 2.3.

### 2.2.2. RESÚMENES NUMÉRICOS

Además de gráficas, los resúmenes numéricos son de mucha utilidad. Para un conjunto de datos univariado, estadísticas descriptivas comunes son la media de la muestra, la desviación estándar y la varianza. Para conjuntos de datos bivariados es importante describir como se relacionan los dos conjuntos de datos. Las estadísticas utilizadas comúnmente son la covarianza y la correlación.

La *media de la muestra*  $\bar{Y}$  se define como la suma de las observaciones en la serie de tiempo sobre la cantidad de observaciones, esto es

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N y_i \quad (2.1)$$

en donde  $N$  es la cantidad de observaciones en la serie de tiempo,  $y_i$  es cada uno de los valores de estas observaciones. Esto proporciona una medida de en que posición se encuentra el centro de la serie de tiempo.

Además del centro también existe una medida para determinar la dispersidad de los datos: se resta la media de la muestra  $\bar{Y}$  de cada observación  $y_i$  para obtener la  $i$ -ésima desviación con respecto a la media ( $\bar{Y} - y_i$ ). Otra medida es la desviación promedio absoluta (MAD por sus siglas en ingles) la cual se obtiene por la formula

$$\text{MAD} = \frac{1}{N} \sum_{i=1}^N |y_i - \bar{Y}|. \quad (2.2)$$

Otra medida utilizada comunmente es la desviación cuadrada media (MSD por sus siglas en ingles), la cual se define como

$$\text{MSD} = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{Y})^2. \quad (2.3)$$



Muy relacionada a la desviación media cuadrada (MSD) se encuentra la *varianza de la muestra*  $S^2$ , la cual se define como la suma de las desviaciones al cuadrado dividido entre el número de observaciones menos 1. Se divide entre  $N - 1$  debido a que este es el número de grados de libertad, el número de grados de libertad se define como el número de observaciones en la muestra menos el número de parámetros estimados. La formula para la varianza es la siguiente:

$$S^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{Y})^2. \quad (2.4)$$

La varianza devuelve una cantidad en unidades cuadradas, para obtener una medida en las mismas unidades que los datos de la muestra se define la *desviación estándar*  $S$ , la cual es la raíz cuadrada de la varianza.

$$S = \sqrt{S^2}. \quad (2.5)$$

Para el caso de series de tiempo bivariadas se tiene la *covarianza*, la cual indica como estas dos variables varían en conjunto. La covarianza se define de la siguiente manera:

$$\text{Cov}_{XY} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{X})(y_i - \bar{Y}) \quad (2.6)$$

en donde  $x_i, y_i$  es el valor de las variables en la muestra y  $\bar{X}, \bar{Y}$  son las medias de cada una de las muestras.

Un problema encontrado con la covarianza son las unidades y escala de los posible valores que puede tomar. Para resolver este problema se define el *coeficiente de correlación*  $r$ , el cual es una medida de la covarianza que se encarga del problema de escala. Si la covarianza ( $\text{Cov}_{XY}$ ) es dividida por dos desviaciones estándar ( $S_X$  y  $S_Y$ ) las unidades se cancelan dejando un numero adimensional, el cual es el

coeficiente de correlación entre  $X$  e  $Y$ . Esto se escribe de la siguiente manera:

$$r_{XY} = \frac{\text{Cov}_{XY}}{S_X S_Y} = \frac{\sum_{i=1}^N (x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (y_i - \bar{Y})^2} \sqrt{\sum_{i=1}^N (x_i - \bar{X})^2}}. \quad (2.7)$$

El efecto de escalar es para restringir el rango de  $r_{XY}$  al intervalo  $[-1,1]$ . Sin embargo al utilizar esta medida se debe recordar que se esta explicando una asociación lineal entre las dos variables, por lo que no es apropiado aplicar esta medida cuando hay una relación no lineal pronunciada.

En series de tiempo además existen la *autocorrelación* y la *autocovarianza*, las cuales tienen el mismo propósito, el cual es identificar relaciones descriptivas en la serie de tiempo. Si se comparan por ejemplo  $Y_t$  (observación al tiempo  $t$ ) con  $Y_{t-1}$  se puede ver como se relacionan las observaciones. La observación  $Y_{t-1}$  es llamada con retardo (inglés: lag).

De la misma manera es posible comparar observaciones con retardos de dos, tres, etc. periodos de tiempo. El nombre de autocovarianza y autocorrelación proviene de que se calculan tomando en cuenta la misma serie de datos como si fueran dos distintas.

$$c_k = \frac{1}{N} \sum_{t=k+1}^N (y_t - \bar{Y})(y_{t-k} - \bar{Y}) \quad (2.8)$$

$$r_k = \frac{\sum_{t=1}^N (y_t - \bar{Y})(y_{t-k} - \bar{Y})}{\sum_{t=1}^N (y_t - \bar{Y})^2} \quad (2.9)$$

en donde  $N$  es la cantidad de datos de la serie de tiempo.

Al calcular la autocorrelación con retardos de 1, 2, ..., se obtiene la *función de autocorrelación* o ACF (por sus siglas en ingles). En vez de analizar una serie de números es más sencillo graficar la autocorrelación contra el retardo. A esta gráfica se

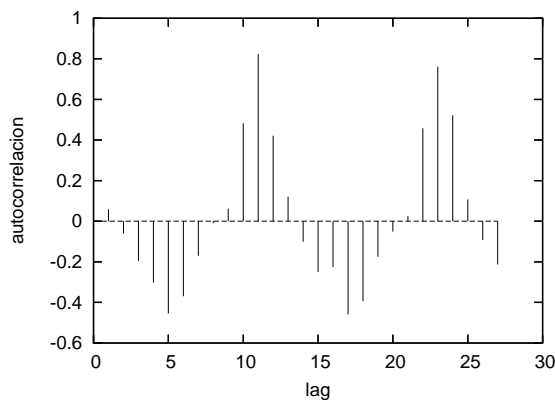


Figura 2.4: El correlograma ayuda visualizar la función de autocorrelación de una manera rápida y sencilla.

le llama *correlograma* y ayuda a visualizar la ACF de una manera rápida y sencilla, como se puede ver en la figura 2.4.

El correlograma es una herramienta estándar en la exploración de una serie de tiempo antes de realizar un pronóstico. Éste provee una verificación útil para estacionalidad, ciclos, y otros patrones. Además la ACF ayuda a identificar si los valores anteriores en la serie de tiempo contienen información acerca de los siguientes valores, o si existe alguna relación entre una observación y la siguiente.

### 2.2.3. MEDIDAS DE ERROR DE PRONÓSTICO

Otro aspecto fundamental es como medir que tan bien se ajusta un pronóstico dado a un conjunto de datos. En algunas ocasiones a esto se le llama *bondad de ajuste*, la cual se refiere a que tan bien se pueden reproducir los datos conocidos por el modelo de pronóstico. Para el usuario de un pronóstico, esto es, la exactitud del pronóstico futuro.

Si  $y_t$  es el valor de la observación actual y  $f_t$  es el pronóstico para el mismo

periodo, se define el *error de pronóstico* como

$$e_t = y_t - f_t. \quad (2.10)$$

Usualmente  $f_t$  se calcula utilizando los datos  $y_1, \dots, y_{t-1}$ . Esto es un *pronóstico a un paso* ya que se pronóstica un periodo adelante de la ultima observación. Consecuentemente se define  $e_t$  como *error de pronóstico a un paso*. Esto es, la diferencia entre la observación  $y_t$  y el pronóstico obtenido.

#### ERROR DEPENDIENTE DE LA ESCALA

El error de pronóstico  $e_t$  se encuentra en la misma escala que los datos de pronóstico. Debido a esto las medidas de error basadas en  $e_t$  son dependientes de la escala. Las medidas mas comunes son *Error medio* ME, *Error absoluto promedio* MAE, *Error cuadrado medio* MSE todos nombrados por sus siglas en ingles:

$$\text{ME} = \frac{1}{N} \sum_{t=1}^N e_t \quad (2.11)$$

$$\text{MAE} = \frac{1}{N} \sum_{t=1}^N |e_t| \quad (2.12)$$

$$\text{MSE} = \frac{1}{N} \sum_{t=1}^N e_t^2 \quad (2.13)$$

donde  $N$  es la cantidad de datos en la serie de tiempo.

Para el caso del ME, al realizar la sumatoria, los errores pueden cancelarse entre sí, por eso en las demás medidas se utiliza valor absoluto o cuadrados para eliminar los valores negativos y prevenir la cancelación. El problema con estos métodos es que son dependientes de la escala por lo que no pueden ser comparados

con diferentes series de datos. Para esto se presentan los errores porcentuales.

#### ERRORES PORCENTUALES

El *porcentaje de error* para el periodo  $t$  esta dado por  $p_t = 100 \frac{e_t}{y_t}$ . Los porcentajes de error tienen la ventaja de ser independientes de la escala. Una medida común de este tipo es *error porcentual absoluto promedio* MAPE por sus siglas en ingles:

$$\text{MAPE} = \frac{1}{N} \sum_{t=1}^N p_t \quad (2.14)$$

en donde  $N$  es la cantidad de datos que se promedia.

Las medidas basadas en porcentaje de error tienen la desventaja de que se vuelven infinitas o indefinidas si hay valores de 0 en la serie de datos, lo que es frecuente para datos intermitentes. Una desventaja del MAPE es que da mas peso a los errores positivos lo que propicia el uso del *error porcentual absoluto promedio simétrico* sMAPE (por sus siglas en ingles) el cual se define como:

$$\text{sMAPE} = \frac{1}{N} \sum_{t=1}^N \frac{|y_t - f_t|}{(y_t + f_t)/2} * 100. \quad (2.15)$$

Sin embargo si el valor de la serie es cercano a cero entonces el valor pronosticado será cercano a cero, lo que propicia una división por un número cercano a cero, además el sMAPE puede tomar valores negativos lo que puede provocar ambigüedad en su interpretación.

## ERRORES LIBRES DE ESCALA

El *error escalado absoluto promedio* MASE (por sus siglas en inglés) fue propuesto por Hyndman y Koehler [10] como una medida de exactitud de pronósticos aplicable sin los problemas encontrados en las otras medidas. Esta medida propone escalar los errores basados en el MAE de la muestra utilizando el método ingenuo el cual consiste en hacer ( $f_t = y_{t-1}$ ) generado para cada punto en la muestra. El error escalado es el siguiente:

$$q_t = \frac{e_t}{\frac{1}{N-1} \sum_{i=2}^N |y_i - y_{i-1}|}. \quad (2.16)$$

Este resultado es independiente de la escala de los datos. Un error escalado es menor a uno si proviene de un error menor que el promedio del pronóstico ingenuo de la muestra. De la misma manera es mayor que uno si el pronóstico es peor. El MASE es el siguiente

$$\text{MASE} = \frac{1}{N} \sum_{t=1}^N |q_t|. \quad (2.17)$$

La única circunstancia bajo la que el MASE puede ser infinito o indefinido es cuando todas las observaciones históricas son iguales. Una ventaja que presenta este método contra los demás es que no hace suposiciones respecto a la media de los datos por lo que puede ser aplicado a series con tendencia, estacionalidad, o demás patrones. Debido a que es libre de escala puede utilizarse para comparar los resultados en diferentes series de tiempo.

En el presente trabajo se utilizará el MASE para calcular el error de pronóstico en las distintas series de tiempo que se utilizarán.

#### 2.2.4. OPERACIONES CON SERIES DE TIEMPO

En ocasiones el ajustar los datos históricos lleva a un modelo de pronóstico más simple e interpretable. Con este fin se realizan distintas transformaciones para procesar o manipular los datos. Este aspecto suele implicar una combinación de algunas operaciones básicas.

*Escalamiento de amplitud.* Considere que  $y(t)$  denota una serie de tiempo, la serie resultante del escalamiento de amplitud aplicado a  $y(t)$  se define mediante

$$y'(t) = cy(t) \quad (2.18)$$

donde  $c$  es el factor de escala. De acuerdo con esta ecuación el valor  $y'(t)$  se obtiene multiplicando el valor correspondiente de  $y(t)$  por el escalar  $c$ .

*Suma.* Se considera que  $y_1(t)$  y  $y_2(t)$  son un par de series de tiempo. La serie  $y'(t)$  obtenida por la suma de  $y_1(t)$  y  $y_2(t)$  está definida por

$$y'(t) = y_1(t) + y_2(t). \quad (2.19)$$

*Multiplicación.* Sean  $y_1(t)$  y  $y_2(t)$  un par de funciones de series de tiempo. La señal  $y'(t)$  resultante de la multiplicación de  $y_1(t)$  y  $y_2(t)$  está definida por

$$y'(t) = y_1(t)y_2(t). \quad (2.20)$$

*Escalamiento de tiempo.* Sea  $y(t)$  una serie de tiempo. La serie  $y'(t)$  obtenida por el escalamiento de tiempo  $t$  por un factor  $a$  se define como

$$y'(t) = y(at). \quad (2.21)$$

Si  $a > 1$ , la serie  $y'(t)$  es una versión *comprimida* de  $y(t)$ . Si por otro lado  $0 < a < 1$ , la señal  $y'(t)$  es una versión *expandida* de  $y(t)$ .

*Reflexión.* Sea  $y(t)$  una serie de tiempo. Sea  $y'(t)$  la serie obtenida al sustituir el tiempo  $t$  con  $-t$ , como se muestra mediante

$$y'(t) = y(-t). \quad (2.22)$$

La señal  $y'(t)$  representa la versión reflejada de  $x(t)$  en torno al eje de la amplitud.

De aquí se derivan los siguiente dos casos

- Señales pares, para las cuales  $y(-t) = y(t)$  para todo  $t$ ; es decir, una señal par es la misma que su versión reflejada.
- Señales impares, para las cuales se tiene que  $y(-t) = -y(t)$  para todo  $t$ ; es decir, una señal impar es el negativo de su versión reflejada.

Otra operación común con series de tiempo es el *corrimiento en el tiempo* la cual se define mediante

$$y'(t) = y(t - t_0) \quad (2.23)$$

donde  $t_0$  es el corrimiento en el tiempo. Si  $t_0 > 0$ , la serie que representa  $y(t)$  se corre intacta hacia la derecha, con respecto al eje del tiempo, Si  $t_0 < 0$ , se corre hacia la izquierda.

Cuando se tiene un conjunto de datos en donde se presenta un incremento en la tendencia, es decir, un patrón estacional muy marcado en la variación con el nivel del tiempo, se debe utilizar un modelo de pronóstico que tome en cuenta estos



patrones. Con el fin de disminuir el efecto de estos patrones en la serie de tiempo se utilizan distintas transformaciones a la serie de tiempo.

Una de estas transformaciones es el aplicar la función *raíz cuadrada*. Esto se realiza calculando la raíz cuadrada a cada uno de los datos de la serie de tiempo. Esta es solo una de las distintas transformaciones que se aplican a series de tiempo, en la práctica las más comunes son la raíz cuadrada y el logaritmo, este último debido a que un cambio en el valor del logaritmo se interpreta como un cambio relativo (porcentaje) a la escala original.

Estas transformaciones son miembros de la familia de *transformaciones potencia*:

$$W_t = \begin{cases} -Y_t^p, & p < 0; \\ \log(Y_t), & p = 0; \\ Y_t^p, & p > 0. \end{cases} \quad (2.24)$$

Es preferible elegir valores sencillos de  $p$  para realizar alguna de estas transformaciones. Los modelos y pronósticos para una serie de tiempo son relativamente insensibles al valor elegido de  $p$  (valores cercanos de  $p$  producen resultados similares). Valores de  $p$  como 0, -1 o 1/2, producen resultados mucho más fáciles de interpretar que un número como  $p = 0,38463$ . En muchas ocasiones no se necesita realizar transformación alguna, esto es  $p = 1$ .

Cuando los datos han sido transformados los intervalos de predicción necesitan ser transformados a su escala original. La manera más simple de proceder es aplicando la transformación inversa a los puntos finales. Es decir, si se utilizó el logaritmo y el pronóstico en la escala de logaritmo es  $f_{t+1}$ , entonces el pronóstico en la escala original es  $e^{f_{t+1}}$ .

### 2.3. MODELOS MATEMATICOS ESTOCÁSTICOS Y DETERMINÍSTICOS

La idea de utilizar un modelo matemático para describir el comportamiento de un fenómeno físico está bien establecida. En particular, en ocasiones es posible derivar un modelo basado en leyes físicas, lo cual nos permite calcular el valor de alguna cantidad dependiente del tiempo, casi a cualquier instante de tiempo. Por consiguiente, podríamos calcular la trayectoria de un misil lanzado en una dirección y velocidad conocidas. Si es posible realizar este cálculo de manera exacta, el modelo sería puramente *determinístico*.

Probablemente ningún fenómeno es totalmente determinístico, debido a que factores desconocidos pueden ocurrir como una velocidad de viento variable que puede conducir a que el misil se desvíe levemente. En muchos problemas hay que considerar un fenómeno dependiente del tiempo, como las ventas mensuales de un periódico, en las cuales hay muchos factores desconocidos para los cuales no es posible escribir un modelo determinístico que permita el cálculo exacto del comportamiento futuro del fenómeno. No obstante, si los valores futuros pueden ser descritos solamente en términos de una distribución de probabilidad, se dice que la serie de tiempo es no-determinista o simplemente es una serie de tiempo *estadística*.

Un fenómeno estadístico que se desarrolla en el tiempo de acuerdo a leyes probabilísticas es llamado *proceso estocástico*. En ocasiones es referido simplemente como "proceso" omitiendo la palabra "estocástico". Entonces la serie de tiempo a analizar puede ser pensada como una *realización* particular producida por el mecanismo de probabilidad subyacente del sistema en estudio. En otras palabras, al analizar la serie de tiempo se toma como una realización de un proceso estocástico.

## 2.4. PROCESOS ESTOCÁSTICOS ESTACIONARIOS

Una clase de proceso estocástico muy especial, llamada *proceso estacionario*, está basado en la suposición que el proceso se encuentra en un estado particular de *equilibrio estadístico*. Se dice que un proceso estocástico es *estrictamente estacionario* si sus propiedades no son afectadas por un cambio en el origen, esto es, si la probabilidad conjunta asociada con  $m$  observaciones  $x_{t_1}, x_{t_2}, \dots, x_{t_m}$ , tomadas en cualquier conjunto de tiempos  $t_1, t_2, \dots, t_m$ , es el mismo que el asociado con las observaciones  $x_{t_1+k}, x_{t_2+k}, \dots, x_{t_m+k}$ , tomado en los tiempos  $t_1+k, t_2+k, \dots, t_m+k$ . Por lo tanto para que un proceso sea estrictamente estacionario, la distribución conjunta de cualquier conjunto de observaciones no debe ser afectado por desplazar los tiempos de observación hacia adelante o atrás en cualquier entero  $k$ .

### 2.4.1. MEDIA Y VARIANZA DE UN PROCESO ESTACIONARIO

Cuando  $m = 1$ , la suposición de estacionariedad implica que la distribución de probabilidad  $p(z_t)$  es la misma para todos los tiempos  $t$  y puede ser escrita  $p(z)$ . De aquí que el proceso estocástico posea una media constante

$$\mu = E[z_t] = \int_{-\infty}^{\infty} zp(z)dx \quad (2.25)$$

la cual define el nivel acerca del que fluctúa, y una varianza constante

$$\sigma_z^2 = E[(z_t - \mu)^2] \quad (2.26)$$

que mide la *dispersidad* cerca de este nivel de tiempo. Debido a que la distribución de probabilidad  $p(z)$  es la misma para todos los tiempos  $t$ , su forma puede ser inferida del histograma de las observaciones  $x_1, x_2, \dots, x_N$ , obtenidas de la serie de tiempo observada. Además, la media  $\mu$  del proceso estocástico puede ser estimado

por la media de la muestra (2.1) de la serie de tiempo, y la varianza  $\sigma_z^2$  del proceso estocástico puede ser estimada por la varianza de la muestra (2.4) de la serie de tiempo.

### 2.4.2. AUTOCOVARIANZA Y COEFICIENTES DE AUTOCORRELACIÓN

La suposición de estacionariedad además implica que la distribución de probabilidad conjunta  $p(z_{t_1}, z_{t_2})$  es la misma para todos los tiempos  $t_1, t_2$ , los cuales forman un intervalo constante independiente. De esto se tiene que la naturaleza de esta distribución conjunta puede ser inferida graficando un diagrama de dispersión que utilice pares de valores  $(z_t, z_{t+k})$  de la serie de tiempo separados por un retardo constante  $k$ . La covarianza entre los valores  $x_t$  y  $x_{t+k}$  separados por  $k$  intervalos de tiempo, los cuales bajo la suposición de estacionariedad deben ser los mismos para todo  $t$ , es llamada la *autocovarianza* en el retardo  $k$  y se define por

$$\gamma_k = \text{cov}[z_t, z_{t+k}] = E[(z_t - \mu)(z_{t+k} - \mu)]. \quad (2.27)$$

Similarmente la *autocorrelación* en el retardo  $k$  es

$$\begin{aligned} \rho_k &= \frac{E[(z_t - \mu)(z_{t+k} - \mu)]}{\sqrt{E[(z_t - \mu)^2]E[(z_{t+k} - \mu)^2]}} \\ &= \frac{E[(z_t - \mu)(z_{t+k} - \mu)]}{\sigma_z^2} \end{aligned} \quad (2.28)$$

donde, para un proceso estacionario, la varianza  $\sigma_z^2 = \gamma_0$  es la misma al tiempo  $t + k$  que en el tiempo  $t$ . Por lo tanto la correlación en el retardo  $k$ , esto es, la correlación entre  $z_t$  y  $z_{t+k}$ , es

$$\rho_k = \frac{\gamma_k}{\gamma_0}, \quad (2.29)$$

lo que implica que  $\rho_0 = 1$ .

### 2.4.3. MATRIZ DE AUTOCOVARIANZA Y DEFINIDA POSITIVA

La matriz de covarianza asociada con un proceso estacionario para las observaciones  $(z_1, z_2, \dots, z_n)$  tomadas en  $n$  tiempos sucesivos es

$$\begin{aligned} \Gamma_n &= \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{n-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \gamma_{n-3} & \cdots & \gamma_0 \end{bmatrix} \\ &= \sigma_z^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & 1 \end{bmatrix} & (2.30) \\ &= \sigma_z^2 P_n. & (2.31) \end{aligned}$$

Una matriz de covarianza  $\Gamma_n$  de esta forma, que es simétrica con elementos constantes en cualquier diagonal será llamada una *matriz de autocovarianza* y la correspondiente matriz de correlación  $P_n$ , será llamada una *matriz de autocorrelación*. Ahora si se considera cualquier función lineal de las variables aleatorias  $z_t, z_{t-1}, \dots, z_{t-n+1}$  se tiene:

$$L_t = l_1 z_t + l_2 z_{t-1} + \cdots + l_n z_{t-n+1}. \quad (2.32)$$

lo cual es llamado un proceso estocástico.

Dada  $cov[z_i, z_j] = \gamma_{|j-i|}$  para un proceso estacionario, la varianza de  $L_t$  es

$$\text{Var}[L_t] = \sum_{i=1}^n \sum_{j=1}^n l_i l_j \gamma_{|j-i|} \quad (2.33)$$

la cual es necesariamente mayor que cero, si las  $l$ 's no son todas cero. De aquí viene que ambas, la matriz de autocovarianza y la matriz de autocorrelación son definidas positivas para un proceso estacionario. Correspondientemente, se tiene que ambas, la función de autocovarianza  $\gamma_k$  y la función de autocorrelación  $\rho_k$ , vistas como funciones del retardo  $k$ , son funciones definidas positivas en el sentido que  $\sum_{i=1}^n \sum_{j=1}^n l_i l_j \gamma_{|j-i|} > 0$  para todo entero positivo  $n$  y todas las constantes  $l_1, \dots, l_n$ .

#### 2.4.4. CONDICIONES SATISFECHAS POR LAS AUTOCORRELACIONES DE UN PROCESO ESTACIONARIO

La definición positiva de la matriz de autocorrelación (eq. 2.30) implica que su determinante y los menores principales son mayores que cero. En particular para  $n = 2$ ,

$$\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix} > 0$$

así que  $1 - \rho_1^2 > 0$ , de esto que  $-1 < \rho_1 < 1$ .

Similarmente, para  $n = 3$  se tiene

$$\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix} > 0, \quad \begin{vmatrix} 1 & \rho_2 \\ \rho_2 & 1 \end{vmatrix} > 0$$

$$\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}$$

lo cual implica que  $-1 < \rho_1 < 1$ ,  $-1 < \rho_2 < 1$ ,  $-1 < \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} < 1$ . Debido a que  $P_n$  debe ser definida positiva para todos los valores de  $n$  las correlaciones de un proceso estacionario deben satisfacer un gran número de condiciones. Estas se mencionan en la sección 2.8.1.

#### 2.4.5. ESTACIONARIEDAD DE FUNCIONES LINEALES

De la definición de estacionariedad se tiene que el proceso  $L_t$ , obtenido realizando operaciones lineales (2.32) en un proceso estacionario  $z_t$  para  $n$  fija y coeficientes  $l_1, \dots, l_n$  fijos es también estacionario. En particular, la primera diferencia  $\nabla z_t = z_t - z_{t-1}$  y mayores  $\nabla^d z_t$ , son estacionarias.

#### 2.4.6. PROCESO GAUSSIANO

Si la probabilidad de observaciones asociada con cualquier conjunto de tiempos es una distribución Normal multivariada, el proceso es llamado Normal o proceso Gaussiano. Dado que la distribución Normal es completamente caracterizada por sus momentos de primero y segundo orden, la existencia de una media fija  $\mu$  y una matriz de autocovarianza  $\Gamma_n$  para todo  $n$  sería suficiente asegurar la estacionariedad de un proceso Gaussiano.

### 2.4.7. ESTACIONARIEDAD DÉBIL

Se ha visto que para que un proceso sea estrictamente estacionario, la estructura de probabilidad debe depender sólo de diferencias de tiempo. Un requisito menos restrictivo, llamado *estacionariedad débil* de orden  $f$  consiste en que los momentos hasta algún orden  $f$  dependen sólo de diferencias en el tiempo. Por ejemplo, la existencia de una media constante  $\mu$  y una matriz de autocovarianza  $\Gamma_n$  de la forma (2.30) es suficiente para asegurar estacionariedad hasta segundo orden. Entonces, estacionariedad de segundo orden, más una suposición de normalidad, son suficientes para producir estacionariedad débil.

## 2.5. FUNCIONES DE AUTOCOVARIANZA Y AUTOCORRELACIÓN

Se ha visto que el coeficiente de autocovarianza  $\gamma_k$ , en el retardo  $k$ , mide la covarianza entre dos valores  $z_t$  y  $z_{t+k}$  a una distancia  $k$ . La gráfica de  $\gamma_k$  contra el retardo  $k$  es llamada *función de autocovarianza*  $\gamma_k$  del proceso estocástico. Similarmente, la gráfica del coeficiente de correlación  $\rho_k$  como función del retardo  $k$  es llamada *función de autocorrelación*  $\rho_k$  del proceso. Se observa que la función de autocorrelación no tiene dimensión, esto es, es independiente de la escala de medición de la serie de tiempo. Dado que  $\gamma_k = \rho_k \sigma_z^2$ , el conocer la función de autocorrelación  $\rho_k$  y la varianza  $\sigma_z^2$  es equivalente a conocer la función de autocovarianza  $\gamma_k$ .

Dado que  $\rho_k = \rho_{-k}$ , la función de autocorrelación es necesariamente simétrica en cero, y en la práctica sólo es necesario graficar la mitad positiva de la función. En ocasiones la función de autocorrelación es llamada *correlograma*.



### 2.5.1. ESTIMACIÓN DE LAS FUNCIONES DE AUTOCOVARIANZA Y AUTOCORRELACIÓN

En la práctica se tiene una serie de tiempo finita  $z_1, z_2, \dots, z_N$  de  $N$  observaciones, de las cuales solo se pueden obtener estimados de la media  $\mu$  y las autocorrelaciones. La media  $\mu = E[z_t]$  es estimada por la media de la muestra (2.1). Se puede ver que  $E[\bar{z}] = \mu$ , de aquí que  $\bar{z}$  es un estimador insesgado de  $\mu$ . Como medida de precisión de  $\bar{z}$  como estimador de  $\mu$  se tiene que

$$\text{Var}[\bar{z}] = \frac{1}{N^2} \sum_{t=1}^N \sum_{s=1}^N \gamma_{t-s} = \frac{\gamma_0}{N} \left[ 1 + 2 \sum_{k=1}^{N-1} \left( 1 - \frac{k}{N} \right) \rho_k \right].$$

Una aproximación de “muestra grande” para esta expresión de la varianza esta dada por  $\text{Var}[\hat{z}] = (\gamma_0/N)(1 + 2 \sum_{k=1}^{\infty} \rho_k)$ , en el sentido que  $N\text{Var}[\hat{z}] \rightarrow \gamma_0(1 + 2 \sum_{k=1}^{\infty} \rho_k)$  cuando  $N \rightarrow \infty$ , suponiendo que  $\sum_{k=-\infty}^{\infty} |\rho_k| < \infty$ .

Los estadistas han sugerido numerosos estimados para la función de autocorrelación, sin embargo, el estimador más aceptado para la autocorrelación  $\rho_k$  en el retardo  $k$ -esimo es

$$r_k = \frac{c_k}{c_0} \quad (2.34)$$

donde

$$c_k = \frac{1}{N} \sum_{t=1}^{N-k} (z_t - \bar{z})(z_{t-k} - \bar{z}) \quad k = 0, 1, 2, \dots, K \quad (2.35)$$

es el estimado de la autocovarianza  $\gamma_k$ , y  $\bar{z}$  es la media de la muestra de la serie de tiempo. Los valores  $r_k$  en (2.34) pueden ser llamados *función de autocorrelación de la muestra*.

## 2.6. REPRESENTACIONES DE FOURIER PARA SERIES DE TIEMPO

En este contexto se considera la función de una serie de tiempo como la superposición ponderada de funciones senoidales complejas. Si una señal de este tipo se aplica en un sistema lineal, entonces la salida del sistema es una superposición ponderada de la respuesta del sistema a cada senoide compleja.

Un proceso físico puede ser descrito en un *dominio de tiempo* por los valores de alguna cantidad  $h$  en función del tiempo  $t$ , por ejemplo  $h(t)$  o también en el *dominio de frecuencias* en donde el proceso se describe por su amplitud  $H$  como función de frecuencia  $f$  esto es  $H(f)$  con  $-\infty < f < \infty$ . Para muchos propósitos es útil suponer que  $h(t)$  y  $H(f)$  son dos distintas *representaciones* de la *misma* función, para las cuales se puede ir y regresar entre estas dos representaciones mediante las ecuaciones de la *Transformada de Fourier*,

$$\begin{aligned} H(f) &= \int_{-\infty}^{\infty} h(t)e^{2\pi ift} dt \\ h(t) &= \int_{-\infty}^{\infty} H(f)e^{-2\pi ift} df. \end{aligned} \quad (2.36)$$

Si  $t$  se mide en segundos, entonces  $f$  en la ecuación 2.36 son ciclos por segundo o Hertz (la unidad de medida de frecuencia). Sin embargo, las ecuaciones también funcionan con otras unidades, si  $f$  es una función de la posición  $x$  (en metros),  $H$  será una función de longitud de onda inversa (ciclos por metro), etc.

De la ecuación 2.36 se tiene que la transformada de Fourier es una operación lineal. La transformada de la suma de dos funciones es igual a la suma de las transformadas. La transformada de una constante que multiplica una función es la constante multiplicando la transformada de la función.

En el dominio de tiempo, la función  $h(t)$  puede tener una o más simetrías, puede ser completamente real o completamente imaginaria, puede ser par  $h(t) = h(-t)$ , impar  $h(t) = -h(-t)$ . En el dominio de frecuencia, estas simetrías conllevan relaciones entre  $H(f)$  y  $H(-f)$ . El siguiente cuadro presenta la correspondencia entre las simetrías en ambos dominios:

Si	entonces...
$h(t)$ es real	$H(-f) = [H(f)]^*$
$h(t)$ es imaginaria	$H(-f) = -[H(f)]^*$
$h(t)$ es par	$H(-f) = H(f)$ ( $H(f)$ es par)
$h(t)$ es impar	$H(-f) = -H(f)$ ( $H(f)$ es impar)
$h(t)$ es real y par	$H(f)$ es real y par
$h(t)$ es real e impar	$H(f)$ es imaginaria e impar
$h(t)$ es imaginaria y par	$H(f)$ es imaginaria y par
$h(t)$ es imaginaria e impar	$H(f)$ es real e impar

Estas propiedades son utilizadas para incrementar la eficiencia computacional de los algoritmos para realizar la transformada de Fourier.

Algunas propiedades de la transformada de Fourier son (el símbolo " $\iff$ " se utilizará para indicar la transformada). Si  $h(t) \iff H(f)$  es una pareja de transformación, entonces las otras parejas de transformación son:

$$\begin{aligned}
 h(at) &\iff \frac{1}{|a|}H\left(\frac{f}{a}\right) && \text{"escalamiento de tiempo"} \\
 \frac{1}{|b|}h\left(\frac{t}{b}\right) &\iff H(bf) && \text{"escalamiento de frecuencia"} \\
 h(t - t_0) &\iff H(f)e^{2\pi i f t_0} && \text{"corrimiento en el tiempo"} \\
 h(t)e^{-2\pi i f_0 t} &\iff H(f - f_0) && \text{"corrimiento de frecuencia"}
 \end{aligned}$$

Con dos funciones  $h(t)$ ,  $g(t)$  y sus correspondientes transformadas de Fourier  $H(f)$ ,  $G(f)$ , se pueden formar dos combinaciones de gran interés. La *convolución* de

ambas funciones, denotada  $g * h$ , se define como

$$g * h \equiv \int_{-\infty}^{\infty} g(\tau)h(t - \tau)d\tau \quad (2.37)$$

la cual expresa el “empalme” de la función  $h$  conforme se desplaza sobre la función  $g$ .

Se observa que  $g * h$  es una función en el dominio de tiempo y que  $g * h = h * g$ . De esto se tiene que  $g * h$  es otra pareja de transformación

$$g * h \iff G(f)H(f) \quad \text{Teorema de convolución} \quad (2.38)$$

En otras palabras, la transformada de Fourier de la convolución es solo el producto de las transformadas de Fourier individuales.

La *correlación* de dos funciones, denotada  $\text{Corr}(g,h)$ , esta definida por

$$\text{Corr}(g,h) \equiv \int_{-\infty}^{\infty} g(\tau + t)h(\tau)d\tau \quad (2.39)$$

La correlación es una función de  $t$ , la cual es llamada el *retardo*. Esta yace en el dominio del tiempo, y se convierte en miembro de las pareja de transformación:

$$\text{Corr}(g,h) \iff G(f)H(-f) \quad \text{Teorema de correlación} \quad (2.40)$$

## 2.7. LA TRANSFORMADA DE FOURIER PARA MUESTRAS TOMADAS DE DATOS DISCRETOS

Es común que la función  $h(t)$  sea muestreada en instantes espaciados en el tiempo. Sea  $\Delta$  el intervalo de tiempo entre muestras consecutivas, tal que la secuencia de valores muestreados es

$$h_n = h(n\Delta) \quad n = \dots, -3, -2, -1, 0, 1, 2, 3, \dots \quad (2.41)$$

El recíproco del intervalo de tiempo  $\Delta$  es llamado *tasa de muestreo*; si  $\Delta$  es medida en segundos, por ejemplo, entonces la tasa de muestreo es el número de muestras obtenidas por segundo.

Para cualquier intervalo de muestreo  $\Delta$ , hay una frecuencia especial  $f_c$ , llamada la *frecuencia crítica de Nyquist*, dada por

$$f_c = \frac{1}{2\Delta}. \quad (2.42)$$

Si una onda seno es muestreada con la frecuencia crítica de Nyquist en su valor pico positivo, entonces la siguiente muestra será a través de su valor negativo, la muestra siguiente estará en el valor positivo, y así sucesivamente. Expresado de otra manera: El muestreo crítico de una onda seno es dos puntos de muestra por ciclo. Frecuentemente se elige medir el tiempo en unidades del intervalo de muestreo  $\Delta$ . En este caso la frecuencia crítica de Nyquist es la constante  $1/2$ .

## 2.8. LA TRANSFORMADA DE FOURIER EN TIEMPO DISCRETO

Para realizar la estimación de la transformada de Fourier de una función en un número finito de sus puntos muestreados, se supone que se tienen  $N$  valores muestreados consecutivos

$$h_k \equiv h(t_k), \quad t_k \equiv k\Delta, \quad k = 0, 1, \dots, N-1 \quad (2.43)$$

tal que el intervalo de muestreo es  $\Delta$ . Con el fin de simplificar las operaciones, se supone que  $N$  es par. Si la función  $h(t)$  es no cero solo en un intervalo de tiempo finito, entonces se supone que todo el intervalo de tiempo está contenido en el rango de los  $N$  puntos obtenidos. Alternativamente, si la función  $h(t)$  es infinita, entonces se supone que los puntos muestreados son al menos "tipicos" de  $h(t)$  en el resto del tiempo.

Con  $N$  números de entrada, evidentemente sólo se producirán  $N$  números de salida. Así que, en vez de tratar de estimar la transformada de Fourier  $H(t)$  en todos los valores de  $f$  en el rango  $-f_c$  a  $f_c$ , se estiman sólo los valores discretos

$$f_n \equiv \frac{n}{N\Delta}, \quad n = -\frac{N}{2}, \dots, \frac{N}{2} \quad (2.44)$$

Los valores extremos de  $n$  en 2.44 corresponden exactamente a los límites superior e inferior del rango de frecuencia crítica de Nyquist.

El siguiente paso es aproximar la integral en la ecuación 2.36 mediante una

suma discreta:

$$H(f_n) = \int_{-\infty}^{\infty} e^{2\pi i f_n t} dt \approx \sum_{k=0}^{N-1} h_k e^{2\pi i f_n t_k} \Delta = \Delta \sum_{k=0}^{N-1} h_k e^{2\pi i k n / N}. \quad (2.45)$$

En la igualdad final se utilizaron las ecuaciones 2.43 y 2.44. La suma de la ecuación final 2.45 es llamada la *transformada discreta de Fourier* de los  $N$  puntos  $h_k$ . Es denotada por  $H_n$

$$H_n = \sum_{k=0}^{N-1} h_k e^{2\pi i k n / N}. \quad (2.46)$$

La transformada discreta de Fourier mapea  $N$  números complejos (los  $h_k$ ) en  $N$  números complejos (las  $H_n$ ). No depende de ningún parámetro dimensional, tal como la escala dimensional  $\Delta$ .

La fórmula para la transformada discreta *inversa* de Fourier, que obtiene el conjunto  $h_k$  exactamente de  $H_k$  es:

$$h_k = \frac{1}{N} \sum_{n=0}^{N-1} H_n e^{-2\pi i k n / N}. \quad (2.47)$$

Se puede observar que las únicas diferencias entre las ecuaciones 2.46 y 2.47 son: el cambio del signo de la exponencial y dividiendo el resultado en  $N$ .

### 2.8.1. PROPIEDADES ESPECTRALES DE MODELOS ESTACIONARIOS

#### EL PERIODOGRAMA DE UNA SERIE DE TIEMPO

El periodograma fue utilizado originalmente para detectar y estimar la amplitud de una componente seno, de frecuencia conocida y envuelta en ruido. Para

ilustrar el cálculo del periodograma, se supone que el número de observaciones  $N = 2q + 1$  es impar. Si se ajusta el modelo de series de Fourier

$$z_t = \alpha_0 + \sum_{i=1}^q (\alpha_i c_{it} + \beta_i s_{it}) + e_t \quad (2.48)$$

en donde  $c_{it} = \cos(2\pi f_i t)$ ,  $s_{it} = \sin(2\pi f_i t)$ , y  $f_i = i/N$  es el  $i$ -ésimo armónico de la frecuencia fundamental  $1/N$ , el estimado de mínimos cuadrados de los coeficientes  $\alpha_0$  y  $(\alpha_i, \beta_i)$  serán

$$a_0 = \bar{z} \quad (2.49)$$

$$\left. \begin{aligned} a_i &= \frac{2}{N} \sum_{t=1}^N z_t c_{it} \\ b_i &= \frac{2}{N} \sum_{t=1}^N z_t s_{it} \end{aligned} \right\} i = 1, 2, \dots, q. \quad (2.50)$$

El periodograma consiste en los  $q = (N - 1)/2$  valores

$$I(f_i) = \frac{N}{2} (a_i^2 + b_i^2) \quad i = 1, 2, \dots, q \quad (2.51)$$

en donde  $I(f_i)$  es llamado la *indensidad* en la frecuencia  $f_i$ .

Cuando  $N$  es par, se hace  $N = 2q$  y las ecuaciones 2.48, 2.49, 2.51, aplican para  $i = 1, 2, \dots, (q - 1)$  excepto

$$a_q = \frac{1}{N} \sum_{t=1}^N (-1)^t z_t, \quad (2.52)$$

$$b_q = 0, \quad (2.53)$$

$$I(f_q) = I(0,5) = N a_q^2. \quad (2.54)$$



## ESPECTRO MUESTRA

La definición 2.52 del periodograma supone que las frecuencias  $f_i = i/N$  con armónicos de la frecuencia fundamental  $1/N$ . Por medio de la introducción del espectro se relaja esta suposición y se permite a la frecuencia  $f$  variar continuamente en el rango de 0 a 0,5. La definición en la ecuación 2.52 del periodograma puede ser modificada a

$$I(f) = \frac{N}{2}(a_f^2 + b_f^2) \quad 0 \leq f \leq \frac{1}{2} \quad (2.55)$$

en donde  $I(f)$  es llamada el *espectro de muestra*.

## ESPECTRO

El periodograma y el espectro de muestra son herramientas apropiadas para analizar series de tiempo compuestas de mezclas de funciones seno y coseno en frecuencias constantes y mezcladas con ruido. Sin embargo si se supone que el espectro de muestra es calculado en una serie de tiempo de  $N$  observaciones, la cual es una realización de un proceso estacionario Normal. Un proceso de este tipo no contiene componentes determinísticas seno o coseno. Sin embargo se podría obtener, mediante un análisis de Fourier, valores de  $(a_f, b_f)$  para cualquier frecuencia dada  $f$ . Si se toman varias muestras de  $N$  observaciones de un proceso estocástico, se puede construir una población de valores para  $a_f, b_f$  y  $I(f)$ . Entonces es posible calcular el valor promedio de  $I(f)$  en realizaciones repetidas de tamaño  $N$ , esto es

$$E[I(f)] = 2 \left[ E[c_0] + 2 \sum_{k=1}^{N-1} E[c_k] \cos(2\pi f k) \right]. \quad (2.56)$$

Se puede demostrar que para  $N$  grande, el valor promedio del estimado del coeficiente de autocovarianza  $c_k$  en muestras repetidas tiende a la autocovarianza

teórica  $\gamma_k$ , esto es

$$\lim_{N \rightarrow \infty} E[c_k] = \gamma_k. \quad (2.57)$$

Al tomar el límite de 2.56 conforme  $N$  tiende a infinito, el *espectro potencia*  $p(f)$  se define como

$$p(f) = \lim_{N \rightarrow \infty} E[I(f)] = 2 \left[ \gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k \cos(2\pi f k) \right] \quad 0 \leq f \leq 1/2. \quad (2.58)$$

De aqui se tiene que

$$|p(f)| \leq 2 \left[ |\gamma_0| + 2 \sum_{k=1}^{\infty} |\gamma_k| |\cos(2\pi f k)| \right] \leq 2 \left( |\gamma_0| + 2 \sum_{k=1}^{\infty} |\gamma_k| \right), \quad (2.59)$$

una condición suficiente para la convergencia del espectro es que  $\gamma_k$  decaiga suficientemente rápido para que la serie de la ecuación 2.59 sea convergente. Debido a que el espectro potencia es la transformada de Fourier de coseno para la función de autocovarianza, el conocimiento de la función de autocovarianza es matemáticamente equivalente a conocer el espectro, y viceversa.

## 2.9. MODELOS NO LINEALES PARA SERIES DE TIEMPO

Los métodos lineales interpretan completamente la estructura regular en un conjunto de datos, tal como una frecuencia dominante a través de correlaciones lineales. Esto significa, que la dinámica intrínseca del sistema está regida por el paradigma lineal que establece que pequeñas causas provocan pequeños efectos. Debido a que las ecuaciones lineales sólo pueden tratar decaimiento (o crecimiento) exponencial o soluciones oscilantes periódicas, el comportamiento irregular del sistema debe ser atribuido a alguna entrada aleatoria en el sistema.

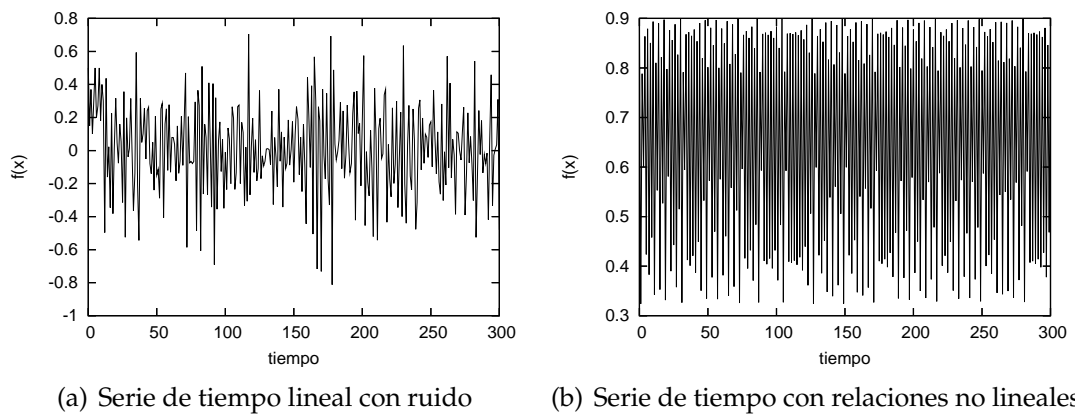


Figura 2.5: Ejemplos de series de tiempo lineales con ruido y no lineales

Como se puede apreciar en la figura 2.5, en ocasiones es difícil realizar una distinción entre una serie de tiempo lineal con ruido y una serie de tiempo con relaciones no lineales, debido a esto se han desarrollado distintas herramientas para tratar de descifrar esta no linealidad en las series de tiempo. Algunas de estas se resumen la siguiente.

### 2.9.1. RESÚMENES GRÁFICOS

Una manera práctica de observar algún patrón es graficando cada dato contra su predecesor, como se muestra en la figura 2.6. Esta gráfica es llamada *retrato de fases*. Esta representación es una aplicación particular de *retardo de tiempo anidado*, la cual es una herramienta básica utilizada frecuentemente en análisis no lineal de series de tiempo.

## 2.10. ESTACIONARIEDAD, PRUEBAS PARA DETERMINARLA

En general, una medida científica de cualquier tipo es en principio más útil conforme más es reproducible. Se necesita saber que los números medidos corresponden a las propiedades de un objeto estudiado, hasta alguna medida de error. En el caso de medidas de series de tiempo, la capacidad de reproducir los datos está conectada a dos diferentes nociones de estacionariedad.

La más débil pero mas evidente forma de *estacionariedad* requiere que todos los parámetros que son relevantes para la dinámica del sistema son constantes durante el periodo de medición. En algunos casos se puede manejar un cambio sencillo en un parámetro una vez que se observa la ocurrencia de tal cambio. Si la calibración del aparato de medición se modifica, se puede intentar reescalar los datos continuamente con el propósito de mantener la media y varianza constante.

Desafortunadamente en muchos casos, no es posible tener acceso al sistema que produce la señal, por lo que no es posible establecer evidencia de que los parámetros son constantes. Debido a esto, es necesario formular un segundo concepto de *estacionariedad*, el cual está basado en los datos obtenidos. Una manera formal de definirla es, que una señal es llamada estacionaria si todas probabilidades conjuntas de encontrar el sistema en un estado y en algún tiempo despues en otro estado son independientes del tiempo en el periodo de observación. Esto implica que el fenómeno asociado con la dinámica del sistema ocurre con suficiente frecuencia, tal que las probabilidades o demás reglas puedan ser inferidas adecuadamente.

Las señales no estacionarias son muy comunes en particular al observar fenómenos naturales o culturales. Tales señales pueden ser objeto interesante de estudio e incluso se puede hacer dinero prediciendo precios de mercado en situaciones que

no tienen paralelo en la historia.

Un problema encontrado con estas series de tiempo es el como determinar si son o no estacionarias. Este problema se torna aún más complicado debido a que los requerimientos de estacionariedad varían de acuerdo a la aplicación. Sin embargo existen distintos requerimientos que una serie de tiempo estacionaria satisface.

El primer requerimiento es que la serie de tiempo debe cubrir un intervalo de tiempo que es mucho mayor que la escala de tiempo de la característica más duradera y que es relevante para la evolución del sistema. Por ejemplo, la concentración de azúcar en la sangre de un humano está regida por el consumo de alimentos y este forma un ciclo aproximado de 24 horas. Si esta cantidad es medida por 24 horas o menos, el proceso puede ser considerado no estacionario, sin importar cuantos puntos sean tomados durante este tiempo.

Las violaciones de los requerimientos básicos de que las propiedades del sistema asociadas no deben cambiar durante el periodo de observación se pueden checar simplemente verificando tales propiedades en diversos segmentos del conjunto de datos.

## 2.11. METODOS DEL ESPACIO DE FASES

### 2.11.1. DETERMINISMO, UNICIDAD EN EL ESPACIO DE FASES

Si se considera un sistema deterministico completamente, una vez que se tiene definido el estado actual será posible determinar los estados en cualquier instante del futuro. Debido a esto es importante establecer un espacio vectorial (llamado *espacio de estados* o *espacio de fase* ) para el sistema, tal que, especificando un pun-

to en este espacio se especifique el estado del sistema y viceversa. Por consiguiente se puede estudiar la dinámica del sistema estudiando la dinámica definida por los puntos correspondientes del espacio de fase. En teoría, los sistemas dinámicos usualmente están definidos por un conjunto de ecuaciones diferenciales de primer orden actuando en el espacio de fase.

El concepto de *estado del sistema* es muy útil aún en sistemas no deterministas. Una gran clase de sistemas puede ser descrito por un conjunto de estados (posiblemente infinito) y algún tipo de reglas de transición que especifican como el sistema puede evolucionar de un estado a otro. Si se restringe al caso en que el espacio de fase es un vector de dimensión finita  $\mathbb{R}^m$ . Un estado está especificado por un vector  $x \in \mathbb{R}^m$ . Se puede describir la dinámica del sistema por medio de un mapa de dimensión  $m$  o por un sistema explícito de  $m$  ecuaciones diferenciales de primer orden. En el primer caso, el tiempo es una variable discreta:

$$x_{n+1} = F(x_n), \quad n \in \mathbb{Z}. \quad (2.60)$$

En el segundo caso, es una expresión continua:

$$\frac{d}{dt}x(t) = f(x(t)), \quad t \in \mathbb{R}. \quad (2.61)$$

La segunda situación es llamada *flujo*. El vector  $f$  en la ecuación 2.61 está definida de tal manera que no depende explícitamente del tiempo, debido a esto es llamada *autónoma*.

## ATRACTOR

Un concepto importante en series de tiempo no lineales es el llamado *atractor*. Existen muchas definiciones para este concepto, en dependencia del autor que lo maneja, sin embargo una definición ampliamente aceptada y que es utilizada en este trabajo consiste en: un *atractor* es el conjunto de puntos a los cuales una gran cantidad de puntos evoluciona conforme el sistema itera, esto es, una región en el espacio de búsqueda a la cual el sistema evoluciona con mayor probabilidad.

### 2.11.2. RECONSTRUCCIÓN DE RETARDOS

Un problema encontrado al realizar un experimento es que lo que se observa no es un objeto en el espacio de fase, sino una serie de tiempo, es decir una secuencia de mediciones escalares. Sin embargo, es necesario convertir las observaciones en vectores de estado. Este es el problema importante de la *reconstrucción de estado de fase* el cual es resuelto básicamente mediante el método de retardos.

Comúnmente una serie de tiempo es una secuencia de medidas escalares de alguna cantidad que depende del estado actual del sistema, tomadas en múltiplos de un tiempo de muestreo fijo:

$$s_n = s(x(n\Delta t)) + n_0 \quad (2.62)$$

Esto es, se observa el sistema a través de alguna función de medida  $s$  y se realizan observaciones solo hasta algunas fluctuaciones aleatorias  $\eta_n$ , el *ruido de medición*.

Una *reconstrucción de retardos* en  $m$  dimensiones está formada por los vectores

$s_n$ , dados como

$$\mathbf{s}_n = (s_{n-(m-1)\tau}, s_{n-(m-2)\tau}, \dots, s_{n-\tau}, s_n). \quad (2.63)$$

La diferencia de tiempo en el número de muestras  $\tau$  (o en unidades de tiempo  $\tau\Delta t$ ) entre componentes adyacentes del retardo es llamado el *retardo* o *tiempo de retardo*. Se puede observar que para  $\tau > 1$ , sólo la ventana de tiempo cubierta por cada vector es incrementada, mientras el número de vectores construidos de la serie de tiempo escalar permanece a grandes rasgos igual. Esto ocurre debido a que se crea un vector por cada observación escalar,  $s_n$ , con  $n > (m-1)\tau$ . Bajo determinadas circunstancias el atractor formado por  $\mathbf{s}_n$  es equivalente al atractor de el espacio desconocido en el cual habita el sistema original si la dimensión  $m$  de la coordenada de retardo es suficientemente grande.

Al iniciar el análisis de una serie de tiempo no se conocen características de la función que la produce, como son la dimensión de conteo de segmentos, la cual es a grandes rasgos, el número de coordenadas necesarias para cubrir el subconjunto invariante en el cual habita la dinámica. El cual es necesario para calcular  $m$ , tampoco se conoce el retardo de tiempo  $\tau$ . Debido a esto se han desarrollado distintos algoritmos para estimar estos componentes desconocidos pero indispensables para determinar la función que genera la serie de tiempo.

### 2.11.3. OBTENCIÓN DE DIMENSIONES DE RECURRENCIA: VECINOS FALSOS

Si se asume que la dinámica en el espacio de fase es representado por un campo vectorial suave, entonces los estados vecinos deben estar sujetos a casi la misma evolución en el tiempo. De aquí que después de un intervalo de tiempo pequeño en el futuro, ambas trayectorias que emergen de ellos deben ser vecinos cercanos.



La idea básica del concepto *vecinos falsos* consiste en buscar puntos en un conjunto de datos que sean vecinos en el espacio incrustado, pero que no deben ser vecinos debido a que la evolución en el tiempo es muy distinta. Se supone que la dimensión anidada correcta para algún conjunto de datos es  $m_0$ , en donde la dimensión anidada es la dimensión de una estructura que se encuentra dentro de la estructura principal de la serie de tiempo. Ahora se estudian los mismos datos en una dimensión de anidamiento menor  $m < m_0$ . La transición de  $m_0$  a  $m$  es una proyección, eliminando algunos ejes del sistema de coordenadas. De aquí, los puntos cuyas coordenadas que son eliminadas por la proyección difieren fuertemente pueden ser considerados *vecinos falsos* en el espacio de dimensión  $m$ . La estadística a estudiar es entonces, para cada punto de la serie de tiempo, tomar su vecino más cercano en  $m$  dimensiones y calcular el cociente de la distancia entre estos dos puntos en  $m + 1$  dimensiones y en  $m$  dimensiones. Si este cociente es mayor que un umbral  $r$ , el vecino es falso. Este umbral debe ser suficientemente grande para permitir divergencia exponencial debido a caos determinístico.

Si se denota la desviación estándar de los datos por  $\sigma$  y se utiliza la norma máxima, la estadística a calcular es

$$X_{\text{fnn}} = \frac{\sum_{n=1}^{N-m-1} \Theta \left( \frac{|\mathbf{s}_n^{(m+1)} - \mathbf{s}_{k(n)}^{(m+1)}|}{|\mathbf{s}_n^{(m)} - \mathbf{s}_{k(n)}^{(m)}|} - r \right) \Theta \left( \frac{\sigma}{r} - |\mathbf{s}_n^{(m)} - \mathbf{s}_{k(n)}^{(m)}| \right)}{\sum_{n=1}^{N-m-1} \Theta \left( \frac{\sigma}{r} - |\mathbf{s}_n^{(m)} - \mathbf{s}_{k(n)}^{(m)}| \right)} \quad (2.64)$$

en donde  $\mathbf{s}_{k(n)}$  es el vecino más cercano a  $\mathbf{s}$  en  $m$  dimensiones, esto es,  $k(n)$  es el índice en la serie de tiempo del elemento dado por  $\arg \min_{k \neq n} |\mathbf{s}_n - \mathbf{s}_k|$ . El primer elemento en el numerador es la unidad si el vecino más cercano es falso, esto es, si la distancia se incrementa en un factor más que  $r$  cuando se incrementa la dimensión anidada en una unidad, mientras la segunda parte elimina los pares cuya distancia inicial es mayor que  $\sigma/r$ .

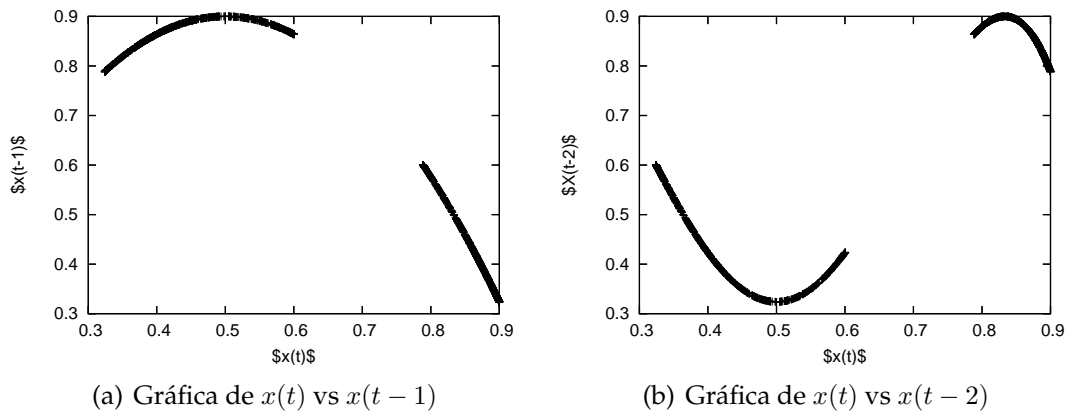


Figura 2.6: Distintas representaciones para una serie de tiempo

## 2.12. INSPECCIÓN VISUAL DE DATOS

A pesar de que existe una gran cantidad de herramientas para la caracterización y análisis de series de tiempo, se debe tener en cuenta que se deben observar de diversas maneras. Una gráfica de la señal en función del tiempo proporciona los primeros indicios de posibles problemas de estacionalidad, como desfases, amplitudes que varían sistemáticamente o escalas de tiempo y eventos atípicos. Esto permite seleccionar la parte de la serie de tiempo que parece ser más estacionaria. Es frecuente encontrar errores en los datos visuales mediante inspección visual.

Como siguiente paso se puede realizar una representación en dos dimensiones como las mostradas en la figura 2.6. Aún cuando no siempre es posible obtener una estructura visible se obtiene una idea de que retardo se podría utilizar para un anidamiento del conjunto de datos. Un análisis visual puede revelar simetrías en los datos o puede ayudar a seleccionar una mejor representación de los datos. En este caso se puede apreciar como al incrementar el retardo con el que se gráfica se van formando más ondas en la curva, lo cual da una idea que un retardo de tamaño 1 es suficiente para describir el sistema.

Figura 2.7: La gráfica de recurrencia permite visualizar los valores redundantes en la serie de tiempo

## 2.13. GRÁFICAS DE RECURRENCIA

Una de las aplicaciones de los análisis de series de tiempo consiste en extraer información significativa de los datos. Esto recae en encontrar redundancias en esta información. Si los datos no son periódicos y no se puede definir alguna regla simple de sus dependencias en el tiempo, repeticiones aproximadas de algunos eventos pueden ayudar en la construcción de reglas más complicadas. Al suponer determinismo y que un espacio de retardos anidado domina el espacio de estado del sistema, una repetición aproximada es llamada *recurrencia*, esto es, el regreso de la trayectoria del espacio de estado al vecindario de un punto en donde ha estado antes. Tales recurrencias existen en el espacio original para todo tipo de movimientos que no son transitorios.

Un método simple para visualizar recurrencias es llamado *gráfica de recurrencia*. Se calcula la matriz

$$M_{ij} = \Theta(\epsilon - |s_i - s_j|) \quad (2.65)$$

en donde  $\Theta(\cdot)$  es la función escalón de Heaviside,  $\epsilon$  es un parámetro de tolerancia, y  $s_i$  son vectores de retardo de alguna dimensión anidada. La matriz es simétrica por construcción. Si la trayectoria en el espacio reconstruido regresa al tiempo  $i$  en el  $\epsilon$ -vecindario de donde estuvo al tiempo  $j$ , entonces  $M_{ij} = 1$ , de otra manera  $M_{ij} = 0$ .  $M_{ij}$  puede ser dibujada como puntos blancos y negros en el plano de índices para inspección visual, esta es la gráfica de recurrencia 2.7.

Dos funciones importantes en este contexto son la función delta, o función impulso, la cual se produce cuando ocurre una señal por un tiempo infinitesimal,

la función que representa esto, contiene una unidad de área, y se escribe:

$$\delta(x) = \begin{cases} \infty & x = 0 \\ 0 & x \neq 0 \end{cases} \quad (2.66)$$

y que está restringida a satisfacer la densidad

$$\int_{-\infty}^{\infty} \delta(x) = 1. \quad (2.67)$$

Debido a que esto es imposible de obtener en la vida real, frecuentemente se define un valor  $\varepsilon$  muy pequeño utilizado de la siguiente manera:

$$\delta(x) = \begin{cases} 0 & x < -\varepsilon/2 \\ 0 & x > \varepsilon/2 \\ 1/\varepsilon & -\varepsilon/2 \leq x \leq \varepsilon/2 \end{cases} \quad (2.68)$$

y que esta restringida a satisfacer la densidad

$$\int_{-\infty}^{\infty} \delta(x) = 1. \quad (2.69)$$

Asociada a la función delta de Dirac, se tiene la función de Heaviside, la cual es la integral de la función delta de Dirac, esta se define de la siguiente manera:

$$H(x) = \int_{-\infty}^{\infty} \delta(t) dt. \quad (2.70)$$

Esta función es llamada función escalon unitario, ya que toma los siguientes valores:

$$H(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}. \quad (2.71)$$

## 2.14. REDES NEURONALES

El trabajo conocido como *redes neuronales* ha sido motivado por el reconocimiento de como el cerebro humano procesa la información en una manera completamente distinta a como lo realiza una computadora digital común. El cerebro es una computadora altamente no lineal, compleja y paralela. Posee la capacidad de organizar sus componentes estructurales conocidos como neuronas, con el propósito de realizar determinados calculos muchas veces mas rápido que la computadora digital más poderosa de la actualidad.

En su forma más general una *red neuronal* es una máquina que esta diseñada para modelar la manera en que el cerebro realiza una tarea en particular; la red es implementada comúnmente utilizando componentes electrónicos o simulada mediante software en una computadora digital. Para alcanzar buenos resultados, las redes neuronales utilizan una interconexión masiva de células de cómputo llamadas neuronas o unidades de procesamiento. Una definición de red neuronal como máquina adaptativa sería la siguiente:

Una red neuronal es un procesador masivamente distribuido compuesto de unidades de procesamiento simple, las cuales por naturaleza son propensas a almacenar conocimiento experimental y ponerlo disponible para su uso. Esto imita al cerebro en los siguientes aspectos:

1. El conocimiento es adquirido por la red a través de un proceso de aprendizaje del ambiente.
2. Se utilizan fuerzas de interconexiones llamados pesos sinápticos, para almacenar el conocimiento adquirido.

El procedimiento utilizado para realizar el procedimiento de aprendizaje es

conocido como *algoritmo de aprendizaje*, su función es modificar los pesos sinápticos en la red de una manera ordenada para obtener el objetivo de diseño deseado.

El uso de redes neuronales ofrece las siguientes propiedades y capacidades útiles:

- No linealidad. Una red neuronal artificial puede ser lineal o no lineal. La red neuronal compuesta de interconexiones de neuronas no lineales es en si, no lineal.
- Mapeo de entrada - salida. Un método de aprendizaje popular llamado aprendizaje con maestro o aprendizaje supervisado implica la modificación de los pesos sinápticos aplicando un conjunto de muestras de entrenamiento. Cada muestra consta de una señal de entrada y una respuesta deseada únicas.
- Adaptatividad. Las redes neuronales poseen la capacidad intrínseca de adaptar los pesos sinápticos a los cambios en el medio ambiente que la rodea.
- Evidencia de respuesta. En el contexto de clasificación de patrones, una red neuronal puede ser diseñada para proveer información no solo de como selecciona un patrón, sino además de la confianza en la decisión tomada.
- Información contextual. Debido a la estructura de la red neuronal el conocimiento es representado por el estado de activación de las neuronas. Cada neurona es afectada por la actividad global de las demás neuronas en la red.
- Tolerancia de errores. Una red neuronal implementada en forma de hardware tiene potencial de ser tolerante a errores por su naturaleza, o capaz de computación robusta; en el sentido que su desempeño se degrada lentamente con forme se presentan condiciones de operación adversas.
- Implementaciones a gran escala. La naturaleza de cómputo masivo en una red neuronal la propicia para realizar ciertas tareas con gran rapidez.

- Uniformidad de análisis y diseño. Una red neuronal es universal a su aplicación, es decir, los componentes de la red neuronal son básicamente los mismos sin importar el dominio de la aplicación.

La manera en la cual las neuronas de una red neuronal están estructuradas se encuentra íntimamente relacionada con el algoritmo de aprendizaje utilizado para entrenarla.

En general se pueden identificar tres diferentes tipos fundamentales de clases de arquitecturas de red:

#### REDES DE CAPA SIMPLE CON FLUJO HACIA ADELANTE

En una red neuronal con capas, las neuronas están organizadas en forma de capas. En la manera más simple de una red por capas, se tiene una *capa de entrada* de nodos origen que se proyectan en una *capa de salida* de neuronas, pero no en la dirección opuesta. En otras palabras el flujo es de tipo hacia adelante o acíclica.

#### REDES MULTICAPA CON FLUJO HACIA ADELANTE

La segunda clase de redes neuronales con flujo hacia adelante se distingue por la presencia de una o más *capas ocultas*, cuyos nodos de cómputo se conocen como *neuronas ocultas* o *unidades ocultas*. La función de las neuronas ocultas es intervenir entre la entrada de datos y la salida de la red en alguna manera útil. Al agregar una o más capas ocultas es posible extraer estadísticas de orden más alto.

## REDES RECURRENTE

Una *red neuronal recurrente* se distingue ya que tiene al menos un ciclo de flujo de datos, es decir, la salida de la red neuronal es utilizada como una entrada de la misma.

## 2.14.1. EL PROBLEMA DE FILTRADO ADAPTATIVO

Considere un sistema dinámico con una caracterización matemática desconocida. Toda la información que se tiene disponible es un conjunto de datos de entrada-salida generados por el sistema en instantes de tiempo discretos con alguna tasa uniforme. Específicamente cuando un estímulo  $m$ -dimensional  $x(i)$  es aplicado a través de  $m$  nodos de entrada del sistema, este responde produciendo una salida escalar  $d(i)$ , en donde  $i = 1, 2, \dots, n, \dots$ . Entonces el comportamiento externo del sistema está descrito por el conjunto de datos  $T : x(i), d(i); i = 1, 2, \dots, n, \dots$  en donde  $x(i) = [x_1(i), x_2(i), \dots, x_m(i)]^T$

Las muestras  $T$  son idénticamente distribuidas de acuerdo a una ley de probabilidad desconocida. La dimensión  $m$  perteneciente al vector de entrada  $x(i)$  es conocida como *dimensionalidad del espacio de entrada* o simplemente *dimensionalidad*.

El estímulo  $x(i)$  puede provenir básicamente de dos distintas maneras, uno espacial y el otro temporal:

- Los  $m$  elementos de  $x(i)$  se originan en distintos puntos en el espacio; en este caso  $x(i)$  es una fotografía de los datos.
- Los  $m$  elementos de  $x(i)$  representan el conjunto presente y  $m - 1$  valores pasados de alguna excitación que están espaciados uniformemente en el tiempo.



El problema encontrado es como diseñar un modelo de múltiples entradas - salida simple para el sistema desconocido. El modelo neuronal opera bajo la influencia de un algoritmo que controla los ajustes necesarios a los pesos sinápticos de las neuronas, con los siguientes puntos en mente:

- El algoritmo inicia de una *configuración arbitraria* de pesos sinápticos.
- Los ajustes a los pesos sinápticos en respuesta a variaciones estadísticas en el comportamiento del sistema se realizan en una base continua.
- Los calculos de los ajustes a los pesos sinápticos se completan en un intervalo de tiempo que dura un periodo de muestreo.

En el caso de trabajar con una serie de tiempo, se tiene que las entradas de la red neuronal están asociadas con observaciones obtenidas en el pasado de la misma manera que el problema del filtrado, sin embargo las salidas de la red neuronal son también observaciones del pasado. Para realizar un pronóstico una red neuronal es entrenada utilizando solo observaciones del pasado. Una vez que la red neuronal ha logrado un buen desempeño al pronosticar los valores conocidos, es probada con valores del futuro.

Existen distintas maneras en que una red neuronal puede ser utilizada para pronóstico:

- Utilizar  $m$  neuronas de entrada y producir una salida solamente. Esto es pronosticar a un horizonte de 1.
- Utilizar  $m$  neuronas de entrada y producir  $n$  salidas, en donde cada salida corresponde a el pronóstico a un horizonte  $i = 1, 2, \dots, n$ .

De esta manera se tiene que debido a la naturaleza no lineal de una red neuronal y capacidad para caracterizar sistemas dinámicos desconocidos, es propicia para realizar un pronóstico en series de tiempo, sobre todo cuando se tiene la sospecha de existencia de relaciones no lineales entre los datos de entrada - salida del sistema tratado.

## MODELO

---

### 3.1. MODELOS LINEALES PARA SERIES DE TIEMPO

Cuando se tiene el caso de que una serie de tiempo presenta sólo relaciones lineales en sus datos, se utilizan modelos lineales. Una de las ventajas que presenta el utilizar modelos lineales es que sólo requiere realizar operaciones lineales con los datos de entrada; lo cual es sinónimo de un costo computacional menor, en comparación con el utilizado para evaluar una función no lineal. Esto permite obtener respuesta en tiempos cortos.

Uno de los problemas encontrados con cualquier modelo de pronóstico es la obtención de los parámetros adecuados, ya que algunos necesitan la estimación de una gran cantidad de éstos. En este trabajo se compara el algoritmo propuesto contra el modelo AR, el cual es un ejemplo clásico de modelo de pronóstico.

### 3.2. EL MODELO AR

Un modelo estocástico que puede ser muy útil en la representación de ciertas series de tiempo que ocurren en la práctica es el modelo *autoregresivo*. En este mode-

lo, el valor actual del proceso se expresa como un agregado lineal y finito de valores anteriores en el proceso y un ruido blanco  $\varepsilon_t$ .

Se denota los valores de un proceso a tiempos igualmente espaciados  $t, t - 1, t - 2, \dots$  por  $z_t, z_{t-1}, z_{t-2}, \dots$ . Además sean  $\tilde{z}_t, \tilde{z}_{t-1}, \tilde{z}_{t-2}, \dots$  las desviaciones de  $\bar{X}$ , por ejemplo,  $\tilde{z} = z_t - \bar{X}$ . Entonces

$$\hat{z}_t = a_1 \tilde{z}_{t-1} + a_2 \tilde{z}_{t-2} + \dots + a_p \tilde{z}_{t-p} + \varepsilon_t \quad (3.1)$$

es llamado *proceso autoregresivo de orden p*. La razón de este nombre es que un modelo lineal

$$\hat{z}_t = a_1 \tilde{x}_1 + a_2 \tilde{x}_2 + \dots + a_p \tilde{x}_p + \varepsilon_t \quad (3.2)$$

relaciona una variable “dependiente”  $z$  con un conjunto de variables “independientes”  $x_1, x_2, \dots, x_p$  más un término de error  $\varepsilon_t$ , es llamado un modelo de *regresión*, y se dice que  $z$  es “regresada” en  $x_1, x_2, \dots, x_p$ . En la ecuación 3.1 la variable  $z$  es regresada en los valores previos de sí misma; de aquí que el modelo reciba el nombre de *autoregresivo*.

El modelo contiene  $p + 2$  parámetros desconocidos  $\bar{X}, a_1, a_2, \dots, a_p, \sigma_a^2$  los cuales en la practica tienen que ser estimados de los datos. El parámetro adicional  $\sigma_a^2$  es la varianza del proceso de ruido blanco  $\varepsilon_t$ .

Si se define un *operador autoregresivo* de orden  $p$

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p, \quad (3.3)$$

el modelo autoregresivo puede ser escrito en una manera compacta

$$\phi(B)\tilde{z}_t = a_t, \quad (3.4)$$

lo cual es equivalente a

$$\tilde{z}_t = \psi(B)a_t, \quad (3.5)$$

en donde

$$\psi(B) = \phi^{-1}(B). \quad (3.6)$$

El proceso autoregresivo puede ser o no estacionario. Para que el proceso sea estacionario, las  $\phi$  deben ser elegidas de manera que los pesos  $\psi_1, \psi_2, \dots$  en  $\psi(B) = \phi^{-1}(B)$  formen una serie convergente. El requerimiento necesario para estacionariedad es que el operador autoregresivo  $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$  considerado como un polinomio de grado  $p$  en  $B$  debe tener todas las raíces de  $\phi(B) = 0$  mayor o igual que 1 en valor absoluto; esto es, todas las raíces deben encontrarse fuera del círculo unitario.

### 3.2.1. RAZÓN DE LA UTILIZACIÓN DEL MODELO AR

Otro modelo clásico para pronóstico en series de tiempo es el modelo ARMA, el cual incorpora una ponderación del error de pronóstico obtenido. La forma general de este modelo es

$$\hat{z}_t = \varepsilon_t + a_1 \tilde{z}_{t-1} + a_2 \tilde{z}_{t-2} + \dots + a_p \tilde{z}_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}, \quad (3.7)$$

en el cual se agregan los terminos  $\theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$  en donde  $\theta_i$  son los parámetros del modelo y  $\varepsilon_{t-i}$  es el error de pronóstico en el horizonte  $i$ . Estos modelos son llamados aurotegresivos con media móvil (ARMA por sus siglas en inglés), su notación es ARMA( $p, q$ ), en donde  $p$  es el orden del modelo autoregresivo y  $q$  es el orden del modelo de medias móviles.

Un problema encontrado en este tipo de modelos es que solamente mejoran el

pronóstico solamente hasta un horizonte  $q$ , el cual es el orden de la sección de media móvil  $MA(q)$ . Debido a que en este trabajo el modelo KNNR es capaz de trabajar con horizontes mucho mayores que el orden  $q$  del modelo  $ARMA(p,q)$ . Se utiliza solamente el modelo  $AR(p)$  como una generalización del modelo  $ARMA(p,q)$ .

### 3.3. EL MODELO KNNR

La identificación de procesos determinísticos en series de tiempo es muy importante para obtención de características básicas y relaciones que se encuentran en los datos. En muchos casos la separación de componentes estocásticas y determinísticas se realiza directamente del espectro de frecuencia de la señal. Sin embargo en muchos casos importantes, la identificación de comportamiento determinístico basado solamente en el espectro de Fourier es imposible de realizar debido a la presencia de ruido intenso, una dinámica caótica o datos insuficientes.

El enfoque presentado en este trabajo involucra solamente operaciones lineales, lo cual es computacionalmente poco costoso en comparación con la mayoría de las técnicas.

El interés principal de este trabajo es el tratar series de tiempo producidas mediante dinámicas deterministas con presencia de ruido aditivo intenso. Para estos casos el modelo estudiado es capaz de obtener resultados comparables y en algunos casos mejor que las técnicas de análisis lineal para series de tiempo.

### 3.3.1. REDES DE CONOCIMIENTO

El método está basado en la noción de una red de conocimiento, un concepto de reciente introducción en el campo de mecánica estadística, utilizado para responder preguntas relacionadas con algunas redes de interacción que provienen de campos como biología y sociología. En una red de conocimiento los nodos son caracterizados por un número interno de grados de libertad, mientras los enlaces transportan el producto escalar de los dos nodos que conectan. Una pregunta muy importante para las aplicaciones de redes de conocimiento ocurre cuando solo es posible observar directamente el producto escalar.

En este caso el problema consiste en extraer alguna información de una base de datos con escasa información de productos escalares. En el contexto de series de tiempo, las cantidades observables son valores de una serie de tiempo estacionaria univariada. Los nodos son las amplitudes y la frecuencia son componentes de la parte determinística de la serie de tiempo, la cual es en general desconocida por anticipación. De hecho, se tiene que con la estimación del número oculto de componentes de Fourier, es posible reconstruir características importantes de los datos.

El método KNNR recae en la noción de una red de conocimiento [1, 16]. Las redes de conocimiento son motivadas originalmente del estudio de algunas estructuras particulares que surgen en el campo de la economía y biología, como puede ser la interacción entre clientes y productos en un mercado o proteínas que presentan interacciones de sustratos [15, 16].

Una red de conocimiento se define como una red en donde los nodos son caracterizados por  $L$  grados de libertad internos, mientras las conexiones corresponden al producto vectorial de los nodos que conectan [16]. Con el propósito de ordenar las ideas se puede considerar el siguiente modelo de red de conocimiento de información de opiniones [1, 16]: se supone que existe una base de datos de opiniones dadas

por agentes en un conjunto dado de productos. Esta base de datos puede ser una matriz dispersa, en donde los datos faltantes corresponden a opiniones faltantes, es decir, agentes que no han expresado su opinión para determinado producto.

En palabras geométricas, las preferencias de un agente son representadas como un vector en un espacio de gustos hipotético, cuya dimensión y vectores base son generalmente desconocidos. Un producto es representado por un vector similar de características, las cuales pueden ser expresadas por el producto escalar entre vectores correspondientes. Se asume que la opinión de un agente respecto a cierto producto es proporcional al cruce entre preferencias y características, las cuales pueden ser expresadas por el producto escalar entre vectores correspondientes. De esta manera los productos actúan como una base, y las opiniones como coordenadas de los agentes en esta base.

Se considera una población de  $M$  agentes interactuando con  $N$  productos. Estos dos conjuntos de vectores yacen en un espacio  $L$ -dimensional,  $\mathbf{a}_n = (a^1, a^2, \dots, a^L)$  y  $\mathbf{b}_m = (b^1, b^2, \dots, b^L)$ , donde  $n = 1, 2, \dots, N$  y  $m = 1, 2, \dots, M$ . En este sentido el cruce de  $y_{m,n} = \mathbf{b}_m \cdot \mathbf{a}_n$  representa la opinión del agente  $\mathbf{b}_m$  en el producto  $\mathbf{a}_n$ . Solo los cruces  $y_{m,n} = \mathbf{b}_m \cdot \mathbf{a}_n$  pueden ser directamente observables. El objetivo es entonces, reconstruir las cantidades ocultas a partir de una fracción de los productos escalares.

Para el caso en el cual  $L$  es conocido, Maslov y Zhang [16] han mostrado la existencia de umbrales para la fracción de  $p$  cruces conocidos, sobre el cual es posible reconstruir en distintos intervalos la información perdida. Bagnoli, Berrones y Franci [1], han generalizado el estudio de Maslov y Zhang a el caso en el cual la dimensionalidad  $L$  es desconocida. El modelo KNNR se basa principalmente en este enfoque.

Se supone que los componentes de  $\mathbf{b}_m$  y  $\mathbf{a}_n$  son variables aleatorias distribui-



das de acuerdo a

$$P(a_n^l, b_m^l) = P_{n,l}(a)P_{m,l}(b) \quad (3.8)$$

y se define  $\langle h \rangle$  como el promedio calculado en el estado estacionario, sobre  $P(a_n^l, b_m^l)$  de una función arbitraria  $h(a_n^l, b_m^l)$ . Para un conjunto de componentes ocultos distribuidos de acuerdo a la ecuación 3.8, las  $y$  son no correlacionadas. Sin embargo las correlaciones se incrementan debido a que  $L$  es finito.

Con el propósito de conservar simplicidad se asume que  $\langle a_n^l \rangle = \langle b_n^l \rangle = 0$ . Promediando sobre la distribución de la ecuación 3.8 la varianza de los cruces se escribe:

$$\langle y^2 \rangle = L \langle a^2 \rangle \langle b^2 \rangle \quad (3.9)$$

Para esta configuración de modelo Bagnoli, Berrones y Franci [1] demostraron que cualquier cruce puede ser expresado en términos de un promedio ponderado de los demás cruces,

$$y_{m,n} = \frac{L}{M-1} \sum_{i=1}^M C_{m,i} y_{i,n} + \epsilon_{L,M,N}, \quad i \neq m \quad (3.10)$$

en donde  $C_{i,j}$  es la correlación entre  $y_i$  y  $y_j$ , específicamente, la correlación calculada sobre las opiniones de los agentes  $i$  y  $j$  en diferentes productos. Esta correlación aintíticamente converge al cruce entre los vectores correspondientes de gustos de los agentes. La cantidad oculta  $L$  puede ser extraída ajustando el factor de proporcionalidad  $\frac{L}{M-1}$ .

El termino de error  $\epsilon$  esta dado por

$$\epsilon \sim \sqrt{\langle a^2 \rangle \langle b^2 \rangle} L^{3/2} \frac{\sqrt{M} + \sqrt{N}}{\sqrt{MN}}. \quad (3.11)$$

Un aspecto de este formalismo importante para aplicaciones es que no existe la necesidad de tener una matriz de opiniones completamente conectada. Los resultados se expanden a conjuntos dispersos simplemente por la definición de los parámetros  $M$  y  $N$  como funciones del par  $(m, n)$ . De esta manera  $M_n$  representa el número de opiniones sobre el producto  $n$  disponibles dadas cualquier agente, y  $N_m$  es el número de opiniones expresadas por el agente  $m$  hacia cualquier producto [1].

### 3.4. REDES DE CONOCIMIENTO Y ANÁLISIS DE SERIES DE TIEMPO

La reducción de ruido e identificación de comportamiento determinístico subyacente son puntos fundamentales de las señales en campos como comunicación [28] y análisis de series de tiempo [5, 11]. Una configuración clásica de modelos relativa a la medición de tales señales [28] considera cada observación en una secuencia  $y_1, y_2, \dots, y_i, \dots, y_T$  que puede ser descompuesta como la suma de una componente determinística y una perturbación aleatoria.

$$y_i = y(t_i) = f(t_i) + \epsilon(t_i). \quad (3.12)$$

Los términos aleatorios  $\epsilon(t_i)$  son estadísticamente independientes entre mediciones. Se considera una señal limpia que puede ser modelada adecuadamente por

una combinación lineal de la forma

$$f(t_i) = \sum_{l=1}^L a_l \varphi(b_l t_i + c_l) \quad (3.13)$$

en donde los términos en la suma son componentes de Fourier. Se proporciona un framework para redes de conocimiento [2] en señales como las descritas por las ecuaciones 3.13 y 3.12 puede ser construido para cierta clase de señales estacionarias.

Sin pérdida de generalidad se puede suponer que la señal y el ruido tienen media cero. Si  $N$  submuestras ordenadas de tamaño  $M$  se extraen de la secuencia observada  $y_1, y_2, \dots, y_i, \dots, y_T$ ,  $y_{m,n}$  se refiere al valor medido en el tiempo  $t_m$  en la submuestra  $n$ , con  $n = 1, 2, \dots, N$  y  $m = 1, 2, \dots, M$ . La presencia de ruido y el tamaño finito de cada submuestra permiten que cada señal extraída pueda ser representada como la combinación lineal de componentes de frecuencia de la señal limpia más los nuevos componentes introducidos por el ruido,

$$y_{m,n} = \sum_{l=1}^L a_{n,l} \varphi_{m,l} + \varepsilon_{m,n}. \quad (3.14)$$

El modelo 3.14 representa la variabilidad en las amplitudes medidas para diferentes frecuencias de Fourier en una muestra finita, tal que

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N a_{n,l} = a_l. \quad (3.15)$$

Otros promedios relacionados con la ecuación 3.14 son los siguientes:

$$\begin{aligned}\lim_{M \rightarrow \infty} \frac{1}{M} \sum_{m=1}^M \varphi_{m,l} &= 0, \\ \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{m=1}^M \varepsilon_{m,n} &= 0, \\ \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \varepsilon_{m,n} &= 0\end{aligned}\tag{3.16}$$

los cuales provienen de las propiedades del ruido dada la estacionalidad supuesta y a que la señal  $y$  tiene media cero. Además

$$\begin{aligned}\lim_{M \rightarrow \infty} \frac{1}{M} \sum_{m=1}^M \varphi_{m,l_1} \varphi_{m,l_2} &= 0, \quad l_1 \neq l_2 \\ \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{m=1}^M \varphi_{m,l_1} \varphi_{m,l_2} &= \langle \varphi_l^2 \rangle, \quad l_1 = l_2.\end{aligned}\tag{3.17}$$

El modelo de señal 3.14 obtiene fuertemente la estructura de los cruces discutidos anteriormente, un enfoque similar puede ser utilizado con el propósito de obtener un formalismo para pronóstico que relaciona la dimensionalidad  $L$  con las fluctuaciones observadas.

Se considera el promedio  $\langle h \rangle$  de una función arbitraria  $h(a_{n,l}, \varphi_{l,m})$  en el límite  $N \rightarrow \infty, M \rightarrow \infty$ ,

$$\langle h \rangle = \frac{1}{L} \sum_{l=1}^L \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{m=1}^M h(a_{n,l}, \varphi_{l,m})\tag{3.18}$$

en este límite la varianza de la señal es expresada por

$$\langle y^2 \rangle = L \langle a^2 \rangle \langle \varphi^2 \rangle + \langle \varepsilon^2 \rangle\tag{3.19}$$

Introduciendo las definiciones

$$A = \begin{pmatrix} a_{1,1} & \dots & a_{N,1} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ a_{1,L} & \dots & a_{N,L} \end{pmatrix}, \Phi = \begin{pmatrix} \varphi_{1,1} & \dots & \varphi_{M,1} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \varphi_{1,L} & \dots & \varphi_{M,L} \end{pmatrix} \quad (3.20)$$

y

$$\Gamma = \begin{pmatrix} \varepsilon_{1,1} & \dots & \varepsilon_{1,N} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \varepsilon_{M,1} & \dots & \varepsilon_{M,N} \end{pmatrix}, \quad (3.21)$$

la configuración del modelo dado por la ecuación 3.14 puede ser escrita en forma matricial como

$$Y = \Phi_n^T A + \Gamma. \quad (3.22)$$

Posteriormente se introduce la operación

$$\hat{Y} = \frac{k}{M} CY, \quad (3.23)$$

en donde  $C$  es la matriz de correlación de las  $y$ . Ahora se muestra que  $\hat{Y} = Y - \Gamma$  en el límite  $N \rightarrow \infty, M \rightarrow \infty$ , utilizando un valor adecuado para el factor  $k$ .

La formula 3.23 se expande:

$$\begin{aligned}
 \hat{Y} &= \frac{k}{M} \frac{YY^{\tau}}{\langle y^2 \rangle} Y & (3.24) \\
 &= \frac{k}{M} \frac{[\Phi^{\tau} A + \Gamma] [A^{\tau} \Phi + \Gamma^{\tau}]}{N [L \langle a^2 \rangle \langle \varphi^2 \rangle + \langle \varepsilon^2 \rangle]} Y \\
 &= \frac{k}{M} \frac{[\Phi^{\tau} A A^{\tau} \Phi \Phi^{\tau} A + \Gamma \Gamma^{\tau} \Phi^{\tau} A]}{N [L \langle a^2 \rangle \langle \varphi^2 \rangle + \langle \varepsilon^2 \rangle]}.
 \end{aligned}$$

Introduciendo los resultados 3.17 en la ecuación 3.24

$$\hat{Y} = \frac{k}{M} \frac{M \langle a^2 \rangle \langle \varphi^2 \rangle + 2 \langle \varepsilon^2 \rangle}{L \langle a^2 \rangle \langle \varphi^2 \rangle + \langle \varepsilon^2 \rangle} \Phi^{\tau} A. \quad (3.25)$$

El factor  $k$  debe ser elegido como

$$k = \frac{M [L \langle a^2 \rangle \langle \varphi^2 \rangle + \langle \varepsilon^2 \rangle]}{M \langle a^2 \rangle \langle \varphi^2 \rangle + 2 \langle \varepsilon^2 \rangle} \quad (3.26)$$

Al utilizar la ecuación 3.19 en la ecuación 3.26 se obtiene una expresión para  $L$  en términos de características medibles

$$L = \frac{\alpha M [\langle y^2 \rangle - \langle \varepsilon^2 \rangle]}{\langle y^2 \rangle - 2\alpha \langle \varepsilon^2 \rangle}. \quad (3.27)$$

en donde  $\alpha = \frac{k}{M}$

Con el propósito de mostrar como se obtienen los términos del lado derecho en la ecuación 3.27 se considera el siguiente algoritmo para reducción de ruido y estimación de complejidad óptima en modelos para señales estacionarias.

La formula de anticipación en esta caso es

$$\hat{y}(t_i) = \frac{k}{M} \sum_{h=1, h \neq i}^M C(t_h) y(t_i - t_h). \quad (3.28)$$

La señal es procesada realizando los siguientes pasos:

1. Calcular la función de autocorrelación  $C(t)$ .
2. Aplicar mínimos cuadrados sobre una muestra de  $M$  puntos consecutivos para estimar el factor  $\alpha = \frac{k}{M}$  en la ecuación 3.28.

El problema de mínimos cuadrados puede ser resuelto exactamente tomando

$$\alpha = \frac{\sum_{i=1}^M y(t_i) \sum_{\tau=1}^M C(t_\tau) y(t_i - t_\tau)}{\sum_{j=1}^M \sum_{\tau_1=1}^M C(t_{\tau_1}) y(t_j - t_{\tau_1}) \sum_{\tau_2=1}^M C(t_{\tau_2}) y(t_j - t_{\tau_2})} \quad (3.29)$$

$$i \neq \tau, j \neq \tau_1, j \neq \tau_2,$$

con  $M$  menor o igual a un medio de la longitud total de la señal. El termino  $\langle \varepsilon^2 \rangle$  es estimado despues del filtrado, utilizando los datos filtrados como aproximación a la señal determinística subyacente y realizando la substracción  $\langle \varepsilon^2 \rangle = \langle y^2 \rangle - \langle f^2 \rangle$ .

3. Utilizando la ecuación 3.27, calcular  $L$  en términos de las cantidades observables

Los pasos 1 - 3 definen lo que es llamado algoritmo de red de conocimiento para reducción de ruido, KNNR por sus siglas en ingles (Knowledge Network for Noise Reduction).

### 3.5. OBTENCIÓN DE FUNCIONES DE DENSIDAD PARA PROBLEMAS DE OPTIMIZACIÓN NO LINEALES

Cada día los modelos no lineales se vuelven más importantes en todos los campos de la ciencia e ingeniería, por ejemplo, se consideran esenciales las interacciones no lineales para la obtención de comportamiento colectivo en sistemas dinámicos extendidos [23]. Por otro lado es ampliamente aceptado [11] que las señales generadas por muchos sistemas complejos pueden ser modelados mediante ecuaciones de evolución no lineales. Otro ejemplo importante es la pregunta general para diseño óptimo en ingeniería, en donde los modelos lineales no son más que aproximaciones burdas [18]. Una limitación en el uso de modelos no lineales es que con el propósito de ser cuantitativamente significativos, los modelos deben ser ajustados a los datos. Este paso implica la optimización de cierta función de costo o verosimilitud, la cual es frecuentemente definida en un espacio de parámetros altamente dimensional.

En este trabajo se presenta la implementación de una técnica estocástica recientemente introducida [1]. El método propuesto está fundamentado en la relación entre los ambientes de trabajo de procesos estocásticos en los cuales se utiliza la ecuación de Langevin y de Fokker - Planck, lo cuales son conocidos en el estudio de sistemas físicos fuera de equilibrio [22, 26]. Dada una función de costo  $V(x_1, x_2, \dots, x_n, \dots, x_N)$ , una búsqueda estocástica puede ser modelada por la ecuación de Langevin

$$\dot{x}_n = -\frac{\partial V}{\partial x_n} + \varepsilon(t) \quad (3.30)$$

en donde  $\varepsilon(t)$  es un ruido aditivo con media cero y segundo momento  $E[\varepsilon(t)\varepsilon(t')] = D\delta(t - t')$ . La cantidad  $D$  es llamada coeficiente de difusión. Con un  $D$  constante, el proceso 3.30 produce una búsqueda mediante difusión. Considerando un coeficiente de difusión que disminuye lentamente con el tiempo la ecuación 3.30 representa



un proceso de recocido simulado [13,25].

Además otras estrategias de búsqueda (por ejemplo, aquellas que involucran memoria) pueden ser modeladas por la ecuación 3.30 después de la adición de correlaciones al ruido aditivo. Se considera que  $\varepsilon(t)$  es ruido blanco Gaussiano con media cero y segundo momento  $E[\varepsilon(t)\varepsilon(t')] = D\delta(t-t')$  en donde  $D$  es constante y  $\delta(t-t')$  es la función delta de Dirac. Con estas suposiciones, la densidad de probabilidad asociada al proceso estocástico de la ecuación 3.30 está dado por la ecuación de Fokker–Planck [22,26].

$$\dot{p} = \frac{\partial}{\partial x} \left[ \frac{\partial V}{\partial x} p \right] + D \frac{\partial^2 p}{\partial x^2} \quad (3.31)$$

El principio fundamental de este enfoque es que para espacios de búsqueda acotados con condiciones reflejantes en los extremos, la densidad  $p$  converge al estado estacionario [7]. De esta manera se puede estudiar el límite  $t \rightarrow \infty$  del proceso de búsqueda 3.30 observando las soluciones estacionarias de la ecuación 3.31, dado esto, el problema de optimización asociado se define en la región

$$L_{1,n} \leq x_n \leq L_{2,n}. \quad (3.32)$$

### 3.6. ALGORITMO FOKKER–PLANCK ESTACIONARIO

En esta sección se tratan las generalidades del algoritmo propuesto en [1]. La proyección unidimensional de la ecuación 3.31 en  $t \rightarrow \infty$  está dada por

$$D \frac{\partial p(x_n | \{x_{j \neq n} = x_j^*\})}{\partial x_n} + p(x_n | \{x_{j \neq n} = x_j^*\}) \frac{\partial V}{\partial x_n} = 0. \quad (3.33)$$

En el estado estacionario la proyección unidimensional proporciona la densidad de probabilidad dependiente que corresponde al proceso  $N$  dimensional original [1]. De la ecuación 3.33 se obtiene una ecuación diferencial de segundo orden para la distribución de probabilidad acumulada para  $x_n$ , denotada por  $y(x_n)$

$$\frac{d^2 y}{dx_n^2} + \frac{1}{D} \frac{\partial V}{\partial x_n} \frac{dy}{dx_n} = 0 \quad (3.34)$$

$$y(L_{1,n}) = 0, \quad y(L_{2,n}) = 1$$

Es posible generar puntos aleatorios provenientes de la densidad  $p(x_n | \{x_{j \neq n} = x_j^*\})$  por el hecho de que  $y$  es una variable aleatoria distribuida uniformemente en el intervalo  $y \in [0,1]$ , lo cual se tiene mediante las leyes de transformación de funciones de probabilidad [1]. Al ver  $y(x_n)$  como una función de la variable aleatoria  $x_n$ , puede ser aproximada mediante una combinación lineal de funciones de un conjunto completo que satisfaga las condiciones de los extremos en el intervalo de interés.

$$\hat{y} = \sum_{l=1}^L a_l \varphi_l(x_n). \quad (3.35)$$

Eligiendo, por ejemplo, una base en la cual  $\varphi_l(0) = 0$ , los  $L$  coeficientes son definidos únicamente por la evaluación de la ecuación 3.34 en  $L - 1$  puntos interiores. De esta manera la aproximación de  $y$  se realiza resolviendo un conjunto de  $L$  ecuaciones algebraicas, implicando  $L - 1$  evaluaciones de la derivada de  $V$ .

El procedimiento propuesto está basado en la interacción de los siguientes pasos

1. Fijar las variables  $x_{j \neq n} = x_j^*$  y aproximar  $y(x_n)$  mediante el uso de las ecuaciones 3.34 y 3.35.

2. Mediante el uso de  $\hat{y}(x_n)$  se obtienen puntos de la distribución estacionaria  $p(x_n | \{x_{j \neq n} = x_j^*\})$ .
3. Actualizar  $x_n = x_n^*$  y repetir el procedimiento para la siguiente variable  $x_{j \neq n}$ .

Este procedimiento básico es llamado Fokker–Planck Estacionario (SFP, por sus siglas en inglés, Stationary Fokker - Planck), en donde el nombre indica la manera en que se aprende la distribución en equilibrio del proceso de búsqueda estocástico dado por 3.30, utilizando la iteración de los pasos anteriores.

El algoritmo tiene sus raíces en la ecuación de Fokker–Planck estacionaria. En contraste con la Máquina de aprendizaje Fokker–Planck [25], la cual usa una versión dependiente del tiempo de la misma ecuación. El procedimiento aquí presentado no trata con la búsqueda de un óptimo global mediante poblaciones de puntos finitas, el cual es el caso de la Máquina de Fokker - Planck Estacionaria y otros métodos relacionados [24, 25], el método presentado estima la expresión para la función de distribución de probabilidad.

La convergencia del algoritmo SFP a la distribución en equilibrio proviene del hecho que el procedimiento representa un muestreo de Gibbs. Bajo condiciones generales [6], los puntos generados mediante el muestreo de Gibbs están asintóticamente distribuidos de acuerdo a la distribución  $N$ -dimensional asociada.

En este trabajo se presenta una implementación del algoritmo que es capaz de generar aproximaciones a la distribución en equilibrio en una manera eficiente desde el punto de vista computacional.

## DESCRIPCIÓN DEL PROBLEMA

---

### 4.1. SERIES DE TIEMPO LINEALES, EL MODELO KNNR

Existen distintos tipos de modelos para series de tiempo que trabajan con relaciones lineales en los datos de entrada, sin embargo es común encontrar que éstos modelos requieren la estimación de una cantidad considerable de parámetros. Esta estimación de parámetros en muchas ocasiones requiere de una cantidad considerable de cálculos, lo cual puede tomar un cantidad de tiempo considerable.

En una situación en donde se debe tomar una decisión en base al pronóstico obtenido el tiempo de respuesta entre la generación de los datos de entrada para la serie de tiempo y el ajuste del modelo es crucial, por ejemplo al realizar el cálculo de la cantidad de combustible que se inyectará a un motor que controla la velocidad en un avión. Otro caso encontrado comúnmente es la ausencia de un grán equipo de computo y almacenamiento como puede ser el caso del control del motor en un refrigerador en donde los recursos de cómputo son muy limitados, debido a que se utilizan microcontroladores con capacidad pequeña por cuestiones de costo.

Este tipo de aplicaciones requieren modelos que contengan una cantidad pequeña de parámetros por estimar además de un algoritmo que pueda estimar los valores con una potencia de cómputo considerada.

Dados estos casos, un modelo de pronóstico que contenga un solo parámetro es de mucha utilidad. Debido a esto se presenta el modelo KNNR, el cual requiere la estimación de solamente un parámetro y que además presenta solamente operaciones lineales para obtener su pronóstico. Además el parámetro se estima mediante mínimos cuadrados con los datos de entrada.

Otra bondad del modelo KNNR contra los demás existentes es que presenta un gran desempeño cuando se cuenta con pocos valores de entrada, obteniendo resultados que incluso mejoran modelos más complejos como el AR.

La importancia del modelo KNNR recae en que sólo requiere la estimación de un parámetro y que además es calculado por medio de operaciones lineales que requieren un esfuerzo computacional considerado. Además de obtener resultados comparables contra métodos complejos.

## 4.2. SERIES DE TIEMPO NO LINEALES, OBTENCIÓN DE FUNCIONES DE DENSIDAD

En la naturaleza es común encontrar modelos no lineales para describir procesos los procesos que ocurren de manera cotidiana, como en el caso de el pronóstico del clima, en donde un cambio pequeño en ciertas condiciones produce resultados complejos a través del sistema. Debido a la naturaleza no lineal de estos modelos, es difícil para una computadora evaluar las funciones de costo asociadas a los mismos. Además de esta evaluación es necesario ajustar los parámetros para lograr un desempeño óptimo. Este último problema requiere la utilización de herramientas que permitan obtener información acerca de como se afecta cada uno de éstos parámetros a la respuesta del sistema.

En este trabajo se presenta un método para la obtención de la función de densidad de probabilidad asociada a cada uno de los parámetros en un modelo de pronóstico no lineal. Este método además utiliza solamente operaciones del tipo lineal para realizar la estimación de la función de densidad, además de que el esfuerzo computacional puede ser regulado en base a que tanta precisión se desea en la función de densidad, es decir, al aplicar mayor esfuerzo se obtiene una función más precisa, sin embargo con muy poco esfuerzo es posible obtener resultados muy buenos, o con una gran precisión.

El obtener una expresión para estos valores permite obtener una aplicación robusta respecto a los cambios en los valores de los parámetros ya que en algunas aplicaciones es imposible controlarlos por completo, ya que pueden ser valores ajustados en base a una señal de entrada, como es el caso de aplicaciones de hardware en las cuales el valor de los parámetros puede ser un voltaje de entrada que puede tener variaciones debido a la naturaleza del sistema.

Con el propósito de ilustrar la amplia gama de aplicaciones que tiene el obtener funciones de densidad para los parámetros de un modelo, éste se aplica en distintos contextos que van desde funciones de benchmark, hasta el entrenamiento de una red neuronal; obteniendo resultados muy buenos utilizando un esfuerzo computacional considerado.

## EXPERIMENTACIÓN

---

### 5.1. COMPARACIÓN DE LOS MODELOS AR Y KNNR

El algoritmo KNNR es probado numéricamente generando datos mediante simulación, a estos datos simulados, representando  $f(t)$ , se agrega un término de ruido blanco Gaussiano  $\varepsilon(t)$  en cada realización. La simulación del ruido está basada en el algoritmo de L'Ecuyer, el cual se sabe que presenta un desempeño adecuado con respecto a las principales pruebas estadísticas, y produce secuencias de números aleatorios con longitud de  $\sim 10^{18}$  [21]. El algoritmo KNNR utiliza los datos con ruido  $y(t) = f(t) + \varepsilon(t)$ .

La función de autocorrelación se calcula utilizando la transformada rápida de Fourier [21]. La función de autocorrelación es calculada para un retardo máximo de tamaño igual a la mitad a la longitud de la señal. Los pasos 2 y 3 del algoritmo KNNR son calculados en la segunda mitad de los datos de entrada.

El algoritmo KNNR provee una caracterización de la complejidad del modelo de señal en términos de  $L$ , el número de componentes de Fourier necesarias para describir la señal, si ésta es separable en el sentido de la ecuación (3.12). Si se cumplen las suficientes suposiciones,  $L$  debería converger a un valor finito conforme el tamaño de muestra crece. Este hecho puede ser utilizado para identificar la presencia

de frecuencias relevantes en la señal.

Se experimentó con cuatro tipos de señales, tres de ellas obtenidas mediante simulación: señales generadas por combinaciones lineales de funciones coseno con ruido aditivo, señales quasiperiodicas generadas por mapas no lineales, señales generadas por procesos estocásticos lineales y datos experimentales. El pronóstico del KNNR es comparado con el pronóstico del mejor modelo AR obtenido de acuerdo al criterio AIC, este modelo es obtenido utilizando el software estadístico R. Para ambos casos se calcula el MSE del pronóstico así como la desviación estándar de los datos de prueba para ser usada como referencia.

### 5.1.1. COMBINACIONES LINEALES DE FUNCIONES COSENO CON RUIDO ADITIVO

Para este caso se generan series de tiempo mediante simulaciones del modelo de la ecuación 3.12, tomando la función  $f(t_i)$  como una combinación lineal de funciones coseno. Los parámetros de amplitud, periodo y fase de las componentes  $C$  provienen de números aleatorios con distribución uniforme. Las amplitudes se encuentran en el intervalo  $[0,1]$ , mientras que los periodos y fases se encuentran en los intervalos  $[T_1, T_2]$  y  $[P_1, P_2]$  respectivamente.

#### RESULTADOS OBTENIDOS

En la figura 5.1 se presentan dos gráficas de ejemplo de los pronósticos obtenidos por los modelos (KNNR y AR). El cuadro (5.1) muestra los resultados obtenidos para distintas serie de tiempo de este tipo.



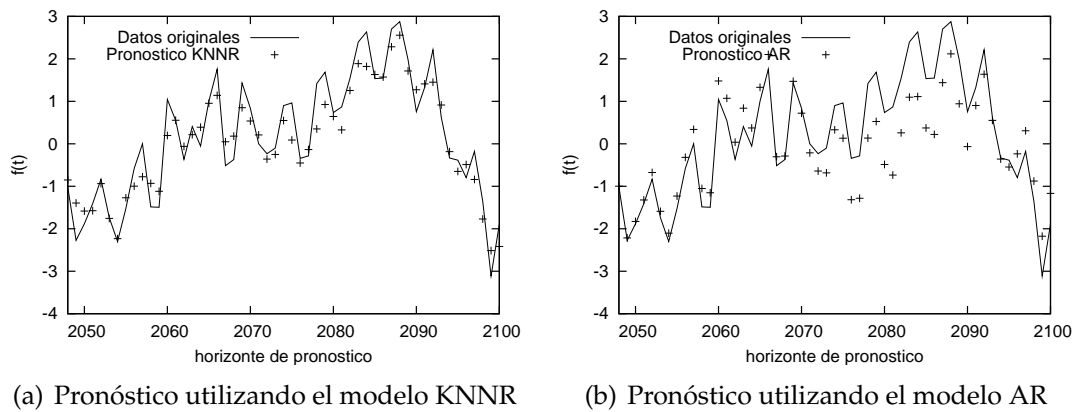


Figura 5.1: Pronóstico para la combinación lineal de diez funciones coseno sin ruido

## CONCLUSIONES

Como puede verse en el cuadro 5.1, hay ocasiones en que el modelo AR es mejor que el modelo KNNR. Esto es debido a que la representación de Fourier para estos datos presenta puntos de alta relevancia, esto es, la serie de tiempo presenta periodicidad notable. Sin embargo en los casos en que no hay manera de distinguir esta periodicidad, el modelo KNNR presenta un menor error.

### 5.1.2. MAPAS NO LINEALES

Bajo la suposición de que es posible encontrar un comportamiento determinista en un mapa caótico, el modelo KNNR es probado con el mapa logístico [27]. Las series de tiempo son generadas utilizando la ecuación del mapa logístico  $x_{n+1} = rx_n(1 - x_n)$  con un valor  $r$  de  $r = 3,6$ , además contiene ruido blanco Gaussiano aditivo.

## RESULTADOS OBTENIDOS

En la figura 5.2 se muestra un ejemplo del resultado de aplicar el modelo KNNR y el AR a series de tiempo generadas por el mapa logístico. En el cuadro 5.2 se presentan los resultados para distintas realizaciones del mapa logístico.

Cuadro 5.1: Resultados del pronóstico con los modelos AR y KNNR para series de tiempo generadas por combinación lineal de cosenos con diferentes horizontes H.

$T_1$	$T_2$	$P_1$	$P_2$	C	Ruido	H	Datos Entrada	Orden del AR	SD	Error KNNR	Error AR
0	1	0	1024	10	0.0	25	1024	29	1.86452	0.39050	0.61770
0	1	0	1024	10	0.0	200	1024	29	2.13162	0.80713	1.62328
0	1	0	4096	10	0.0	250	4096	36	2.15395	0.61242	1.21589
0	1	0	4096	10	0.7	250	4096	33	2.36488	2.08362	2.37548
50	100	0	4096	15	0.7	250	4096	36	2.93307	1.24024	2.4665

Cuadro 5.2: Resultados del pronóstico con los modelos AR y KNNR, para el mapa logístico.

Datos de entrada	Horizonte	Ruido	Orden AR	SD	KNNR	AR
8192	50	0.0	23	0.04993	0.00445	0.00448
8192	100	0.0	23	0.04943	0.00445	0.0045
8192	10	0.1	38	0.06244	0.0139	0.01404
8192	100	0.1	38	0.05777	0.01388	0.01433
16384	100	0.7	42	0.52894	0.48319	0.5073

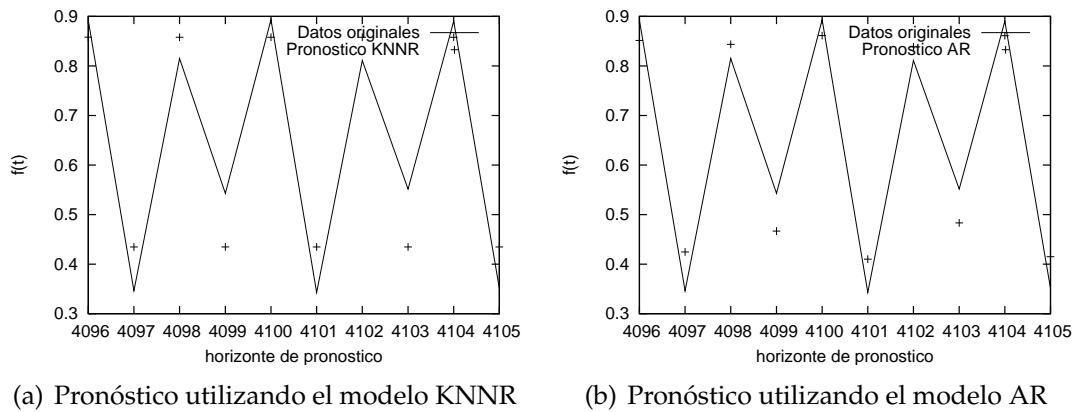


Figura 5.2: Pronóstico para el mapa logístico con parámetro  $r = 3,6$  sin ruido.

## CONCLUSIONES

La diferencia en el error de pronóstico para ambos modelos es muy pequeña. Sin embargo es importante resaltar que el modelo KNNR solo requiere la estimación de un parámetro, en contraste con el modelo AR el cual en todas las series de tiempo generadas por el mapa logístico tiene un orden “grande”.

Debido a esto el modelo AR requiere la estimación de una cantidad de parámetros considerable, además de que presenta la necesidad de un criterio para determinar cual es el mejor orden para el modelo AR. Todo esto implica un mayor número de cálculos computacionales para obtener un modelo adecuado a los datos.

### 5.1.3. PROCESOS ESTOCÁSTICOS LINEALES

Con el propósito de comparar el desempeño del modelo KNNR contra la metodología para obtener el mejor modelo AR se generan series de tiempo mediante un modelo AR de orden 12, con ruido blanco Gaussiano.

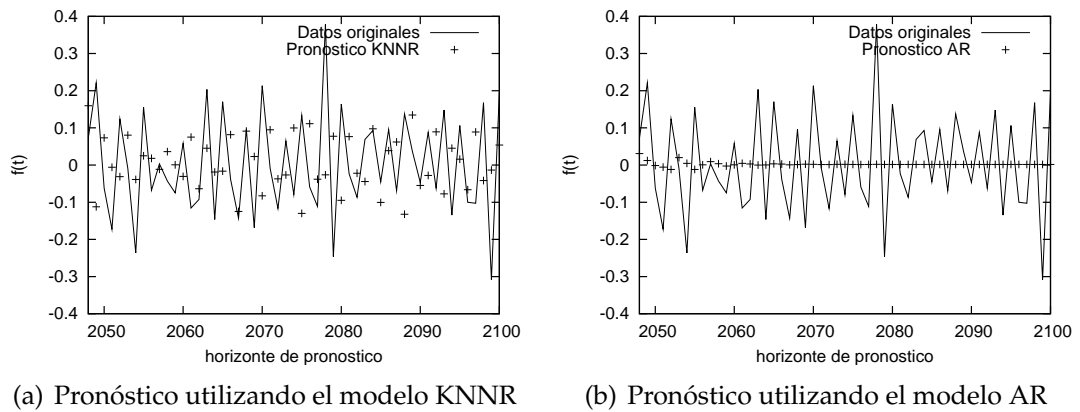


Figura 5.3: Resultado del pronóstico para una serie de tiempo generada mediante un modelo AR de orden 23 y con intensidad de ruido de 0,1

#### RESULTADOS OBTENIDOS

La figura 5.3 presenta un ejemplo del resultado de aplicar los modelos AR y KNNR a series de tiempo generadas por el modelo AR. El cuadro 5.3 presenta los resultados obtenidos para distintas series de tiempo generadas por el modelo AR.

#### CONCLUSIONES

Como se puede apreciar en el cuadro 5.3, los resultados de ambos modelos son muy similares. En ocasiones el error del modelo KNNR es inferior al del modelo AR. Esto se presenta cuando se tienen pocos datos de entrada, esto es de importancia ya que en aplicaciones de la vida real es común encontrar series de tiempo con pocos datos. En estas situaciones el modelo KNNR aparentemente es un buen modelo de pronóstico, además de las ventajas de que solo requiere la estimación de un parámetro.

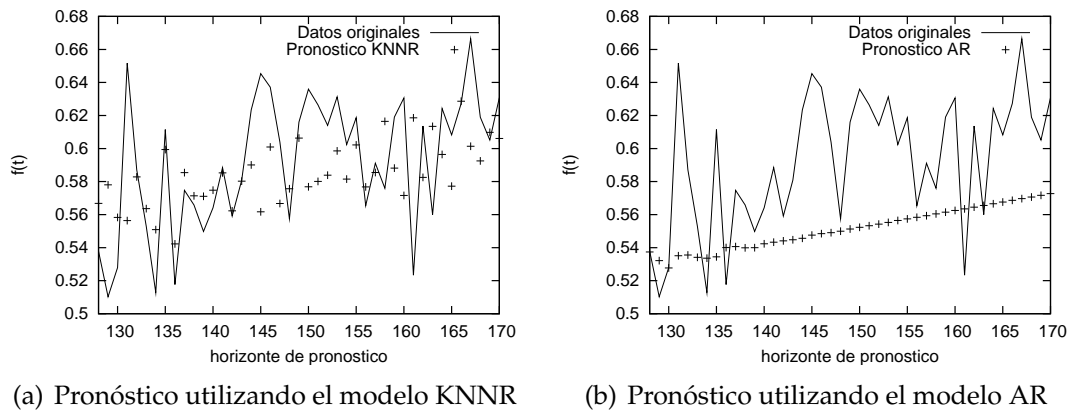


Figura 5.4: Pronóstico de los modelos AR y KNNR para el producto 1 de los datos experimentales

#### 5.1.4. DATOS EXPERIMENTALES

Debido a la importancia de utilizar datos de la vida real, además de los datos generados mediante simulación se utilizan datos provenientes de una industria local del ramo alimenticio. Estos datos son ventas por semana a nivel nacional. Los productos 1, 2 y 3 provienen de la misma familia, mientras el producto 4 es una clase distinta de productos. En todos los casos se aplican diferencias con retardo de 1, con el propósito de eliminar la tendencia de la serie de tiempo.

#### RESULTADOS OBTENIDOS

En la figura 5.4 se presenta un ejemplo de las series de tiempo obtenidas de datos de la vida real. El cuadro 5.4 presenta los resultados del pronóstico obtenido mediante los modelos AR y KNNR.

Cuadro 5.3: Resultados del pronóstico mediante los modelos AR y KNNR para series de tiempo generadas mediante el modelo AR.

Datos de entrada	Horizonte	Ruido	Orden AR	SD	KNNR	AR
128	50	0.1	5	0.02182	0.01975	0.01975
256	100	0.1	5	0.02321	0.01667	0.01667
128	50	0.2	5	0.08727	0.07643	0.07643
256	100	0.2	5	0.09283	0.06639	0.06639
512	10	0.2	11	0.061	0.06557	0.05538

Cuadro 5.4: Resultados del pronóstico para los modelos AR y KNNR con datos de la vida real.

Producto	Datos de entrada	Horizonte	Orden AR	SD	KNNR	AR
Producto 1	32	10	4	0.02573	0.00364	0.0037
Producto 1	128	50	8	0.00372	0.0019992	0.002
Producto 2	32	10	3	0.03855	0.0055	0.00557
Producto 3	128	50	9	0.00495	0.0036	0.00361
Producto 4	64	30	3	0.00116	0.00055	0.00055

## CONCLUSIONES

A pesar de que las series de tiempo no son completamente estacionarias, el modelo KNNR puede pronosticar una porción de la forma de los datos. Esto es importante ya que el modelo es lineal, cuando es probable que la serie de tiempo contenga relaciones no lineales. Esto muestra la flexibilidad del modelo a los distintos tipos de series de tiempo que se presentan en la vida real.

## 5.2. OBTENCIÓN DE FUNCIONES DE DENSIDAD PARA PROBLEMAS DE OPTIMIZACIÓN NO LINEALES

Existe una gran variedad de series de tiempo para las cuales los modelos lineales presentan un error de pronóstico muy grande. Las posibles causas de esto son:

1. No existen dependencias en la serie que se puedan aprovechar para hacer pronósticos.
2. Existen dependencias no-lineales.

Cuando los datos caen en el segundo caso, es necesario utilizar modelos que incluyan la no linealidad de los datos. Sin embargo la utilización de estos modelos presenta la desventaja de que para encontrar los valores adecuados para sus parámetros, es necesario resolver un problema de optimización global. Dada esta necesidad, el algoritmo SFP busca proporcionar una herramienta más para la solución de problemas de optimización global, la cual es la función de densidad de probabilidad asociada al problema tratado.



El algoritmo SFP es aplicado en distintos problemas de optimización no lineal utilizandolo como generador de puntos iniciales para algoritmos determinísticos. Esto con el propósito de mostrar que al utilizar puntos generados mediante la función de densidad obtenida por SFP se presenta una mejora notoria en la calidad de las soluciones obtenidas por los algoritmos determinísticos. Esto solo es posible si la densidad obtenida mediante SFP es acorde con la densidad del problema de optimización tratado.

Otra de las ventajas de SFP es que se puede controlar la complejidad computacional que se desea. El algoritmo requiere de  $2(L - 1)MN$  evaluaciones de la función objetivo por lo que se pueden obtener funciones de densidad tan “ajustadas” como se desee.

El algoritmo determinístico utilizado en la experimentación es el algoritmo de Powell, el cual se basa en direcciones conjugadas, teniendo la ventaja de no necesitar calcular el gradiente de la función objetivo.

En este trabajo se utiliza el algoritmo SFP para calcular la función de densidad para optimizar los siguientes problemas: función de Levy No 5 5.1, una red neuronal para aprender la tabla XOR, una red neuronal que aprende el mapa logístico las cuales son estructuras altamente no lineales.

### 5.2.1. FUNCIÓN DE LEVY NO. 5

La función de Levy No. 5 es una comúnmente utilizada como prueba para algoritmos de solución de problemas de optimización no lineal [19]. La función de

Levy No. 5 es la siguiente:

$$f(x) = \sum_{i=1}^5 i \cos((i-1)x_1 + i) \sum_{j=1}^5 j \cos((j+1)x_2 + j) + (x_1 + 1,42513)^2 + (x_2 + 0,80032)^2, \quad (5.1)$$

con espacio de búsqueda definido en el hipercubo  $[-10,10]$ , contiene 760 mínimos locales y un óptimo global con valor de  $f(x^*) = -176,1375$ , en el punto  $x^* = (-1,3068, -1,4248)^T$ . El gran número de mínimos locales la hace difícil para cualquier método para encontrar el mínimo global.

La implementación de la búsqueda estocástica a través de la ecuación 3.30 implica la simulación de un sistema estocástico dinámico compuesto por dos partículas con interacciones altamente no lineales. Utilizando la metodología de SFP en contraste, se obtienen funciones de densidad utilizando operaciones lineales además de solo realizar una cantidad moderada de evaluaciones de la función de costo.

En la figura 5.5 se muestran las funciones de densidad generadas por 10 iteraciones de SFP con parámetros  $L = 50$  y  $D = 200$ . Las densidades obtenidas son perfectamente consistentes con las propiedades globales del problema, dado que el punto óptimo  $x^* = (-1,3068, -1,4248)^T$  conocido se encuentra en la región de máxima probabilidad. El esfuerzo computacional es bajo en el sentido que la cantidad de evaluaciones de la función objetivo es  $2(L-1)MN = 4900$ . Esto es comparable con el esfuerzo requerido por técnicas avanzadas basadas en poblaciones que obtienen soluciones de buena calidad para este problema [19].

El experimento realizado consiste en generar la función de densidad mediante SFP. Posteriormente se utiliza como semilla para el algoritmo de Powell el punto con máxima probabilidad proveniente de la densidad generada, con el propósito de el valor óptimo para cada una de las variables de decisión. Esto se repite en 100

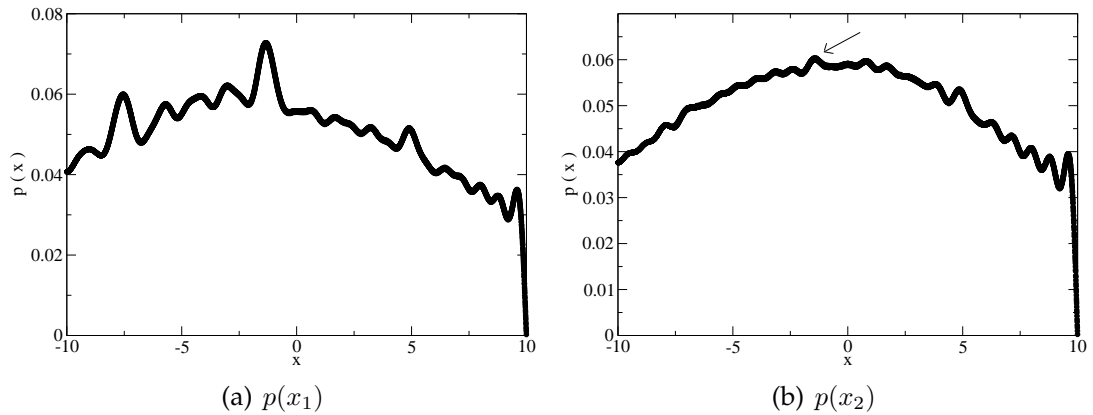


Figura 5.5: Densidades de probabilidad generadas por 10 iteraciones de SFP para la función de Levy No. 5. Los parámetros de SFP son  $L = 50$  y  $D = 200$ . El óptimo global se encuentra en la región de máxima probabilidad.

Cuadro 5.5: Tasa de éxitos para el problema de Levy No. 5, utilizando el punto con máxima probabilidad como punto inicial para el algoritmo de Powell. Cada realización consiste de 100 ejecuciones del algoritmo de Powell. Se considera éxito si el valor objetivo difiere en 0,0001 con el óptimo global.

Realización	1	2	3	4	5	6	7	8	9	10
Tasa de éxito	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

ocasiones y se calcula un promedio de éxito. Este procedimiento se realiza 10 veces. De la misma manera se utiliza una función de densidad uniforme en lugar de la generada por SFP y se ejecuta el algoritmo Powell para encontrar el óptimo global.

## RESULTADOS OBTENIDOS

El cuadro 5.2.1 presenta los resultados obtenidos utilizando la función de densidad generada mediante el algoritmo SFP. En el cuadro 5.2.1 se presentan los resultados obtenidos utilizando la función de densidad uniforme como generador de números aleatorios para el algoritmo de Powell. En ambos casos se considera éxito si el valor objetivo difiere en 0,0001 con el óptimo global.

Cuadro 5.6: Tasa de éxitos para el problema de Levy No. 5, utilizando puntos generados de una distribución uniforme como puntos iniciales para el algoritmo de Powell. Cada realización consiste de 100 ejecuciones del algoritmo de Powell. Se considera éxito si el valor objetivo difiere en 0,0001 con el óptimo global.

Realización	1	2	3	4	5	6	7	8	9	10
Tasa de éxito	0.00	0.04	0.01	0.05	0.02	0.03	0.07	0.08	0.08	0.05

## CONCLUSIONES

La tasa de éxito promedio para el algoritmo de Powell utilizando puntos generados mediante la función de densidad obtenida por SFP es 1,00, en contraste. La tasa de éxito promedio utilizando puntos iniciales generados mediante la distribución uniforme es de solo 0,043. Esto implica que SFP fue capaz de aprender la función de densidad con tan solo 9800 evaluaciones de la función objetivo.

### 5.2.2. PROBLEMA XOR

La optimización de los pesos de una red neuronal que aprende la tabla XOR es un ejemplo clásico que muestra las dificultades en el ajuste de parámetros de modelos no lineales. Se tiene una arquitectura definida por dos nodos lineales de entrada, dos nodos en la capa oculta y un nodo en la capa de salida utilizando la función de activación logística. El problema de optimización es:

$$\text{mín } f, \quad (5.2)$$

Cuadro 5.7: Tasa de éxitos para el problema XOR, utilizando puntos iniciales para el algoritmo de Powell generados a partir de la función de densidad obtenida mediante SFP. Cada realización consiste de 100 ejecuciones del algoritmo de Powell, cada una de ellas con una semilla distinta.

Realización	1	2	3	4	5	6	7	8	9	10
Tasa de éxito	0.79	0.48	0.27	0.86	0.7	1.00	0.83	0.36	0.75	0.72

$$\begin{aligned}
 f = & \left\{ 1 + \exp \left( -\frac{x_7}{1 + \exp(-x_1 - x_2 - x_5)} - \frac{x_8}{1 + \exp(-x_3 - x_4 - x_6)} - x_9 \right) \right\}^{-2} \\
 & + \left\{ 1 + \exp \left( -\frac{x_7}{1 + \exp(-x_5)} - \frac{x_8}{1 + \exp(-x_6)} - x_9 \right) \right\}^{-2} \\
 & + \left\{ 1 - \left[ 1 + \exp \left( -\frac{x_7}{1 + \exp(-x_1 - x_5)} - \frac{x_8}{1 + \exp(-x_3 - x_6)} - x_9 \right) \right]^{-1} \right\}^2 \\
 & + \left\{ 1 - \left[ 1 + \exp \left( -\frac{x_7}{1 + \exp(-x_2 - x_5)} - \frac{x_8}{1 + \exp(-x_4 - x_6)} - x_9 \right) \right]^{-1} \right\}^2
 \end{aligned}$$

El experimento consiste en utilizar la densidad generada por SFP como generador de números aleatorios y utilizarlos como semillas para el algoritmo determinístico de Powell, se generan 100 semillas distintas y se ejecuta el algoritmo de Powell con cada una de ellas para encontrar el óptimo global. Este procedimiento se repite en 10 ocasiones. La función de densidad es obtenida por 10 iteraciones de SFP, con los parámetros  $L = 60$ ,  $D = 0,03$ . Este mismo procedimiento se repite utilizando puntos aleatorios generados por una distribución uniforme. El espacio de búsqueda se define en el intervalo  $[-10,10]$ .

## RESULTADOS OBTENIDOS

El resultado de cada una de las 10 iteraciones independientes se resume en el cuadro 5.7 para puntos generados a partir de la función de densidad obtenida por SFP y en el cuadro 5.8 para puntos generados por una distribución uniforme.

Cuadro 5.8: Tasa de éxitos para el problema XOR, utilizando puntos iniciales para el algoritmo de Powell generados a partir una distribución uniforme. Cada realización consiste de 100 ejecuciones del algoritmo de Powell, cada una de ellas con una semilla distinta.

<b>Realización</b>	1	2	3	4	5	6	7	8	9	10
<b>Tasa de éxito</b>	0.27	0.21	0.18	0.27	0.18	0.15	0.24	0.33	0.19	0.26

## CONCLUSIONES

Al utilizar el algoritmo SFP para generar números aleatorios y utilizarlos con el algoritmo de Powell se tiene una tasa de éxito promedio de 0,675, a diferencia de si se utilizan números aleatorios provenientes de una distribución uniforme se tiene una tasa de éxito promedio de solo 0,228, lo cual presenta una notable mejora en los resultados, lo cual sugiere que el algoritmo SFP aprendió correctamente la función de densidad de probabilidad asociada con el problema XOR.

### 5.2.3. RED NEURONAL PARA APRENDER EL MAPA LOGÍSTICO

Además de las aplicaciones anteriores SFP es utilizado para aprender las funciones de densidad asociadas a los pesos de una red neuronal que aprende el mapa logístico. Un problema encontrado es el determinar la arquitectura apropiada para la red neuronal utilizada, en este caso se utilizó el algoritmo Vecinos falsos [12] para determinar la cantidad de neuronas a colocar en la capa de entrada, el cual trata de identificar la dimensión mínima suficiente embebida en los datos. La cantidad de neuronas en la capa oculta fue determinada experimentalmente. La importancia de utilizar una arquitectura de red neuronal adecuada recae en que es necesario optimizar los pesos de esta red; esto implica resolver problemas con alta no linealidad.

En este caso se utiliza el algoritmo SFP con los parámetros  $M = 10, L = 750, D = 0,002$  con intervalo de búsqueda  $[-10,10]$ , se utilizan solamente 64 datos para apren-

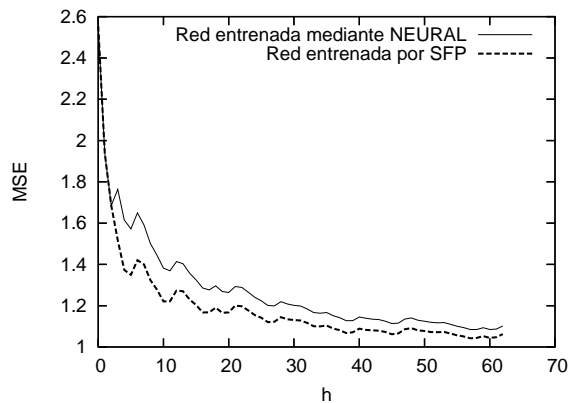


Figura 5.6: Gráfica del MSE en el horizonte  $h$  obtenido al pronosticar datos por una red neuronal entrenada mediante SFP y el software NEURAL

der la función de densidad teniendo como función objetivo minimizar el MSE de los datos de entrenamiento. Por otro lado se utiliza el software “neural” encontrado en la librería de Numerical Recipes for Neural Networks [17].

## RESULTADOS OBTENIDOS

La gráfica de la función de error al utilizar ambas redes neuronales para pronosticar el mapa logístico se presenta en la figura 5.6. La función de error es el MSE de los datos que se han pronosticado hasta el horizonte  $h$ .

## CONCLUSIONES

Al utilizar SFP para entrenar la red neuronal la función de error obtenida se encuentra por “abajo” de la función de error en el caso en que la red neuronal es entrenada mediante el software NEURAL. Esto proporciona evidencia de que la función de densidad obtenida mediante SFP es acorde con el problema de optimización, además de que se tiene una expresión para esta densidad, teniendo por ello, propiedades estadísticas que se pueden aprovechar como en este caso fue el primer

---

momento para cada una de las variables, el cual fue el valor utilizado como los pesos de la red neuronal.



## CONCLUSIONES

---

En este trabajo se presentan herramientas para trabajar distintos tipos de series de tiempo, desde el tipo lineal hasta series de tiempo que presentan relaciones no lineales.

En el aspecto lineal se presenta una herramienta para pronóstico de series de tiempo que utiliza solamente un parámetro para realizar sus calculos, además este es obtenido en base a los datos de la serie de tiempo tratada. Además de utilizar solamente un parámetro los resultados de este modelos son comparables a los obtenidos mediante modelos más complejos, como es el caso del modelo AR.

Se realiza una comparación en el desempeño del modelo KNNR contra el modelo AR utilizando distintas series de tiempo, una de las cuales proviene de datos experimentales tomados de una empresa del ramo alimenticio de la localidad. En todos los casos se obtienen resultados que son comparables a los obtenidos por el modelo AR, y en algunas ocasiones se superan los resultados del mismo, mostrando así, que el modelo KNNR es capaz de encontrar las relaciones ocultas en las series de tiempo utilizando series de tiempo con pocos datos.

Para el caso de series de tiempo no lineales, se presenta un algoritmo para encontrar una expresión de la función de densidad de probabilidad asociada con el

problema de ajuste de parámetros en modelos no lineales. Esto presenta una serie de ventajas debido a que al obtener una expresión para esta función de densidad utilizando la cantidad de cálculos definida por el usuario, se puede obtener información del comportamiento de los parámetros con un costo computacional considerado.

Este algoritmo es probado con distintos modelos de benchmark, además de ser utilizado como algoritmo de entrenamiento para una red neuronal. Una característica importante de la red neuronal utilizada, es que se utilizan métodos para identificar patrones en series de tiempo no lineales, con esto se obtiene una arquitectura de red neuronal más eficiente de lo que se podría obtener si no se utilizan estos métodos.

## 6.1. TRABAJO FUTURO

Como se ha mostrado, existe un gran número de aplicaciones para estos modelos. En el caso de series de tiempo con relaciones lineales, el utilizar un modelo con solamente un parámetro tiene una gran área de aplicaciones en donde la capacidad de cómputo es limitada, como lo es la electrónica.

Para el caso de series de tiempo no lineales, se presentó una herramienta capaz de obtener una expresión para la función de densidad asociada con el problema de optimización, esto abre una gran variedad de aplicaciones y variaciones del algoritmo. Algunas de las posibles aplicaciones son: creación de una heurística basada en este algoritmo utilizando probabilidades para determinar la localización del punto óptimo, utilizarla como generación de puntos iniciales en distintos algoritmos del tipo poblacional. Además de que es necesario agregar al algoritmo la capacidad de trabajar con problemas de optimización con restricciones, no continuos y con variables enteras.

---

## BIBLIOGRAFÍA

---

- [1] BAGNOLI, F., BERRONES, A. Y FRANCI, F. Degustibus disputandum (forecasting opinions by knowledge networks). *Physica* 332 (2004), 509–518.
- [2] BERRONES, A. Filtering by sparsely connected networks under the presence of strong additive noise. *Proc. Seventh Mexican International Conference on Computer Science* (2006).
- [3] Berrones, A., Peña, D. Identification of Relevant Frequencies and Forecasting of Short Time Series Under Strong Additive Noise. *En preparación*.
- [4] Berrones, A., Peña, D. y Sanchez, R. Stationary Density of Stochastic Search Process. *Encyclopedia of Artificial Intelligence* 2008.
- [5] BOX, G. E. P. Y JENKINS, G. M. *Time Series Analysis: Forecasting and Control*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 1994.
- [6] GEMAN, S. Y GEMAN, D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6 (1984), 721–741.
- [7] GRASMAN, J. Y VAN HERWAARDEN, O. A. *Asymptotic Methods for the Fokker-Planck Equation and the Exit Problem in Applications*. Springer, Berlin, 1999.
- [8] HAYKIN, S. *Neural Networks: A Comprehensive Foundation*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 1994.

- 
- [9] HAYKIN, S. Y VAN VEEN, B. *Señales y sistemas*. México : Limusa, 2001.
- [10] HYNDMAN, R. J. Y KOEHLER, A. B. Another look at measures of forecast accuracy. Monash Econometrics and Business Statistics Working Papers 13/05, Monash University, Department of Econometrics and Business Statistics, Mayo 2005.
- [11] KANTZ, H. Y SCHREIBER, T. *Nonlinear time series analysis*. Cambridge University Press, New York, NY, USA, 1997.
- [12] KENNEL, M. B. Y ABARBANEL, H. D. I. False neighbors and false strands: a reliable minimum embedding dimension algorithm. *Physical Review E* 66, 2 (2002), 026209.
- [13] KIRKPATRICK, S., GELATT, J. C. D. Y VECCHI, M. P. Optimization by simulated annealing. *Science* 220, 4598 (Mayo 1983), 671–680.
- [14] MAKRIDAKIS, S. G., WHEELWRIGHT, S. C. Y HYNDMAN, R. J. *Forecasting: Methods and Applications*, 3 ed. John Wiley and Sons, USA, 1998.
- [15] MASLOV, S. Y SNEPPEN, K. Specificity and stability in topology of protein networks. *Science* 296, 5569 (Mayo 2002), 910–913.
- [16] MASLOV, S. Y ZHANG, Y.-C. Extracting hidden information from knowledge networks. *Physical Review Letters* 87, 248701 (2001).
- [17] MASTERS, T. *Practical neural network recipes in C++*. Academic Press Professional, Inc., San Diego, CA, USA, 1993.
- [18] PARDALOS, P. M. Y FABIO, S. Recent advances and trends in global optimization: Deterministic and stochastic methods. In *Proceedings of the Sixth International Conference on Foundations of Computer–Aided Process Design* (2004), vol. 1, pp. 119 – 131.

- [19] PARSOPOULOS, K. E. Y VRAHATIS, M. Recent approaches to global optimization problems through particle swarm optimization. *Natural Computing* 1 (2002), 235–206.
- [20] PEÑA, D., SÁNCHEZ, R. Y BERRONES, A. Stationary Fokker–Planck Learning for the Optimization of Parameters in Nonlinear Models. MICAI 2007: 94–104
- [21] PRESS, W. H., VETTERLING, W. T., TEUKOLSKY, S. A. Y FLANNERY, B. P. *Numerical Recipes in C++: the art of scientific computing*. 2002.
- [22] RISKEN, H. *The Fokker-Planck equation: methods of solution and applications*. Segundo, Springer-Verlag, 1989.
- [23] SCHWEITZER, F. *Brownian Agents and Active Particles: Collective Dynamics in the Natural and Social Sciences*. SpringerVerlag, 2003.
- [24] SUYKENS, J. A., VANDEWALLE, J. Y MOOR, B. L. R. D. *Artificial Neural Networks for Modelling and Control of Non-Linear Systems*. Kluwer Academic Publishers, Norwell, MA, USA, 1996.
- [25] SUYKENS, J. A., VERRELST, H. Y VANDEWALLE, J. On-line learning fokker-planck machine. *Neural Processing Letters* 7, 2 (1998), 81–89.
- [26] VAN KAMPEN, N. G. *Stochastic Processes in Physics and Chemistry*. North-Holland, Enero 1981.
- [27] VERHULST, P. F. Notice sur la loi que la population poursuit dans son accroissement. *Correspondance mathématique et physique*, 10 (1838), 113–121.
- [28] WIENER, N. *Cibernética o el control y comunicación en animales y máquinas*, 2 ed. Tusquets Editores, Constitución, 19 - 08014 Barcelona, 1998.

---

# ÍNDICE DE FIGURAS

---

2.1. Distintos patrones encontrados en una serie de tiempo . . . . .	12
2.2. La gráfica estacional permite visualizar claramente los patrones estacionales, en este caso se ve como los periodos 9 y 15 presentan incrementos en ambas gráficas . . . . .	13
2.3. La gráfica de dispersión muestra la relación entre dos factores . . . . .	13
2.4. El correlograma ayuda visualizar la función de autocorrelación de una manera rápida y sencilla. . . . .	17
2.5. Ejemplos de series de tiempo lineales con ruido y no lineales . . . . .	41
2.6. Distintas representaciones para una serie de tiempo . . . . .	48
2.7. La gráfica de recurrencia permite visualizar los valores redundantes en la serie de tiempo . . . . .	49
5.1. Pronóstico para la combinación lineal de diez funciones coseno sin ruido . . . . .	79
5.2. Pronóstico para el mapa logístico con parametro $r = 3,6$ sin ruido. . .	82

- 
- 5.3. Resultado del pronóstico para una serie de tiempo generada mediante un modelo AR de orden 23 y con intensidad de ruido de 0,1 . . . . 83
- 5.4. Pronostico de los modelos AR y KNNR para el producto 1 de los datos experimentales . . . . . 84
- 5.5. Densidades de probabilidad generadas por 10 iteraciones de SFP para la función de Levy No. 5. Los parámetros de SFP son  $L = 50$  y  $D = 200$ . El óptimo global se encuentra en la región de máxima probabilidad. 89
- 5.6. Gráfica del MSE en el horizonte  $h$  obtenido al pronosticar datos por una red neuronal entrenada mediante SFP y el software NEURAL . . 93

---

# ÍNDICE DE CUADROS

---

5.1. Resultados del pronóstico con los modelos AR y KNNR para series de tiempo generadas por combinación lineal de cosenos con diferentes horizontes H. . . . .	81
5.2. Resultados del pronóstico con los modelos AR y KNNR, para el mapa logístico. . . . .	81
5.3. Resultados del pronóstico mediante los modelos AR y KNNR para series de tiempo generadas mediante el modelo AR. . . . .	85
5.4. Resultados del pronóstico para los modelo AR y KNNR con datos de la vida real. . . . .	85
5.5. Tasa de éxitos para el problema de Levy No. 5, utilizando el punto con máxima probabilidad como punto inicial para el algoritmo de Powell. Cada realización consiste de 100 ejecuciones del algoritmo de Powell. Se considera éxito si el valor objetivo difiere en 0,0001 con el óptimo global. . . . .	89



<p>5.6. Tasa de éxitos para el problema de Levy No. 5, utilizando puntos generados de una distribución uniforme como puntos iniciales para el algoritmo de Powell. Cada realización consiste de 100 ejecuciones del algoritmo de Powell. Se considera éxito si el valor objetivo difiere en 0,0001 con el óptimo global. . . . .</p>	<p>90</p>
<p>5.7. Tasa de éxitos para el problema XOR, utilizando puntos iniciales para el algoritmo de Powell generados a partir de la función de densidad obtenida mediante SFP. Cada realización consiste de 100 ejecuciones del algoritmo de Powell, cada una de ellas con una semilla distinta. .</p>	<p>91</p>
<p>5.8. Tasa de éxitos para el problema XOR, utilizando puntos iniciales para el algoritmo de Powell generados a partir una distribución uniforme. Cada realización consiste de 100 ejecuciones del algoritmo de Powell, cada una de ellas con una semilla distinta. . . . .</p>	<p>92</p>

---

## LISTA DE SÍMBOLOS

---

$\bar{X}$	Media de la muestra.
$S^2$	Varianza de la muestra.
$S$	Desviación estándar de la muestra.
$\text{Cov}_{XY}$	Covarianza de las variables $X, Y$ .
$r_{XY}$	Correlación de las variables $X, Y$ .
$c_k$	Autocovarianza para el retardo $k$ .
$r_k$	Autocorrelación para el retardo $k$ .
$y_t$	Observación al tiempo $t$ .
$f_t$	Pronóstico al tiempo $t$ .
$e_t$	Error de pronóstico al tiempo $t$ .

---

## LISTA DE ABREVIACIONES

---

MAD	Desviación promedio absoluta.
MSD	Desviación cuadrada media.
ACF	Función de autocorrelación.
ME	Error medio.
MSE	Error cuadrado medio.
MAE	Error absoluto promedio.
MAPE	Error porcentual absoluto promedio.
sMAPE	Error porcentual absoluto promedio simétrico.
MASE	Error escalado absoluto promedio.
AR	Autoregresivo.
ARMA	Autoregresivo con media móvil.
MA	Media móvil.
KNNR	Redes de conocimiento para reducción de ruido.
SFP	Fokker-Planck estacionario.

---

# AUTOBIOGRAFÍA

---

Dexmont Alejandro Peña Carrillo

Candidato para el grado de Maestro en Ciencias en Ingeniería de Sistemas

Universidad Autónoma de Nuevo León

Facultad de Ingeniería Mecánica y Eléctrica

Tesis

NUEVAS TÉCNICAS DE PROCESOS ESTOCÁSTICOS Y APRENDIZAJE ESTADÍSTICO

PARA SERIES DE TIEMPO

Nació en Monterrey, Nuevo León. Primer hijo de Marcial Peña Ovalle y Ma. Concepción Carrillo Muñoz. Graduado de la Universidad Autónoma de Nuevo León en la carrera de Lic. en Ciencias Computacionales (2000 - 2004). Laboró dos años en el área de consultoría y desarrollo de software para distintas firmas de la localidad y extranjeras. Posteriormente inició sus estudios de Maestría en Ciencias en Ingeniería de Sistemas en enero del 2006.