

ANALISIS ESPECTRAL UNIVARIADO

WILLIAM CASTILLO ELIZONDO

**SERIE: Trabajos de Metodología #1
Centro de Investigaciones Históricas
Universidad de Costa Rica**

519.5

C352a Castillo Elizondo, William.

Análisis espectral univariado / William Castillo Elizondo. — San José, C. R. : Oficina de Publicaciones de la Universidad de Costa Rica, 1990.

p. : il.

ISBN 9977-917-92-2

1. Análisis espectral. I. Título.

CCC/BUCR-215



PRESENTACION

El análisis de datos cuantitativos es un campo complejo y en expansión que requiere, como pocos, del trabajo interdisciplinario. Matemáticas, estadística y computación se conjugan así al servicio de otras disciplinas, poniendo a disposición de los investigadores un valioso repertorio de herramientas listas para el estudio aplicado.

Hay, sin embargo, muchos obstáculos que se interponen entre los métodos disponibles y sus posibles usuarios. El primero de todos es lo que podríamos llamar la barrera de la especialidad. Siempre se necesita de un conocimiento teórico mínimo sobre cada método pero escasean los textos adecuados para adquirirlo. La teoría estadística y matemática que fundamenta técnicas como el análisis espectral, el análisis de correspondencia o los métodos de Box y Jenkins, -para indicar sólo algunos ejemplos-, es lo suficientemente compleja y difícil como para que sea casi inaccesible a los que no son especialistas en la materia.

El segundo obstáculo es el dilema de la elección. ¿Cuál método es el más apropiado para el problema concreto que nos interesa resolver? En muchos casos tenemos, al menos como punto de partida, una pluralidad de opciones posibles, y es necesario poder definir ciertos criterios precisos para efectuar una elección apropiada. El tercer y último problema se refiere al dilema del "software". De todos los "paquetes" y programas disponibles, ¿cuál es el más adecuado para nuestras necesidades e intereses de investigación?

La serie de publicaciones que inauguramos con el texto de William Castillo Elizondo, "Análisis espectral univariado", se orienta precisamente a responder a estas inquietudes e intenta salvar estos obstáculos. Aunque concebida dentro del programa de investigaciones sobre "Cuantificación e historia interdisciplinaria", y en un marco que es el del uso de la cuantificación en las Ciencias Sociales, esta publicación, al igual que las que vendrán en un próximo futuro, también serán de utilidad a todos los interesados en el análisis de datos cuantitativos sin más.

Héctor Pérez Brignoli
Coordinador del Programa: "Cuantificación
e historia interdisciplinaria"
Centro de Investigaciones Históricas

Este trabajo ha sido elaborado en el marco del Programa "Cuantificación e Historia Interdisciplinaria" del Centro de Investigaciones Históricas de la Universidad de Costa Rica.

El interés central es contribuir a la difusión del análisis espectral en tanto que método para el estudio de series temporales, particularmente en lo que concierne a los movimientos cíclicos.

Pese a la complicación teórica de la materia tratada y a las dificultades que involucra la aplicación del análisis espectral, se ha procurado dar explicaciones sencillas e intuitivas sobre los conceptos y problemas abordados, recurriendo al ejemplo y a las ilustraciones gráficas. Todo ello con el fin de facilitar la lectura.

Desde luego, lo hecho representa un primer paso para el logro del objetivo planteado y mucho queda por hacer en este campo. Por ejemplo, el diseño de una base de datos y la producción de materiales orientados especialmente al análisis de series reales, es una tarea de gran valor.

Finalmente, deseo agradecer los oportunos y valiosos comentarios y sugerencias de los investigadores del Programa: Héctor Pérez B., Arodys Robles y Rolando Cuenca. Lógicamente, la responsabilidad por el contenido del trabajo, es enteramente del autor.

El autor.

SIMBOLOS MAS UTILIZADOS

	pag
(x_i) : serie temporal cuyos términos son x_1, \dots, x_n	2
$\{X_t\}$: proceso estocástico con índice discreto t	3
$\{e_t\}$: proceso puramente aleatorio (ruido blanco)	5
μ_n : estimador de la media μ , de un proceso aleatorio ...	10
$\hat{\mu}$: media de la serie (x_i)	11
$S^2(n)$: estimador de la varianza de un proceso aleatorio ...	11
S^2 : varianza de la serie (x_i)	11
$cor(h)$: correlación entre las variables X_t y X_{t+h}	11
$Aut(h)$: autocorrelación de la serie (x_i)	11
h_k : amplitud de los componentes cíclicos en la serie	24
$I_n(w_k)$: periodograma de la serie (x_i)	25
$f_n(w_k)$: periodograma ponderado de la serie (x_i)	33
$L(j)$: ventana	33
$L(j)$: ventana rectangular	34
$L(j)$: ventana triangular	34
$m = 2d + 1$: ancho de la ventana	34

CONTENIDO

1. INTRODUCCION	1
2. SERIES TEMPORALES Y PROCESOS ESTOCASTICOS	2
2.1 Serie temporal	2
2.2 Proceso estocástico	2
2.2.1 Líneas de espera	2
2.2.2 Procesos de pruebas independientes	3
3. ESTACIONARIDAD	5
3.1 Procesos puramente aleatorios	5
3.2 Procesos estocásticos estacionarios	7
3.2.1 Procesos autorregresivos	7
3.2.2 Procesos de medias móviles	9
3.2.3 Procesos autorregresivos de medias móviles	9
3.2.4 Procesos lineales generados por un ruido blanco	9
3.3 Series estacionarias	10
3.4 Estimadores	10
4. SERIES NO ESTACIONARIAS	13
4.1 Tendencia en la media	13
4.2 Eliminación de la tendencia	15
4.2.1 Método de las diferencias	15
4.2.2 Método de mínimos cuadrados	18
4.3 Otros métodos y tendencias	20
4.4 Estabilización de la varianza	20
5. EL PERIODOGRAMA	23
5.1 Representación trigonométrica de una serie	23
5.2 Descomposición de la varianza	25
5.3 Concepto de periodograma	25
5.4 Forma compleja del periodograma	28
5.5 Una prueba estadística	30
6. ESTIMACION DEL ESPECTRO	33
6.1 El periodograma ponderado	33
6.2 Concepto de ventana	33
6.3 Ancho de la ventana	34
6.4 Bandas de confianza aproximadas	36

7. "SOFTWARE" UTILIZADO	39
7.1 El paquete RATS	39
7.2 El paquete STATGRAPHICS	44
8. COMENTARIO FINAL	45
ANEXO I: DEFINICIONES	46
ANEXO II: RESULTADOS	49
ANEXO III: TABLAS	52
BIBLIOGRAFIA BASICA	57

1. INTRODUCCION

El desarrollo teórico del análisis espectral se fundamenta en métodos y conocimientos estadístico - matemáticos que no siempre están al alcance de las personas e instituciones que eventualmente podrían usarlo para el tratamiento de sus datos. Obviamente este es un factor que entorpece la divulgación y aceptación del análisis espectral, como un método apropiado para el estudio de series temporales.

En razón de lo anterior y por la importancia del tema, nos pareció pertinente escribir un documento que sirva de consulta para el mayor número de usuarios del análisis espectral.

Con ese propósito, hemos simplificado la exposición, evitando en lo posible, el formalismo matemático. Asimismo, enfatizamos más los aspectos metodológicos que la justificación y explicación teórica del método.

De modo que, el principal interés se orienta al estudio, no formal, de la estacionaridad y de los métodos de cálculo y usos del periodograma, y en forma breve, un procedimiento de estimación del espectro. Estos temas son las partes integrantes fundamentales de un análisis espectral univariado.

En la sección 2 se introducen los conceptos de serie temporal y proceso estocástico.

La idea de proceso estocástico estacionario y serie temporal estacionaria, se discute en la sección 3. Particular interés tienen los procesos lineales estacionarios generados por un ruido blanco. La sección 4 trata algunos métodos de estacionarización de series temporales.

El concepto de periodograma y su expresión compleja, se desarrollan en la sección 5. Se precisa el procedimiento para calcularlo y se ilustra su significado. Para ello usamos la serie de bautizos mensuales de Cartago, correspondiente al periodo que va de enero de 1800 a diciembre de 1814.

En la sección 6 se estudia un método de estimación del espectro y se aplica a la serie anterior.

La sección 7 trata sobre el "software" utilizado.

Los anexos I y II contienen respectivamente, definiciones y resultados básicos del análisis espectral.

En el anexo III se han organizado, en forma de tablas, las series temporales usadas en el texto, así como otras series generadas a partir de ellas.

2. SERIES TEMPORALES Y PROCESOS ESTOCÁSTICOS

2.1 Serie temporal

Una serie temporal es cualquier secuencia finita de mediciones de una variable, realizadas a iguales intervalos de tiempo (horas, días, meses, etc.). Por ejemplo, el desempleo mensual en Costa Rica durante el período 1970 - 1979, es una serie temporal que consta de 120 mediciones (u observaciones), de la variable "desempleo mensual en Costa Rica".

Algunas veces usaremos la palabra "serie" para referirnos a una serie temporal. Una serie será denotada por x_1, \dots, x_n o, abreviadamente por (x_i) .

Por lo general los problemas o fenómenos a investigar, son influenciados por factores desconocidos o por variables no observadas. Así por ejemplo, el desempleo será afectado por el grado de desarrollo de las organizaciones sindicales, las políticas económicas del gobierno, el nivel de educación, etc.

Todos los factores que inciden en el fenómeno y de los cuales no se tiene información suficiente o son del todo desconocidos, determinan el carácter aleatorio del mismo. De aquí deriva la necesidad de considerar cada término de la serie como el resultado de la acción de múltiples factores, algunos de ellos no determinados; y por tanto, como el valor asumido por una variable aleatoria, con una ley de probabilidad asociada. Este enfoque del problema conduce al tratamiento estadístico de las series temporales, concebidas como el resultado de la observación de un proceso estocástico, durante un período de tiempo.

2.2 Proceso estocástico

Con el fin de facilitar la comprensión de este concepto, se presentan dos ejemplos: Líneas de espera y el proceso de pruebas independientes. Luego se resume la idea de proceso estocástico, señalando sus elementos más importantes.

2.2.1 Líneas de espera

Las líneas de espera (filas o colas), constituyen un fenómeno cotidiano que se presenta de muy diversas formas: llamadas telefónicas originadas en una localidad, accidentes de tránsito en una intersección, clientes ingresando a una tienda, llegadas de barcos a un muelle, etc. En síntesis, una línea de espera es un sistema constituido por uno o varios canales de servicio como ventanillas, muelles, circuitos telefónicos, etc., y por la llegada de unidades conforme a ciertos intervalos de tiempo, al centro de servicio (lugar donde están los canales de servicio).

Si definimos X_t = número de unidades que llegan al centro de servicio durante el intervalo de tiempo $[0, t]$, entonces tenemos un primer ejemplo de proceso estocástico formado por la secuencia de variables aleatorias $\{X_t\}$ cuyo índice continuo t varía en el

intervalo $[0, +\infty[$. La ley de probabilidad asociada a cada X_t es particular a cada caso, pero es usual hacer hipótesis simplificadas (adecuadas para muchas situaciones reales), de acuerdo con las cuales es posible deducir formalmente que X_t sigue la ley de Poisson. De acuerdo con esta ley la probabilidad de que lleguen n unidades al centro de servicio desde el instante 0 hasta el instante t , es

$$P_n(t) = (\tau t)^n e^{-\tau t} \div n! \quad \text{donde } n = 0, 1, 2, \dots$$

En esta fórmula, τ es la tasa media de llegadas, la cual se estima a partir de una muestra previa. Por otra parte, $n!$ se lee "ene factorial" y se define por $n! = 1 \cdot 2 \cdot \dots \cdot n$. Este proceso se llama proceso de Poisson.

2.2.2 Proceso de pruebas independientes

Supongamos que es posible repetir un experimento cualquier número de veces, en las mismas condiciones. Para mayor simplificación supongamos además que el experimento sólo tiene dos resultados posibles: $E = \text{éxito}$ y $F = \text{fracaso}$.

Sea $P(E) = p$ y $P(F) = q$ las probabilidades de éxito y fracaso respectivamente, tales que $p + q = 1$. Si definimos $X_n = \text{número de éxitos obtenidos tras } n \text{ repeticiones del experimento}$, entonces X_n es una variable aleatoria cuya ley de probabilidad f se define como:

$$f(x) = P(X_n = x) = [(n!) \div x!(n-x)!] p^x q^{n-x}.$$

El valor $f(x)$ expresa la probabilidad de alcanzar x éxitos después de n ensayos. De esta manera queda definido un proceso estocástico $\{X_n\}$ cuya ley de probabilidad f , se conoce con el nombre de ley binomial. El número n es el índice del proceso e indica el número de veces que se repitió el experimento. Este proceso es de índice discreto.

En los ejemplos anteriores hemos visto que la idea de proceso estocástico surge ante la necesidad de analizar los cambios de estado de un fenómeno o sistema. La evolución del fenómeno se supone gobernada por ciertas relaciones de naturaleza probabilística especificadas en cada caso particular. De modo que la definición de un proceso estocástico consiste en precisar una secuencia de variables aleatorias cuantificadoras de los cambios de estado, así como la indicación de una ley de probabilidad asociada a estas variables.

Generalmente usamos la palabra "proceso" para referirnos a un proceso estocástico. Un proceso será denotado por $\{X_t\}$, donde por lo general se asumirá que el subíndice " t " varía en el conjunto de los números enteros.

Una serie temporal, recordésmolo, es considerada como una secuencia de valores probables, x_1, \dots, x_n , asumidos por las variables aleatorias X_1, \dots, X_n del proceso $\{X_t\}$. Más aún, la serie ocurre con una probabilidad dada por la distribución conjunta de

las variables X_1, \dots, X_n . En ese sentido, usaremos expresiones como "el proceso asociado a la serie" o "la serie generada por el proceso".

3. ESTACIONARIDAD

En el plano teórico, la hipótesis básica del análisis espectral reside en el hecho de que el proceso estocástico sea estacionario. Por eso, y por las implicaciones que dicha hipótesis tiene a nivel del análisis de series temporales, se aborda en esta sección el estudio de la estacionaridad en relación con los procesos estocásticos y las series temporales. Primero se introduce en 3.1 la idea de proceso estacionario en su versión más simple: los procesos puramente aleatorios. Luego se amplía a procesos estocásticos más generales (ver 3.2).

Finalmente, en 3.3 y 3.4 se plantean algunas cuestiones en relación con las series estacionarias.

En esta sección y en el resto del texto, se hace referencia a conceptos como esperanza matemática, varianza, covarianza, etc. Siempre se supone que tales valores existen. Las definiciones se pueden consultar en el anexo I

3.1 Procesos puramente aleatorios

Los procesos puramente aleatorios, pese a su extrema simplicidad, constituyen la pieza principal sobre la cual se edifica la teoría de modelos de series temporales y la teoría del análisis espectral. Constituyen el ejemplo más simplificado de proceso estacionario.

Un proceso estocástico $\{e_t\}$ es un proceso puramente aleatorio si satisface las siguientes propiedades:

- a. Todos los términos e_t del proceso tienen igual esperanza matemática. Su valor común es la media del proceso y se denota con la letra μ . Este, siempre se supondrá que vale cero.
- b. Todas las variables aleatorias e_t tienen igual varianza. Esta, se denota con σ^2 , y es la varianza del proceso.
- c. Dos variables aleatorias cualesquiera son no correlacionadas.

A los procesos puramente aleatorios también se les conoce con el nombre de ruido blanco. La razón es que estos procesos, debido a sus propiedades especiales, se los puede concebir como la "suma" lesde una infinitud de componentes sinusoidales (cíclicos), todos ellos aportando por igual a la varianza del proceso. En ese sentido se dice que todas las frecuencias (una por cada componente cíclico), son igualmente importantes. La luz blanca tiene la propiedad análoga, toda vez que en ella los colores aparecen en igual canlestidad (todas las frecuencias son igualmente importantes). De esta analogía se deriva el nombre de ruido blanco con el que también se designa a los procesos puramente aleatorios.

Un proceso puramente aleatorio, como su nombre lo indica, tiene un comportamiento completamente errático, no obedece a ningún patrón regular de variación. Para distinguir mejor sus propieda-

des se puede usar el recurso de la simulación. Luego de la definición de procesos estacionarios, se explicará como se realizan las simulaciones de procesos lineales, y en la sección 7 sobre software se presenta un programa RATS para hacer los cálculos.

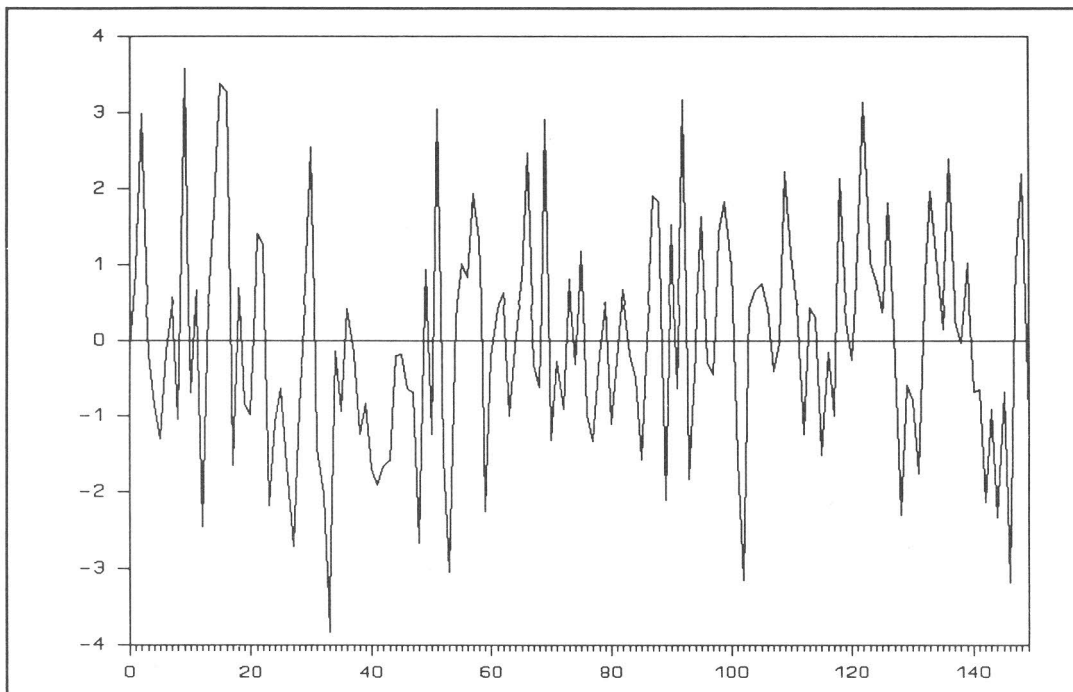


GRAFICO 1. Serie de 150 puntos generados por un proceso del tipo ruido blanco, normal con varianza igual a 2

El GRAFICO 1, corresponde a una simulación de 150 puntos, de un ruido blanco normal $N(0,2)$. Es decir, a cada variable aleatoria e_t , se le asigna una ley de probabilidad normal con media igual a cero y varianza igual a 2. Notemos la coherencia de las propiedades del ruido blanco con las características observadas en el gráfico. En efecto, la forma del gráfico está determinada por variaciones muy irregulares alrededor de una línea horizontal que pasa por cero. Lo cual se asocia con la hipótesis de no correlación de las variables y media igual a cero. Por otra parte, como los valores fueron generados con una distribución normal $N(0,2)$, las ordenadas en el GRAFICO 1 varían en un 95%, entre $-2/\sqrt{2}$ y $2/\sqrt{2}$. Es decir, el 95% de los puntos se encuentran en una franja horizontal centrada en cero, cuyo ancho aproximado es $4/\sqrt{2}$.

3.2 Procesos estocásticos estacionarios

Luego de haber estudiado el concepto de estacionaridad en su forma más simplificada (el ruido blanco), conviene resumir intuitivamente este concepto y luego dar una definición de proceso estacionario. Como vimos en el caso del ruido blanco, la estacionaridad expresa una especie de equilibrio temporal del proceso. Este, varía siempre cerca de un nivel promedio constante. Las fluctuaciones aleatorias alrededor de ese nivel, obedecen a una ley de densidad conjunta invariable en el tiempo, en un sentido que se precisa en el Anexo I, definiciones 10 y 11.

Un proceso estocástico $\{X_t\}$ es estacionario¹ si verifica las siguientes propiedades:

- a. Todas las variables X_t tienen igual esperanza matemática. Su valor común es la media del proceso y se denota por μ .
- b. Todas las variables X_t tienen igual varianza. Este valor es la varianza del proceso y se denota por σ^2 .
- c. La covarianza entre dos variables cualesquiera X_k y X_{k+h} sólo depende de la "separación" entre ellas. Es decir, de h .

Notemos que la diferencia de esta definición con la de ruido blanco reside en la propiedad c. En efecto, la estructura de covarianzas en un proceso estacionario es más general que en el caso del ruido blanco.

Revisemos brevemente algunos procesos estacionarios muy usados en teoría de modelos lineales de series temporales, llamados modelos de Box - Jenkins.

3.2.1 Procesos autorregresivos

Un proceso $\{X_t\}$ es un proceso autorregresivo de orden p con coeficientes a_1, \dots, a_p , si

$$(1) \quad X_t = \sum_{k=1}^p a_k X_{t-k} + e_t$$

donde $a_p \neq 0$ y $\{e_t\}$ es un ruido blanco. Supongamos por ejemplo que X_t representa la producción anual de un producto. Según la ecuación (1), la producción actual será una suma ponderada de la producción de los "p" años anteriores, salvo por un término de "error" tipo

¹ En el Anexo 1 se indica que el concepto de proceso estacionario (def. 11.), se deduce de la definición de proceso estrictamente estacionario (def. 10.). En ese sentido se acostumbra usar en lugar del adjetivo estacionario, otros nombres como, débilmente estacionario, estacionario de segundo orden, estacionario en sentido amplio, y de covarianza estacionaria.

ruido blanco. De una manera general, la ecuación (1), expresa que cada variable X_t depende linealmente de las p variables anteriores, X_{t-1}, \dots, X_{t-p} , más un término de "error", ruido blanco, asociado al instante " t ".

Bajo ciertas condiciones sobre los a_k , un proceso autorregresivo es estacionario. En efecto, a cada proceso definido por (1) se le asocia una ecuación polinomial de la forma

$$z^p - a_1 z^{p-1} - \dots - a_{p-1} z - a_p = 0.$$

Si toda solución " s " de esta ecuación es tal que $|s| < 1$, entonces el proceso es estacionario. Por tanto es posible simular procesos autorregresivos estacionarios, escogiendo un valor de " p " pequeño y valores apropiados de a_1, \dots, a_p .

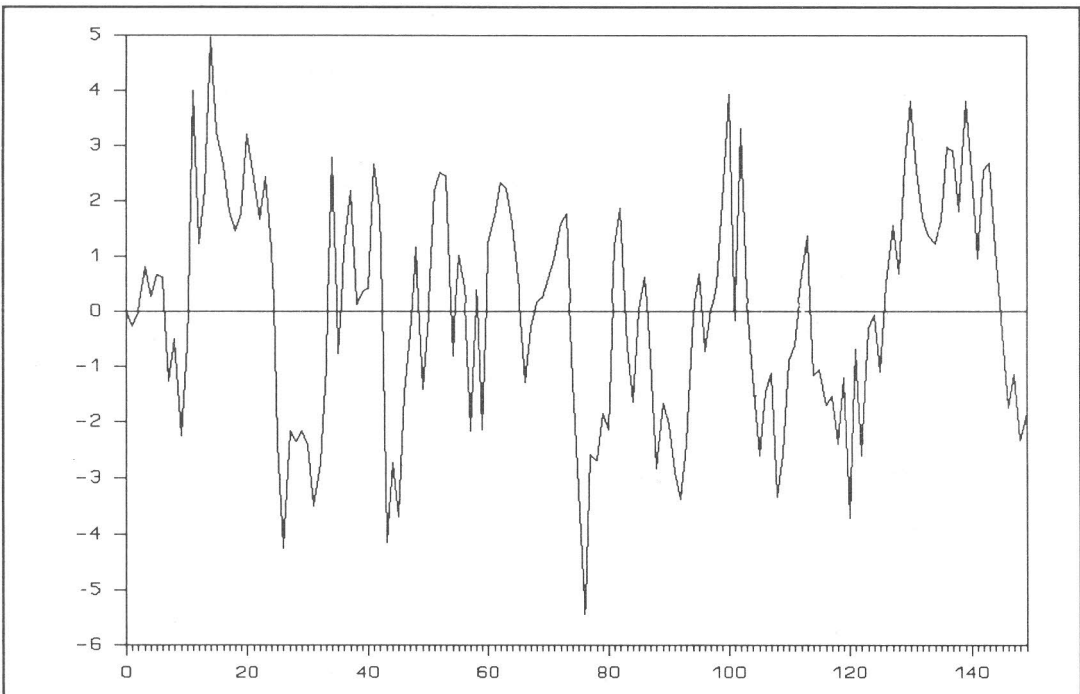


GRAFICO 2. Serie estacionaria de 150 puntos, generada con la ecuación $x_t = 0.6x_{t-1} + e_t$ donde $\{e_t\}$ es un ruido blanco normal con varianza igual a 2.

Ejemplo

El GRAFICO 2, consta de 150 puntos, generados con la ecuación

$$x_t = 0.6 \cdot x_{t-1} + e_t$$

donde $\{e_t\}$ es un ruido blanco normal $N(0,4)$. La ecuación polinomial asociada al proceso es: $z - 0.6 = 0$. Su solución $s = 0.6$ es menor que 1. Luego el proceso autorregresivo es estacionario.

Un ruido blanco normal $N(0, \sigma^2)$, se puede simular indicando un valor de σ^2 , y definiendo $a_1 = \dots = a_p = 0$ para algún p . Por ejemplo, $p = 1$ (ver sección 7, programa 4).

3.2.2 Procesos de medias móviles

Un proceso estocástico $\{X_t\}$ es un proceso de medias móviles de orden q con coeficientes b_1, \dots, b_q , si

$$X_t = \sum_{k=1}^q b_k e_{t-k} + e_t$$

donde $\{e_t\}$ es un ruido blanco y $b_q \neq 0$.

Estos procesos son siempre estacionarios. Igual que los procesos autorregresivos, se pueden simular indicando el valor de q, b_1, \dots, b_q , la media y la varianza del ruido blanco.

3.2.3 Procesos autorregresivos de medias móviles

Una generalización natural de los procesos anteriores son los procesos autorregresivos de medias móviles de orden (p, q) . Estos procesos se definen por la ecuación

$$X_t = \sum_{k=1}^p a_k X_{t-k} + \sum_{i=1}^q b_i e_{t-i} + e_t$$

donde los a_k, b_i, p, q son conocidos; $a_p \neq 0 \neq b_q$ y $\{e_t\}$ es un ruido blanco. La condición de estacionaridad es la misma que para los procesos autorregresivos.

3.2.4 Procesos lineales generados por un ruido blanco

Una clase de procesos muy importantes en la teoría de series temporales, son los procesos lineales generados por un ruido blanco. Un proceso de esta clase se define por

$$X_t = e_t + \alpha_1 e_{t-1} + \dots + \alpha_k e_{t-k} + \dots$$

donde $\{e_t\}$ es un ruido blanco. La definición de este proceso se entiende en el sentido de que cada variable X_t es un límite en la media cuadrática, es decir:

$$E[|X_t - \sum_{j=1}^n \alpha_j e_{t-j}|^2] \rightarrow 0 \quad \text{si } n \rightarrow \infty.$$

Si los α_j verifican la condición

$$(|\alpha_1| + \dots + |\alpha_n|) \rightarrow L \text{ si } n \rightarrow \infty$$

con L un número real, entonces el proceso $\{X_t\}$ es estacionario.

Es claro que estos procesos generalizan la idea de proceso de medias móviles (3.2.2). Por otra parte, se sabe que todo proceso autorregresivo estacionario se puede expresar como un proceso lineal generado por un ruido blanco. Por tanto, si los procesos definidos anteriormente (3.2.1, 3.2.2 y 3.2.3), son estacionarios, todos pueden ser considerados como procesos lineales generados por un ruido blanco.

3.3 Series estacionarias

Como vimos, una serie es la observación de un proceso estocástico, durante un periodo determinado. Si este proceso es estacionario, entonces diremos que la serie **es estacionaria**.

Ejemplos de series estacionarias generadas por simulación, son el GRAFICO 1 y el GRAFICO 2. En estos dos casos la estacionaridad deriva del hecho de que el proceso estocástico asociado es estacionario.

Sin embargo, esta definición de serie estacionaria no es operacional, puesto que, en las aplicaciones lo que conocemos es la serie y no el proceso. Entonces, ¿cómo y por cuáles medios obtener información sobre la estacionaridad de la serie? Esta pregunta nos conduce directamente a uno de los temas más importantes en relación con el problema: definir un mecanismo apropiado para estimar los parámetros que intervienen en la definición de proceso estacionario. En todo caso, cuando se trata con datos reales, la estacionaridad será alcanzada solamente como una aproximación. Esto no debe sorprender, ya que los modelos cuando son confrontados con la realidad por medio de los datos, producen descripciones solamente aproximativas del problema estudiado.

3.4 Estimadores

Concretamente hay que definir tres funciones llamadas estimadores de la media, la varianza y las autocorrelaciones del proceso. El valor de esos estimadores en la muestra (la serie), son los valores estimados de los parámetros. Bajo hipótesis apropiadas sobre el proceso, y para una serie "suficientemente" larga, los valores estimados son aproximadamente iguales a los parámetros.

Consideremos una serie temporal x_1, \dots, x_n generada por el proceso $\{X_t\}$, el cual tiene media μ y varianza σ^2 . Se definen los siguientes estimadores:

1. Un estimador de μ es

$$\mu_n = (1/n) \sum_{t=1}^n X_t$$

La media \hat{u} de la serie, es el resultado de evaluar μ_n en x_1, \dots, x_n :

$$\hat{u} = (1/n) \sum_{i=1}^n x_i$$

2. Un estimador de σ^2 es

$$S^2(n) = (1/n) \sum_{t=1}^n (X_t - \mu_n)^2$$

Similarmente, la varianza S^2 de la serie, se obtiene como resultado de evaluar $S^2(n)$ en x_1, \dots, x_n :

$$S^2 = (1/n) \sum_{i=1}^n (x_i - \hat{u})^2$$

3. Un estimador de la correlación entre dos variables X_t y X_{t+h} es

$$\text{cor}(h) = [1 \div nS^2(n)] \sum_{t=1}^{n-h} (X_t - \mu_n)(X_{t+h} - \mu_n)$$

Evaluando $\text{cor}(h)$ en la serie x_1, \dots, x_n se obtiene la h -ésima autocorrelación:

$$\text{Aut}(h) = (1 \div nS^2) \sum_{t=1}^{n-h} (x_t - \hat{u})(x_{t+h} - \hat{u})$$

El autocorrelograma de la serie es el gráfico de las autocorrelaciones $\text{Aut}(1), \dots, \text{Aut}(r)$, construido en la forma usual: sobre el eje horizontal (eje X) se colocan los valores $1, 2, \dots, r$. Sobre el eje vertical (eje Y) se colocan los correspondientes valores $\text{Aut}(1), \text{Aut}(2), \dots, \text{Aut}(r)$. Luego los puntos $(1, \text{Aut}(1)), \dots, (r, \text{Aut}(r))$ se unen por medio de segmentos.

El autocorrelograma ha sido usado en teoría de modelos lineales, como un indicador de la no estacionaridad de la serie (ver 4.2.1, C.), y para especificar el orden tentativo de un proceso de medias móviles, pero este tema no será objeto de discusión.

Ejemplo

Un ejemplo de una serie temporal de datos reales es, la serie de bautizos mensuales de Cartago correspondiente al período que va de enero de 1800 a dic. de 1814. Sus valores se encuentran en la tabla 1, anexo III. El GRAFICO 3 corresponde a esta serie y presenta características cercanas a la estacionaridad: observando el gráfico notamos que la serie varía siempre alrededor de un nivel fijo, y la varianza parece constante a lo largo de todo el período.

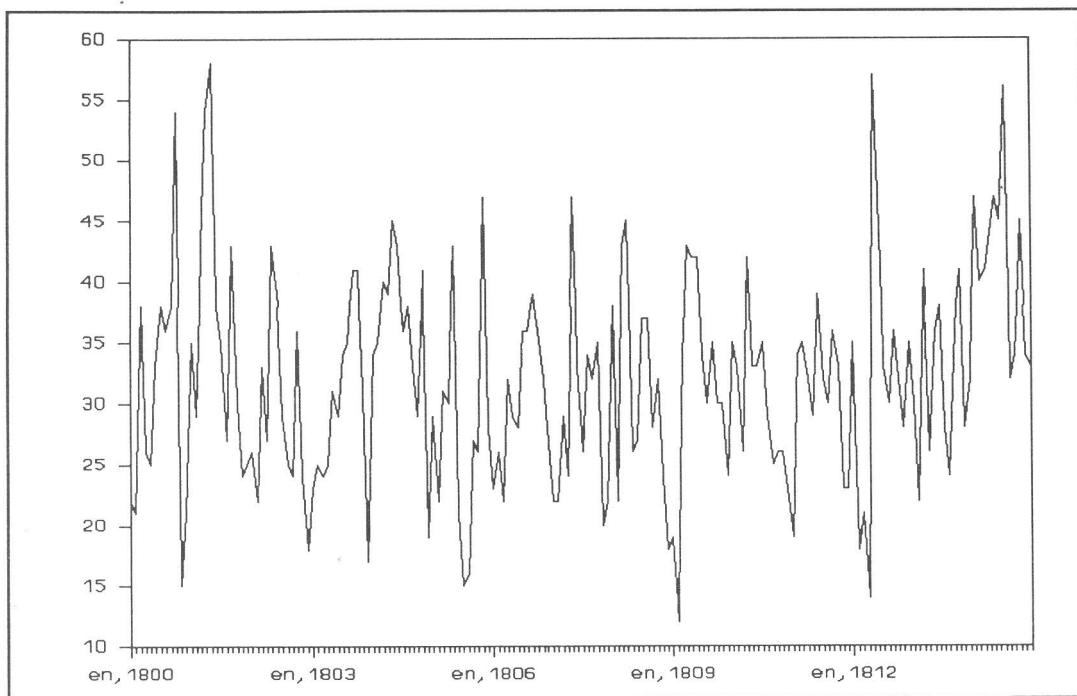


GRAFICO 3. Bautizos mensuales de Cartago,
enero 1800 - diciembre 1814

4. SERIES NO ESTACIONARIAS

En la sección previa se estudió con algún detalle el concepto de estacionaridad a nivel de proceso estocástico y de serie temporal. Asimismo, señalamos su importancia en tanto que hipótesis básica en análisis espectral y se incluyeron algunos ejemplos a modo de ilustración. No obstante, nuestro desarrollo ha sido básicamente a un nivel teórico.

La estacionaridad en datos reales, es poco común. Procede por consecuencia, discutir la no estacionaridad a nivel de las series temporales, en relación con:

1. Los procedimientos para investigar si el proceso es o no, estacionario.
2. Algunos tratamientos usuales de la serie temporal, en caso de no estacionaridad.

Para el análisis de la serie vamos a considerar principalmente dos formas de variación no estacionaria que son al mismo tiempo las más frecuentes: la tendencia en la media y las variaciones sistemáticas en la varianza de la serie. Sin embargo, la estacionarización de la serie constituye el problema más delicado en la utilización de análisis espectral, dado que las tres condiciones exigidas no son siempre satisfechas. Por ejemplo, la media puede depender del tiempo, y si esta dependencia siendo desigual en toda la serie, se la describe con una misma fórmula (un polinomio, por ejemplo), ello introduce un sesgo. En todo caso, el sólo hecho de trabajar con los datos transformados, ya representa un problema adicional, puesto que los estimadores involucrados se afectan.

4.1 Tendencia en la media

Para aproximarnos intuitivamente a la idea de tendencia en la media, comparemos la forma de variación global de las series correspondientes al GRAFICO 3 y al GRAFICO 4. Este último presenta una tendencia al crecimiento, mientras que el primero exhibe una forma muy "horizontalizada". En el primer caso (GRAFICO 4) encontramos evidencia gráfica de una tendencia, posiblemente lineal. En el segundo caso la serie varía alrededor de un nivel fijo.

El término tendencia adquiere un significado más claro cuando se aplica algún procedimiento para removerla y/o estimarla. En ese sentido, es preferible que la forma de una tendencia en la media sea especificada en cada caso, y no intentar dar una definición general.

Aunque no existen recetas ni procedimientos contundentes para distinguir la presencia y naturaleza de una tendencia, se acostumbra combinar una exploración gráfica de la serie con métodos sencillos para removerla.

Exploración gráfica

Una práctica habitual es hacer un estudio gráfico de la serie, como una primera fase del análisis. La idea es que una exploración gráfica nos permita descubrir la presencia de tendencias simples y usuales, como rectas o polinomios de segundo grado.

Cuando se trata de series de observaciones diarias, semanales, o mensuales, muy largas, se puede definir una nueva serie sumando los términos adecuadamente, de modo que se pueda estudiar la variación temporal digamos, mensual y no diariamente. Por ejemplo para series mensuales se suman los términos de 12 en 12 hasta producir un serie anual. Este es el caso de la serie anual de bautizos de Cartago, la cual fue calculada de este modo. Como se observa en el GRAFICO 4, esta serie presenta una tendencia lineal bastante clara, la cual luego será estimada por mínimos cuadrados.

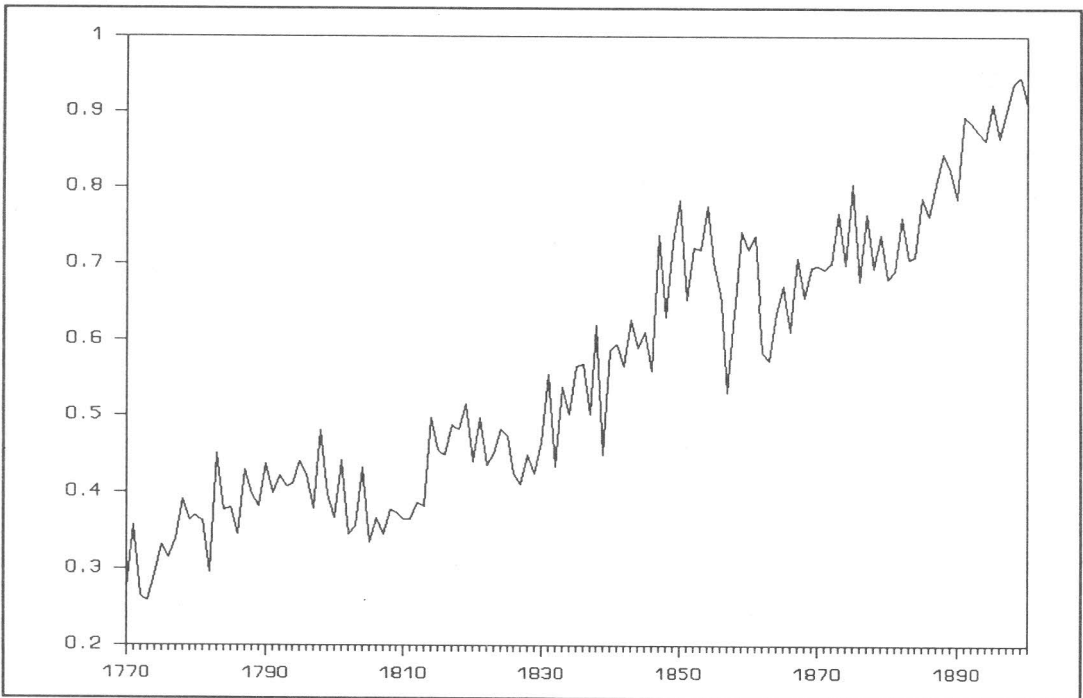


GRAFICO 4. Bautizos anuales de Cartago
(1770 - 1900)

La tendencia en la media puede cambiar su forma por tramos. Por ejemplo, una tendencia puede ser lineal en una parte de la serie, y en otra, cuadrática. En tales circunstancias es de vital importancia poder graficar la serie entera, sin ninguna transformación previa.

4.2 Eliminación de la tendencia

Cualquier procedimiento para remover una tendencia requiere la especificación de un modelo. Por ejemplo, para el caso de variables económicas, ha recibido especial atención el modelo clásico aditivo. Este modelo supone que el proceso es la suma de una tendencia T_t , un componente estacional S_t , y un proceso estacionario Z_t :

$$X_t = T_t + S_t + Z_t$$

En muchas aplicaciones la tendencia T_t es un polinomio de bajo grado, es decir

$$T_t = a_0 + a_1 t + \dots + a_p t^p, \text{ con } a_p \neq 0$$

Varios procedimientos han sido aplicados, ya sea para estabilizar la serie o para estimar la tendencia. Nos referiremos a dos de ellos: el método de las diferencias y el método de mínimos cuadrados.

4.2.1 Método de las diferencias

Este método, que es muy fácil de usar, se utiliza como una forma experimental de examinar la estacionariedad de la serie, sin estimar la tendencia. Dada la serie x_1, \dots, x_n , se definen las diferencias así:

La primera diferencia de (x_i) es la serie (z_i) definida por

$$z_i = x_{i+1} - x_i, \quad i = 1, \dots, n-1$$

La segunda diferencia de (x_i) es la serie (y_i) definida por

$$y_i = z_{i+1} - z_i, \quad i = 1, \dots, n-2$$

Así sucesivamente se pueden definir diferencias de orden mayor.

Sea $\{X_t\}$ el proceso asociado a la serie (x_i) . Algunos modelos para los cuales, por medio de la diferenciación se puede estacionarizar la serie, son los siguientes:

A. Sea $X_t = a + bt + (Y_1 + \dots + Y_t)$, donde $\{Y_t\}$ es un proceso estacionario. Entonces se tiene que la primera diferencia

$$X_{t+1} - X_t = b + Y_{t+1}$$

es un proceso estacionario. En este caso hay una tendencia polinomial de grado 1 (una recta), y la primera diferencia podría estacionarizar la serie.

Ejemplo

Calculando la primera diferencia de la serie de bautizos anuales de Cartago, se obtiene una serie (z_i) tal que

$$z_i = x_{i+1} - x_i, \quad i = 1, \dots, 130$$

Esta serie es representada por el GRAFICO 5. Cada punto en el gráfico es de la forma $(i, x_{i+1} - x_i)$, donde $i = 1, \dots, 130$. Sobre el eje X se colocan los años y sobre el eje Y, los valores de la serie $z_i = x_{i+1} - x_i$. Comparando el GRAFICO 4 de la serie original, con el GRAFICO 5, se nota el efecto de la diferenciación. La nueva serie varía en torno a un nivel constante igual a cero, lo que indica que la tendencia lineal desapareció.

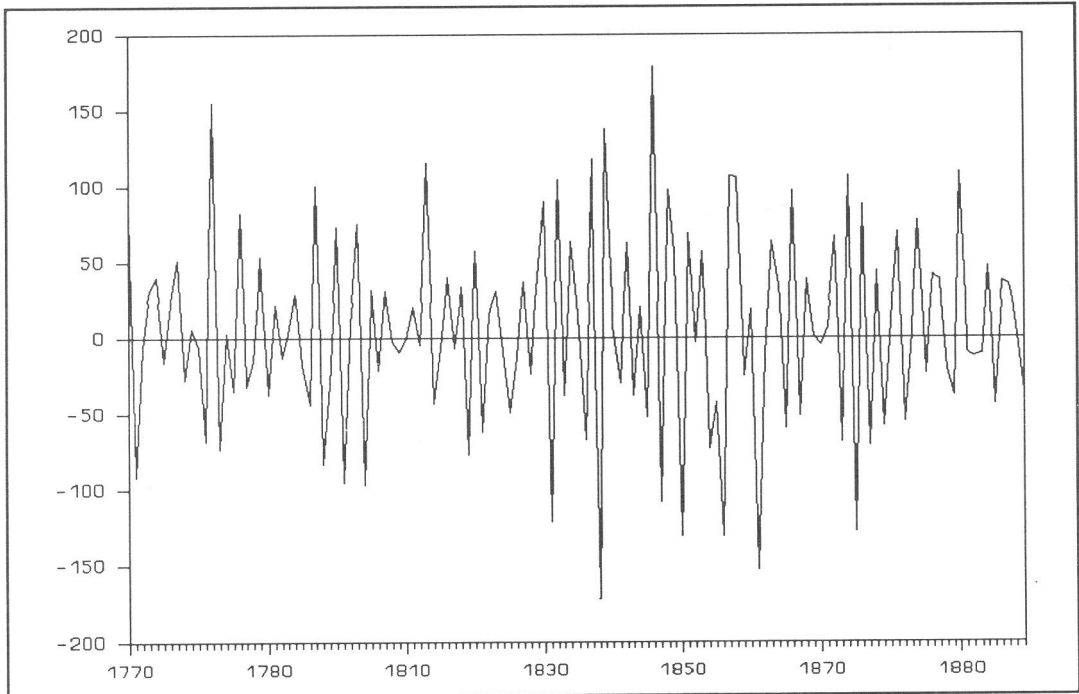


GRAFICO 5. Serie de bautizos anuales de Cartago. La tendencia fue eliminada calculando la primera diferencia.

B. Sea $X_t = a + bt + ct^2 + \sum_{r=1}^t (Y_1 + \dots + Y_r)$, donde $\{Y_t\}$ es un proceso estacionario. En este caso la serie deberá exhibir una tendencia polinomial de grado dos (forma de parábola). Calculando la segunda diferencia del proceso $\{X_t\}$ se tiene:

$$W_t = Z_{t+1} - Z_t \quad \text{donde} \quad Z_t = X_{t+1} - X_t$$

y se comprueba fácilmente que $W_t = 2c + Y_{t+2}$.

Por lo tanto la segunda diferencia podría ser suficiente para obtener una serie aproximadamente estacionaria. De acuerdo con la forma de estos modelos, es claro que se pueden generalizar a tendencias polinomiales de grado "p".

C. Si una serie temporal es generada por un proceso autorregresivo de medias móviles, entonces el autocorrelograma de la serie puede ser útil para investigar si la serie es o no, estacionaria. Además el autocorrelograma se puede usar combinadamente con el método de las diferencias, para estacionarizar la serie.

En efecto, si el proceso $\{X_t\}$ definido por

$$X_t = \sum_{k=1}^p a_k X_{t-k} + \sum_{i=1}^q b_i e_{t-i} + e_t$$

es estacionario y si todas las soluciones de su ecuación asociada

$$z^p - a_1 z^{p-1} - \dots - a_{p-1} z - a_p = 0.$$

son simples (es decir, el polinomio del lado izquierdo de la ecuación tiene exactamente p raíces distintos), entonces las correlaciones $\text{cor}(h)$ del proceso tienden rápidamente a cero, cuando h crece.

Si el proceso no es estacionario, debido a que hay soluciones reales de valor absoluto aproximadamente igual a 1, entonces los valores de $\text{cor}(h)$ tienden lentamente a cero y de manera aproximadamente lineal.

De lo dicho se desprende que si las autocorrelaciones $\text{Aut}(h)$, de la serie, se comportan como los valores $\text{cor}(h)$, entonces el autocorrelograma puede ser usado para estudiar la estacionaridad de la serie. Si la ecuación asociada tiene soluciones aproximadamente iguales a uno, entonces el método de las diferencias es frecuentemente usado para estacionarizar la serie. Si la serie es estacionaria, entonces el autocorrelograma debería tender rápidamente a cero.

Recordemos que las autocorrelaciones de la serie se calculan con la fórmula

$$\text{Aut}(r) = \left[\sum_{i=1}^{n-r} (x_i - \hat{\mu})(x_{i+r} - \hat{\mu}) \right] \div \sum_{i=1}^n (x_i - \hat{\mu})^2$$

Se acostumbra calcular entre 25 y 40 autocorrelaciones, dependiendo de la longitud de la serie. Entre más larga sea ésta, mejores son las estimaciones.

Ejemplo

Consideremos la serie de bautizos anuales de Cartago, citada previamente. Para esta serie se calcularon las 30 autocorrelaciones $\text{Aut}(1), \dots, \text{Aut}(30)$. Estos valores son representados en el GRAFICO 6. Sobre el eje horizontal se indicaron los valores 1,2, ... ,30,

y sobre el vertical, los valores de las autocorrelaciones. Nótese la forma casi lineal del autocorrelograma aproximadamente a partir del valor 2 sobre el eje horizontal.

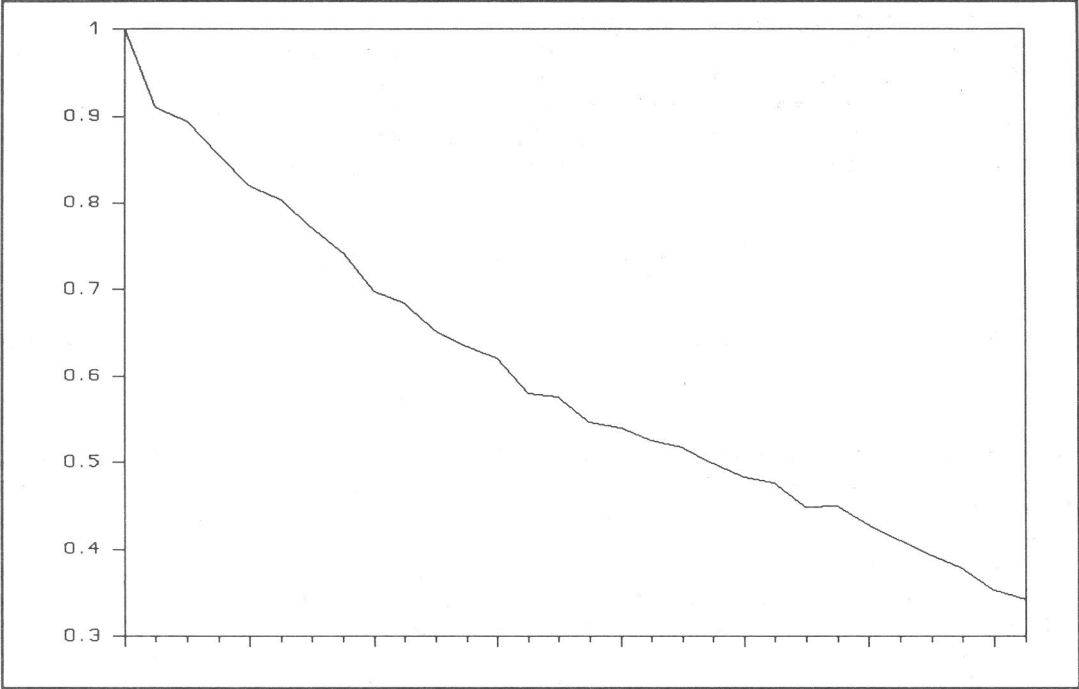


GRAFICO 6. Autocorrelograma de la serie de bautizos anuales de Cartago.

4.2.2 Método de mínimos cuadrados

Se supone que el proceso $\{X_t\}$ tiene una tendencia polinomial de grado r ; específicamente

$$X_t = \alpha_0 + \alpha_1 t + \dots + \alpha_r t^r + Y_t$$

donde $\{Y_t\}$ es un proceso estacionario con media cero. Los parámetros $\alpha_0, \dots, \alpha_r$ son estimados por mínimos cuadrados. Es decir, si x_1, \dots, x_n es la serie, entonces los valores estimados de $\alpha_0, \dots, \alpha_r$, que denotamos con a_0, \dots, a_r ; son aquellos para los cuales la expresión

$$\sum_{t=1}^n (x_t - \alpha_0 - \dots - \alpha_r t^r)^2$$

es mínima.

Si $\{Y_t\}$ es un proceso lineal generado por un ruido blanco, y bajo algunas hipótesis adicionales, se tiene que a_i es aproximada-

mente igual a α_i , $i = 1, \dots, r$; cuando n es suficientemente grande. Además las autocorrelaciones de la serie residual (z_t) definida por

$$z_t = x_t - a_0 - \dots - a_r t^r$$

son aproximadamente iguales a las autocorrelaciones del proceso correspondiente $\{Z_t\}$ definido por

$$Z_t = X_t - \alpha_0 - \dots - \alpha_r t^r$$

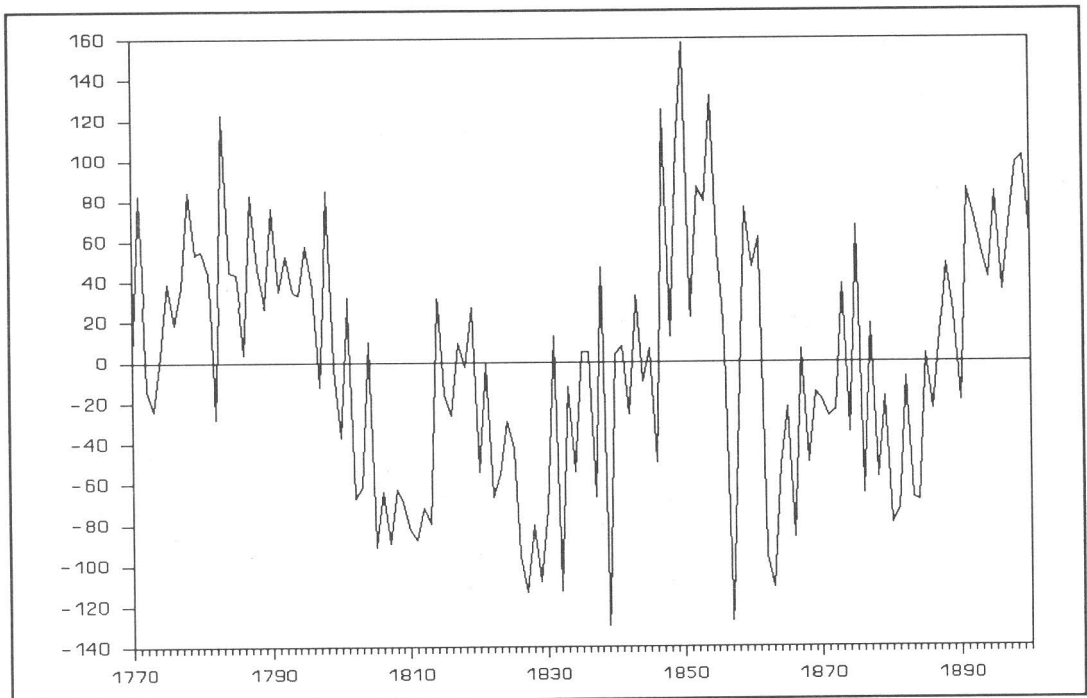


GRAFICO 7. Serie de bautizos anuales de Cartago. La tendencia lineal fue eliminada por mínimos cuadrados.

Ejemplo

Suponiendo que la serie de bautizos anuales de Cartago, sea acorde con el modelo

$$X_t = \alpha_0 + \alpha_1 t + \{Y_t\}$$

donde $\{Y_t\}$ es un proceso estacionario, se estiman los parámetros α_0 y α_1 , utilizando los valores de la serie (ver **Tabla N° 2**). Los valores estimados son

$$a_0 = 266.33 \quad \text{y} \quad a_1 = 4.4$$

La serie residual $w_i = x_i - (266.33 + 4.4 i)$; $i = 1, \dots, 131$, es representada en el GRAFICO 7. Se nota que la tendencia lineal desapareció y la nueva serie varía alrededor del nivel cero, con varianza aproximadamente constante.

Comparando GRAFICO 7 con GRAFICO 5 se nota que las series difieren, lo cual puede comprobarse algebraicamente, restando las series (e_i) y (z_i) antes definidas. De modo que, se pueden usar diferentes métodos de estacionarización para investigar la estabilidad de los resultados del análisis espectral.

4.3 Otros métodos y tendencias

Las medias móviles es otro método muy usado para remover tendencias polinomiales. En particular, las medias móviles entradas para el caso de tendencias polinomiales de grado p , se definen por

$$z_t = \sum_{j=-m}^m w_j x_{t+j}$$

donde (x_t) es la serie original y w_j , $j = -m, -m+1, \dots, m$ son $2m+1$ números tales que

$$\sum_{j=-m}^m w_j j^r = 0, \quad r = 0, 1, \dots, p$$

Las medias móviles juegan un papel importante en los métodos de desestacionalización de series económicas; particularmente en el método desarrollado por el Bureau of the Census de los Estados Unidos, cuya décima primera versión es llamada X-11.

En cuanto a las tendencias, hay otras menos usuales como la logística

$$f(t) = (k + ab^t)^{-1}, \quad b > 0$$

que presenta mayores complicaciones para estimar los parámetros a , b y k ; y la tendencia exponencial definida por

$$g(t) = e^{k \cdot (r/t)}.$$

Para ampliar más sobre el tema de las tendencias, se puede consultar Granger (1980) y Granger y Hatanaka (1964).

4.4 Estabilización de la varianza

Una serie estacionaria necesariamente tiene media y varianza independientes del tiempo. Sin embargo puede ocurrir que en el gráfico de una serie se distingan cambios notorios en la varianza a través del período observado. En tales circunstancias es usual transformar la serie con el fin de estabilizar su varianza. El

logaritmo y la raíz cuadrada son transformaciones frecuentemente utilizadas con este fin. En todo caso, se transforma la serie original en otra, denotada por (z_t) , donde

$$z_t = \log(x_t) \quad \text{o} \quad z_t = (x_t)^{1/2}.$$

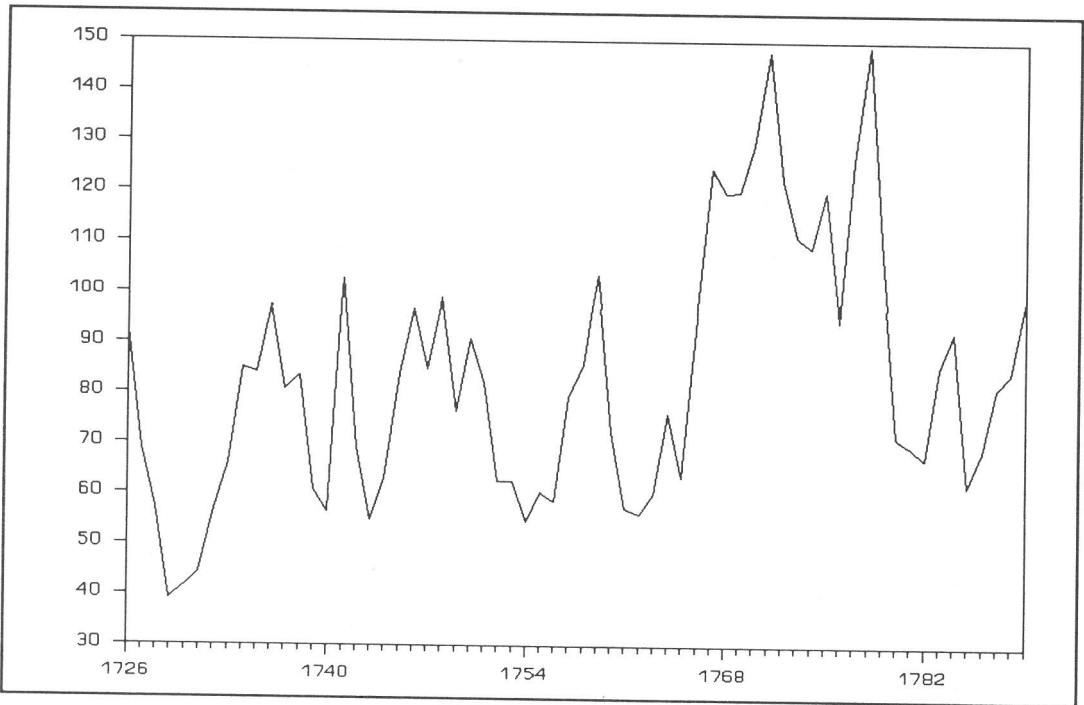


GRAFICO 8(a). Indices del precio del vino en Francia
1726 - 1789

Ejemplo

La serie de índices del precio del vino en Francia durante los años 1726 - 1789 exhibe una varianza aproximadamente estable en los dos primeros tercios del periodo considerado. En el último tercio se observa un crecimiento en la varianza (ver GRAFICO 8(a)).

En el GRAFICO 8(b) se representa el logaritmo de esta serie; como puede verse, la varianza es ahora más estable a lo largo de todo el periodo².

² Los datos fueron tomados de Labrousse, Ernest. Fluctuaciones económicas e Historia Social, Editorial Tecnos, S.A. 1967, Madrid, pp. 143 - 144.

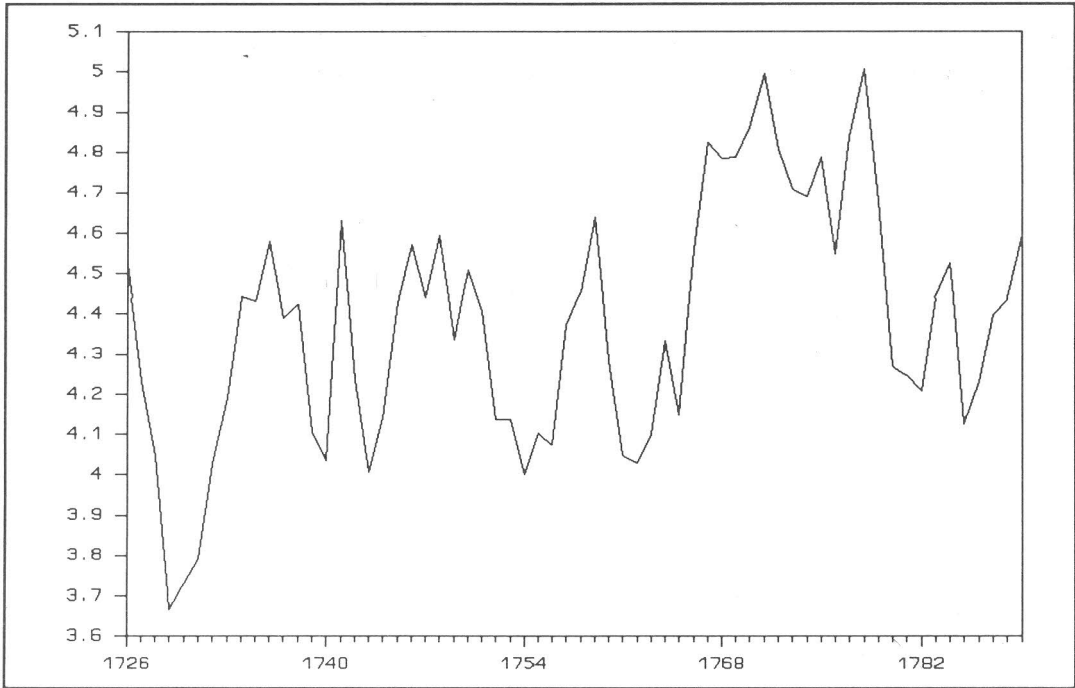


GRAFICO 8(b). Logaritmo de los índices del precio del vino en Francia, 1726 - 1789.

5. EL PERIODOGRAMA

Un análisis espectral puede ser aplicado ya sea en el proceso de búsqueda de un modelo que se ajuste adecuadamente a los datos, o para detectar ciertas regularidades llamadas ciclos o tendencias cíclicas, posiblemente presentes en la serie, pero que no son plenamente determinables bajo la inspección gráfica. Aunque estos objetivos pueden ser complementarios, de momento se deja de lado el primero, limitándonos al estudio de los ciclos.

Siendo el periodograma el instrumento fundamental para dicho estudio, esta sección se dedica con detalle, a la presentación del concepto, cálculo y significado de aquél.

Las relaciones (9) y (10), son especialmente importantes, puesto que sirven de base para la introducción e interpretación del periodograma.

5.1 Representación trigonométrica de una serie

En esta subsección se presenta un resumen del procedimiento matemático para deducir la fórmula (9).

Con las fórmulas (2) a (5) se obtiene la relación (6) y con ésta, la representación (9) de la serie x_1, \dots, x_n .

Consideremos las funciones

$$y_k(t) = \cos(w_k t) \quad \text{y} \quad z_k(t) = \text{sen}(w_k t), \quad t = 1, 2, \dots, n$$

donde $k = 1, 2, \dots, n/2$ (suponemos que n es par)³. Estas funciones satisfacen para $w_k = 2\pi k/n$, las siguientes relaciones:

$$(2) \quad \sum_{t=1}^n \cos(w_k t) \cos(w_j t) = \begin{cases} 0 & \text{si } k \neq j \\ n & \text{si } k = j = n/2 \\ n/2 & \text{si } k = j \neq n/2 \end{cases}$$

$$(3) \quad \sum_{t=1}^n \text{sen}(w_k t) \text{sen}(w_j t) = \begin{cases} 0 & \text{si } k \neq j \text{ o } k = j = n/2 \\ n/2 & \text{si } k = j \neq n/2 \end{cases}$$

$$(4) \quad \sum_{t=1}^n \cos(w_k t) \text{sen}(w_j t) = 0 ; \quad k, j = 1, 2, \dots, n/2.$$

³ Si n es impar, se prueba de igual modo que la serie se escribe en la forma:

$$x_t = \hat{u} + \sum_{k=1}^h a_k \cos(w_k t) + b_k \text{sen}(w_k t)$$

donde $t = 1, \dots, n$ y $h = (n-1)/2$.

$$(5) \quad \sum_{t=1}^n \cos(w_k t) = \sum_{t=1}^n \operatorname{sen}(w_k t) = 0 ; k = 1, 2, \dots, n/2.$$

Utilizando las relaciones (2) a (5) se obtiene la representación trigonométrica de la serie $x_1, x_2, x_3, \dots, x_n$:

$$(6) \quad x_t = \hat{u} + \sum_{k=1}^{n/2} [a_k \cos(w_k t) + b_k \operatorname{sen}(w_k t)]; t = 1, 2, \dots, n$$

donde⁴

$$\hat{u} = (1/n) [x_1 + \dots + x_n]$$

$$(7) \quad a_k = (2/n) \sum_{t=1}^n x_t \cos(w_k t).$$

$$(8) \quad b_k = (2/n) \sum_{t=1}^n x_t \operatorname{sen}(w_k t).$$

De (6) se deducirá la representación (9), que muestra más claramente el significado de las cantidades estimadas. Para $k = 1, \dots, n/2$, se definen los números h_k y u_k de modo que:

$$a_k = h_k \cos(u_k) \quad \text{y} \quad b_k = -h_k \operatorname{sen}(u_k).$$

Entonces los nuevos parámetros h_k y u_k asumen la forma

$$h_k = (a_k^2 + b_k^2)^{1/2} \quad \text{y} \quad u_k = \operatorname{tangible}^{-1}(-b_k/a_k).$$

Sustituyendo a_k y b_k en la relación (6), se obtiene

$$(9) \quad x_t = \hat{u} + \sum_{k=1}^{n/2} h_k \cos(w_k t + u_k), t = 1, \dots, n$$

La representación (9) de la serie (x_t) dice que la variación en la serie es el resultado de una suma de componentes cíclicos

⁴ Los valores \hat{u} , a_k , b_k tienen la propiedad de ser los estimados de la siguiente regresión sin término constante: se toma como variable dependiente la serie x_1, \dots, x_n . Las variables independientes son las $n-1$ funciones definidas por:

$$y_k(t) = \cos(w_k t); k = 1, \dots, n/2 \quad \text{y} \\ z_k(t) = \operatorname{sen}(w_k t); k = 1, \dots, (n/2)-1$$

donde $t = 1, \dots, n$

(cosenos) cuyas amplitudes son los valores h_k asociados a las respectivas frecuencias k/n . Estas frecuencias están dadas por unidad de tiempo : fracción de ciclos por segundo, por día o por mes, dependiendo si n está dado en segundos, días o meses respectivamente. El período (n/k) es la "longitud" del ciclo asociado a la frecuencia k/n .

5.2 Descomposición de la varianza

La idea es expresar la varianza de la serie x_1, \dots, x_n en una suma, donde cada término tiene un significado específico, lo cual permitirá establecer en 5.3, la definición del periodograma.

El análisis de las fluctuaciones cíclicas de la serie x_1, \dots, x_n , descansa sobre la descomposición de su varianza en una suma, en la cual cada término expresa aquella parte de la varianza debida a la presencia en los datos, de un ciclo de una frecuencia determinada. Veremos enseguida que esta descomposición nos conduce al concepto de periodograma.

Usando la representación (6) de la serie y las relaciones (2) a (5) se puede deducir la relación

$$(10) \quad S^2 = (1/n) \sum_{t=1}^n (x_t - \hat{u})^2 = \sum_{k=1}^p (a_k^2 + b_k^2)/2 + a_{n/2}^2$$

donde $p = (n/2) - 1$ y S^2 es la varianza de la serie.

Esta relación (10) nos dice que la varianza de la serie es el resultado de sumar los cuadrados de las amplitudes en las primeras p frecuencias y dividir por dos, más el cuadrado de la amplitud en la frecuencia $1/2$. Esta descomposición va a permitir, en la sección 5.3, representar gráficamente las contribuciones a la varianza de la serie (que denotamos con h_k^2), en cada frecuencia k/n ; de modo que el área bajo la curva sea aproximadamente igual a la varianza total de la serie.

5.3 Concepto de periodograma

Usando la idea geométrica de área y la descomposición de la varianza de la serie (relación (10)), para introducir el concepto de periodograma, se facilita la interpretación de éste. El periodograma de la serie x_1, \dots, x_n es una función que denotamos con $I_n(w_k)$, $w_k = 2\pi k/n$, $k = 1, \dots, n/2$; cuyo gráfico es la curva formada por los segmentos que unen los puntos $(w_k, I_n(w_k))$, como se indica en el dibujo siguiente, donde se dividió el intervalo $[\pi/n, \pi + \pi/n]$ en $n/2$ subintervalos de la forma

$$R_k = [w_k - \pi/n, w_k + \pi/n], \quad k = 1, \dots, n/2.$$

y donde se denota $x_k = w_k - \pi/n$, $k = 1, \dots, (n/2 + 1)$

- b) $L(0)$ es el valor máximo de la función L .
 c) L es una función par : $L(j) = L(-j)$ para todo j .

Entre los diversos tipos de ventanas usados para calcular el periodograma ponderado, se encuentran las ventanas rectangular y triangular. Estas se distinguen porque son definidas por medio de fórmulas muy simples, lo cual facilita el cálculo. Veamos:

1. Ventana lisa o rectangular: esta ventana se define por la fórmula

$$L(j) = 1/m, \text{ si } |j| \leq d \text{ y}$$

$$L(j) = 0, \text{ si } |j| > d$$

donde $m = 2d + 1$.

En este caso el estimador del espectro es:

$$f_n(w_k) = (1/m) \sum_{j=-d}^d I_n(w_{k+j})$$

Como se ve en esta fórmula, el estimador espectral es un promedio aritmético de las ordenadas del periodograma. En este promedio a todas las ordenadas se les da la misma importancia (el peso $1/m$). Como esto puede no ser adecuado se definen también otras ventanas.

2. Ventana triangular: Esta ventana se define por

$$L(j) = [1/(d+1)] - |j|/(d+1)^2 \text{ si } |j| \leq d \text{ y}$$

$$L(j) = 0 \text{ si } |j| > d.$$

El estimador correspondiente es $f_n(w_k) = \sum_{j=-d}^d L(j) I_n(w_{k+j})$.

En este promedio ponderado, entre más cerca esté w_{k+j} de w_k , más importancia se le da a las ordenadas del periodograma $I_n(w_{k+j})$. Esto es así porque $L(j)$ decrece cuando j varía desde cero hasta d , teniendo su gráfico una forma triangular.

6.3 Ancho de la ventana

El número $m = 2d + 1$ llamado ancho de la ventana indica el número de ordenadas del periodograma que se usa para calcular el estimador espectral f_n .

El ancho de la ventana se escoge por un proceso más que todo empírico, en razón de que la teoría no es conclusiva acerca del valor de m adecuado para mantener un sesgo y varianza aceptables. Se debe tener en cuenta que, cuando m aumenta, la varianza del estimador espectral $f_n(w_k)$, decrece, y su sesgo crece. Por lo tanto, escoger m implica controlar dos efectos que actúan en sentido contrario: la varianza y el sesgo.

Generalmente la escogencia de m se hace manteniendo el cociente d/m tan pequeño como sea posible, para acercarse mejor a las hipótesis de la teoría.

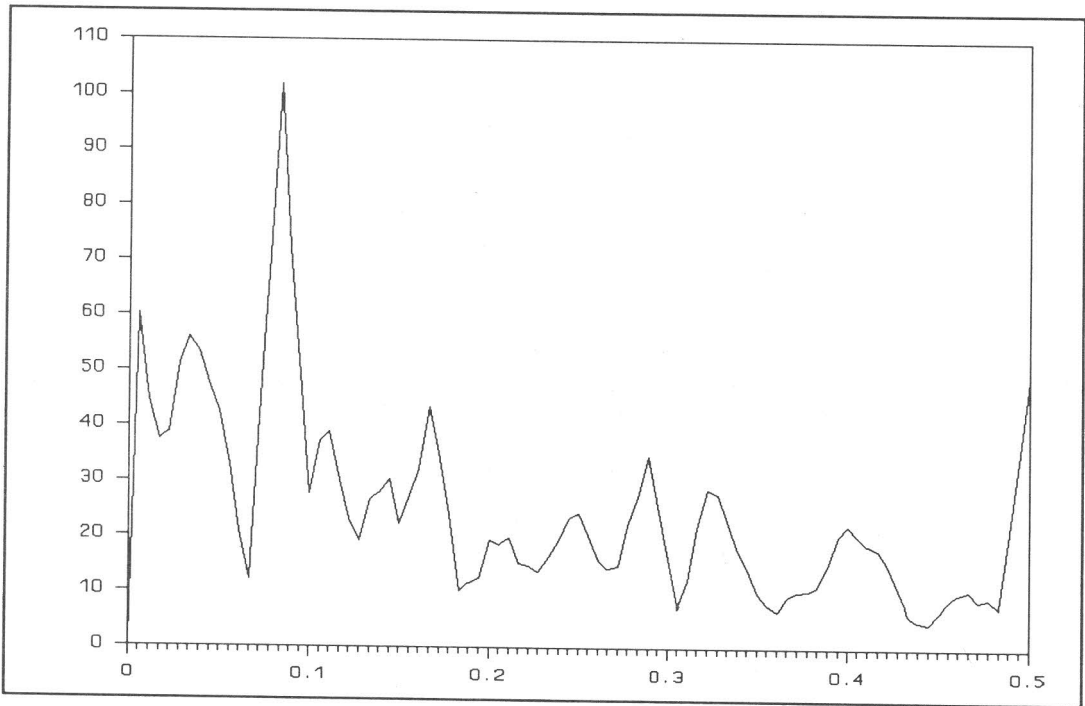


GRAFICO 10. Periodograma ponderado. Se aplicó una ventana triangular de ancho 5.

Ejemplo

Con propósitos de ilustración vamos a aplicar el método de estimación descrito, al periodograma de la serie de bautizos mensuales de Cartago (ver GRAFICO 3 y GRAFICO 9). En este ejemplo el estimador del espectro ha sido calculado utilizando una ventana triangular de ancho 5, por lo cual la fórmula se escribe así:

$$f_{180}(w_k) = \sum_{j=-2}^2 L(j) \cdot I_{180}(w_{k+j}) \quad \text{donde } L(j) = (1/3) - (1/9)|j|.$$

Los valores calculados con esta fórmula son ordenados en la Tabla N° 5.

Consideremos ahora el GRAFICO 10. Sobre el eje de las abscisas se colocan los valores de $k/180$, $k = 1, \dots, 90$; y sobre el eje de las ordenadas, los valores del periodograma ponderado $f_{180}(w_k)$, $k = 1, \dots, 90$.

El GRAFICO 10 muestra dos características muy importantes. La primera, que es una curva más "regular" que el gráfico del periodograma (GRAFICO 9). Este es uno de los objetivos buscados puesto que nos interesa un estimador con varianza decreciente. La segunda, que se conservan los rasgos más importantes del periodograma, observados en el GRAFICO 9. Resalta en particular, la clara permanencia del máximo en la misma frecuencia 15/180.

6.4 Bandas de confianza aproximadas

Las propiedades distribucionales del periodograma ponderado f_n , permiten calcular intervalos de confianza aproximados para el espectro. Sea g el espectro del proceso $\{X_t\}$, la variable aleatoria

$$Y(w_k) = \tau \cdot f_n(w_k) / g(w_k)$$

es aproximadamente distribuida como una variable aleatoria chi - cuadrado con τ grados de libertad, donde

$$\tau = 2 / \left(\sum_{j=-d}^d L(j)^2 \right).$$

Se puede calcular sin dificultad que, para el caso de una ventana triangular o rectangular, los grados de libertad vienen dados respectivamente por:

$$\tau = \frac{6(d+1)^3}{2d^2 + 4d + 3} \quad \text{y} \quad \tau = 2(2d+1).$$

Para cada w_k se puede calcular un intervalo de confianza aproximado, de la siguiente manera:

Dado un nivel de significación α , por lo general $\alpha = 0.05$ ó $\alpha = 0.01$, se calculan con la ayuda de una tabla, los valores c_1 y c_2 tales que

$$P(Y(w_k) \leq c_1) = P(Y(w_k) \geq c_2) = \alpha/2.$$

En consecuencia se tiene que:

$$P(c_1 \leq Y(w_k) \leq c_2) = 1 - \alpha.$$

A partir de $c_1 \leq Y(w_k) \leq c_2$, se obtiene la siguiente doble desigualdad:

$$m_k = \tau f_n(w_k) / c_2 \leq g(w_k) \leq \tau f_n(w_k) / c_1 = M_k.$$

Si en un mismo sistema de coordenadas se representan los valores m_k , M_k y $f_n(w_k)$ para $k = 1, \dots, n/2$, se obtiene un gráfico de las bandas de confianza aproximadas.

El GRAFICO 11 muestra las bandas de confianza para el espectro ponderado. La curva de abajo es la banda inferior, la cual corresponde concretamente a los puntos $(k/180, m_k)$; $k = 1, \dots, 90$.

De igual modo, la curva de arriba es la banda superior formada por los puntos $(k/180, M_k)$; $k = 1, \dots, 90$. La curva central es el periodograma suavizado (el mismo del GRAFICO 10). Las bandas de confianza para los 2 primeros y los 2 últimos valores no son adecuadas, dados los ajustes que necesariamente hubo que introducir en la definición de f_n , al inicio de esta sección. Pero constituyen una minoría de puntos en el gráfico, ocasionando por lo tanto una imprecisión que no tiene mayores consecuencias.

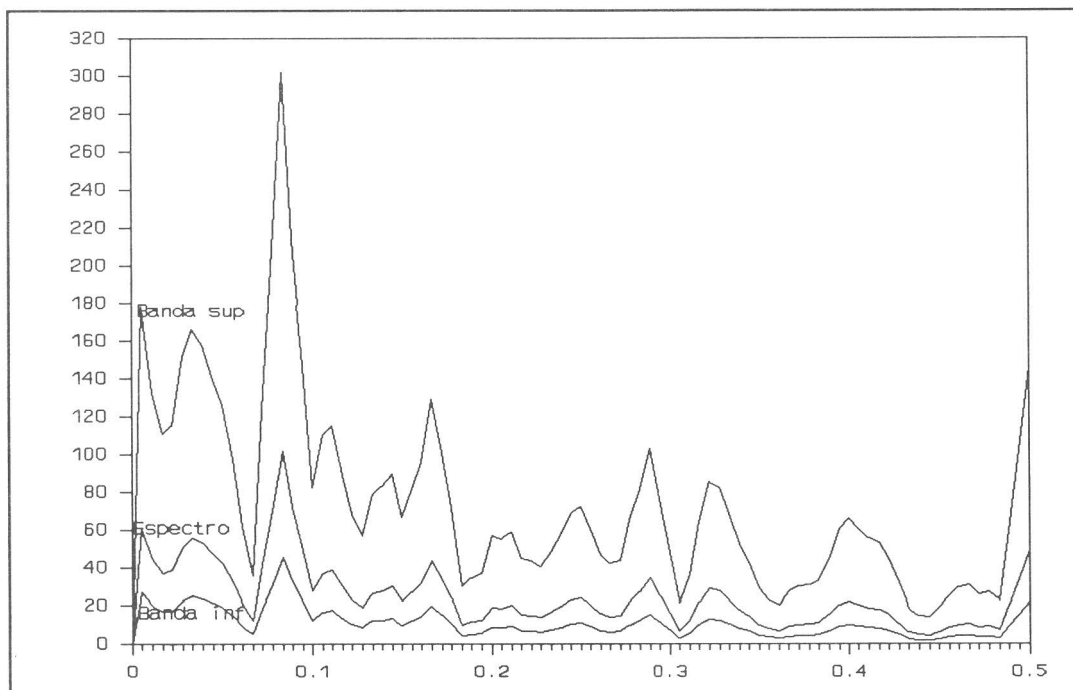


GRAFICO 11. Bandas de confianza aproximadas del periodograma ponderada.

Observación

La estimación del espectro tiene importancia principalmente en dos aspectos:

(a) Es un recurso usado para estudiar la estabilidad de los puntos (frecuencias) de máxima contribución a la varianza de la serie. Aplicando diferentes tipos de ventanas y con diferentes anchos, se puede observar cuales de los puntos más altos en la curva se conservan. En el proceso de estudio de un ciclo, es también útil realizar varias formas de estacionarización de la serie para observar la estabilidad del ciclo.

(b) Por las propiedades asintóticas del periodograma ponderado (insesgado y de varianza decreciente a cero), es un medio para estudiar la forma de la función espectral (o espectro) del proceso estocástico. Esto es de particular interés cuando se desea encontrar un modelo generador de la serie, por ejemplo de tipo Box - Jenkins.

7. "SOFTWARE" UTILIZADO

Esta sección contiene los programas "RATS" usados en los cálculos, y algunas explicaciones sobre la función de las diferentes instrucciones. La sección finaliza con una breve mención de las facilidades que ofrece el STATGRAPHICS para realizar un análisis espectral.

7.1 El paquete RATS

RATS versión 2.01, para microcomputador IBM o compatibles, es un conjunto de programas orientados a la Econometría, los pronósticos y el análisis estadístico. Para la redacción de los programas necesarios para hacer los cálculos espectrales, se usó el manual de RATS versión 2.0, especialmente los capítulos 1 y 13. Cada término de una serie se denomina con el nombre de "entrada".

Al final de cada programa se incluyen las explicaciones que se consideraron pertinentes, sobre la función de las instrucciones.

Observaciones

(a) Los números que aparecen delante de algunas instrucciones, son para facilitar la referencia.

(b) Un asterisco (*) delante de una instrucción indica que ésta no será ejecutada. Es el caso de la instrucción 3., **Programa 3.**

(c) Las líneas en los programas que empiezan con el símbolo #, se vinculan con la instrucción inmediatamente anterior y forman parte de ella. Ver por ejemplo, las instrucciones 7 y 11 del **Programa 1** y las instrucciones 4 y 5 del **Programa 4.**

Programa 1

Calcula el periodograma de la serie estacionaria de longitud 180, contenida en el archivo "BAUTIZOS.PRN. Grava el periodograma en el archivo "PER.PRN".

```
01. OPEN DATA C:BUTIZOS.PRN
02. CAL 1800 1 12
03. ALL 0 1814:12
04. DATA / SERIE
05. DIFFERENCE SERIE / 0 SERIE1

06. FREQ 2 180
07. RTOC
   # SERIE1
   # 1
08. FFT 1
09. SMPL 1800:2 1807:7
10. CMULT(SCALE=1./(PI*180)) 1 1
```

```

11. CTOR
    # 1
    # PERIODOG
12. OPEN COPY C:PER.PRN
13. COPY(ORG=OBS) 1800:2 1807:7 PERIODOG
    END

```

Significado de las instrucciones:

01. Abre el archivo BAUTIZOS.PRN que contiene la serie de bautizos mensuales de Cartago correspondiente al período enero de 1800 a diciembre de 1814.

02. CAL es la abreviación de CALENDAR. Esta instrucción pone la fecha de la primera entrada de la serie (enero de 1800). El número 12 significa que hay 12 entradas por periodo. Es decir, la serie es mensual con 12 entradas por año. Si la serie fuera trimestral debe escribirse un 4 en lugar del 12. Por ejemplo: CAL 1900 2 4 significa que la serie empieza en el segundo trimestre del año 1900.

03. ALL es la abreviación de ALLOCATE. La instrucción sirve para decirle a RATS cual es la última entrada de la serie, la cual en el caso presente, corresponde a diciembre de 1814. Se escribe un cero después de ALL para advertir a RATS que no se usarán series numeradas.

04. Lee la serie y le asigna el nombre SERIE.

05. Le resta a cada entrada de SERIE la media de la serie. A la serie centrada le llama SERIE1. Si en lugar de un 0 se usa un número entero positivo N, esta instrucción calcula la N - ésima diferencia de la serie. Generalmente $N = 1$ o $N = 2$.

06. Especifica el número de entradas de las series complejas que van a ser creadas. En este caso se puede crear hasta dos series complejas con 180 entradas.

Cuando sea necesario completar la serie con ceros para que el cálculo de la transformada finita de Fourier, se realice en menos tiempo, se escribe en FREQ el número deseado. RATS automáticamente pone ceros hasta completar la longitud especificada.

07. Permite a RATS tratar las entradas de SERIE1 como números complejos con parte imaginaria igual a cero. La serie compleja la denomina con el número 1.

08. Calcula la transformada finita de Fourier de la serie 1. La serie que contiene estos valores complejos, se la denomina también con el número 1.

09. Indica las entradas de la serie 1 que van a ser usadas por la instrucción siguiente. Sólo se toman en cuenta las primeras 90

entradas a partir de la segunda, ya que la transformada finita de Fourier es periódica. Además la primera entrada será cero siempre que la serie original sea centrada (instrucción 05).

10. Cada entrada de la serie 1, comprendida en el rango indicado en la instrucción anterior, se multiplica por su conjugado y el resultado se divide por $180 \cdot \pi$. A la nueva serie se le asigna el número 1. Esta serie tiene todas sus entradas complejas con parte imaginaria igual a cero.

11. Análoga a la instrucción 07. Las entradas de la serie 1, que son complejos con parte imaginaria nula, serán tratadas como números reales. La nueva serie de números reales se llama PERIODOG.

12. Abre un archivo con el nombre PER.PRN.

13. Grava en el archivo PER.PRN las entradas de la serie PERIODOG, comprendidas en el rango 1800:2 1807:7.

Programa 2

Remueve la tendencia lineal, sea por regresión o calculando una diferencia a la serie. En este caso, elimina la tendencia lineal por regresión, de la serie de longitud 131, contenida en el archivo BANUALES.PRN. Grava la serie resultante SINTEND, en el archivo SINTEND.PRN.

```
OPEN DATA C:BANUALES.PRN
CAL 1770 1 1
ALL 0 1900:1
DATA / SERIE
```

1. SET AÑO = T
2. OLS BANUALES / SINTEND
CONSTANT AÑO
3. *DIFFERENCE BANUALES / 1 SINTEND

```
OPEN COPY C:SINTEND.PRN
COPY(ORG=OBS) 1770:1 1900:1 SINTEND
END
```

Significado de las instrucciones:

1. Crea la serie 1,2, ... ,131 y le da el nombre de AÑO.

2. Hace una regresión en la cual la variable dependiente es la serie BANUALES. La variable independiente es la serie AÑO y la constante es indicada con la palabra reservada CONSTANT. Además, crea la serie de residuos de la regresión con el nombre SINTEND.

3. La instrucción no es ejecutada. Si se quita el asterisco, RATS calcula la primera diferencia de la serie BANUALES y a la serie resultante le asigna el nombre SINTEND.

Programa 3

Calcula el periodograma ponderado a partir del periodograma calculado con el **Programa 1**, cuyos valores se encuentran en la tabla N° 4, Anexo 2. Se usó una ventana triangular de ancho 5. En el texto se explicó que el estimador así obtenido debe ser modificado en aquellas ordenadas cercanas a cero y π .

La serie resultante (periodograma ponderado), se grava en el archivo con nombre "ESPECTRO.PRN".

```

OPEN DATA C:PER.PRN
CAL 1 1 1
ALL 0 90:1
DATA / PERIODOG
FREQ 2 90
RTOC
# PERIODOG
# 1
WINDOW(TYPE=TENT) 1 1:1 90:1 5 2 1:1
CTOR
# 2
# ESPECTRO
OPEN COPY C:ESPECTRO.PRN
COPY(ORG=OBS) 1:1 90:1 ESPECTRO
END

```

Significado de la instrucción WINDOW: calcula el periodograma ponderado de la serie 1 en el rango 1:1 90:1. Usa una ventana triangular de ancho 5. A la nueva serie se le asigna el número 2.

Note que se debe escribir 1:1 al final de la instrucción, para decirle a RATS que la serie 2 empieza en la primera entrada.

Programa 4

Calcula 150 puntos de la serie autorregresiva definida por la ecuación

$$y_t = 0.6 y_{t-1} + e_t$$

donde $\{e_t\}$ es un ruido blanco normal, con media cero y varianza igual a 2.

```

1. ALL 0 151
2. ZERO Y
3. EQUATION 1 Y 1 0
4. ASSOCIATE 1 0 0 2
# 0 0.6
5. SIMULATE(SETUP) 1 150 2
# 1 Y

```

```
6. SIMULATE
   OPEN COPY SERIE.PRN
   COPY(ORG=OBS) 2 151 Y
   END
```

Significado de las instrucciones:

1. Establece el número máximo de entradas de la serie Y. Para este caso son 151

2. Pone las entradas de Y en cero

3. Define la ecuación autorregresiva asociada a la variable Y, cuyos valores se van a simular. Los números leídos de izquierda a derecha, expresan a RATS que se trata de la ecuación N° 1, con un parámetro autorregresivo, y cero parámetros de promedios móviles.

RATS puede hacer simulaciones con sistemas de ecuaciones, por ello prevee una enumeración de éstas. En nuestro caso sólo hay una ecuación, por eso se le asignó el número 1.

4. Esta instrucción consta de dos líneas. En la primera se le dice a RATS que se van a especificar los parámetros de la ecuación N° 1 (es el significado del 1), y que el ruido blanco en esta ecuación, tiene varianza igual a 2. El significado de los dos ceros se explica más abajo.

En la segunda línea, precedidos por el símbolo #, se escriben los valores de los coeficientes de la ecuación, empezando con el término constante, que en este caso es nulo. El coeficiente autorregresivo es igual a 0.6.

Para comprender mejor el significado de esta instrucción, veamos su sintaxis general:

```
ASSOCIATE N C RESID V
# coeficientes de la ecuación.
```

donde los parámetros N, C, RESID, y V se definen por:

N : número de la ecuación cuyos coeficientes van a ser especificados.

C : es el nombre del vector que contiene los coeficientes de la ecuación. Hay dos formas de especificar los coeficientes de la ecuación: por medio del vector C o usando la segunda línea de la instrucción. Si se hace de la última forma entonces se escribe un 0 en lugar del vector C.

RESID : es el nombre de la serie en que se guardan los residuos para luego hacer cálculos. Siempre que esto no sea necesario, se escribe un 0.

V : es un número real que indica la varianza del ruido blanco. Sólo se usa si RESID y C son cero.

Cuando se usa C = 0, los coeficientes de la ecuación van en la segunda línea precedidos por el símbolo #, en el orden siguiente:

primero el término constante, luego los coeficientes autorregresivos y por último los coeficientes de promedios móviles.

5. Igual que la instrucción 4, tiene dos líneas. En la primera se indica que el sistema consta de una ecuación, lo cual se expresa con el número 1. Luego aparecen los números 150 y 2, para decirle a RATS que simule 150 entradas de la serie empezando en la entrada 2. Se empieza en 2 porque la ecuación tiene un coeficiente autorregresivo, lo que obliga a darle el valor cero a la primera entrada de la serie para poder calcular las siguientes. Si tuviese 2 coeficientes autorregresivos, entonces las dos primeras entradas serían iguales a cero y las simulaciones empezarían en 3, y así sucesivamente.

En la segunda línea, el número 1 significa que RATS debe usar la ecuación número 1 para hacer las simulaciones. Naturalmente, si hubiesen más ecuaciones en el sistema, este número podría ser diferente de 1.

Los resultados se guardan en Y, empezando en la entrada N° 2.

6. Ordena ejecutar las simulaciones.

7.2 El paquete STATGRAPHICS

Este paquete en su versión 2.0 (1986) es bastante apropiado para realizar el análisis de la estacionaridad de una serie y para hacer los cálculos espectrales. Además tiene la ventaja de ser interactivo, lo cual lógicamente facilita su uso.

En lo que concierne al análisis de la estacionaridad, se pueden estimar tendencias polinomiales, aplicar promedios móviles y diferenciación de la serie. También se pueden tratar series con "ciclo estacional". Las series transformadas pueden ser graficadas fácilmente.

Con respecto a los cálculos espectrales, se usa el procedimiento "PERIODOGRAM" para calcular el periodograma de la serie con la fórmula $2\pi I_n(w_k)$. Hace un gráfico de los puntos $(k/n, 2\pi I_n(w_k))$ con $k = 1, \dots, n/2$.

Se pueden obtener estimaciones del espectro, recurriendo a las distintas variantes de promedios móviles disponibles. Hay dos alternativas:

- a) Escoger un promedio móvil simple, lo que corresponde a usar una ventana rectangular. Se debe indicar el ancho de la ventana. Es decir, el número de ordenadas del periodograma que se promedian.
- b) Escoger un promedio móvil ponderado. En este caso hay que indicar el ancho y los valores de la ventana.

Para una lectura más amplia sobre el uso de este "software", se debe consultar el manual (ver BIBLIOGRAFIA BASICA, 10.).

8. COMENTARIO FINAL

Se han planteado en las páginas anteriores tres temáticas principales: los elementos fundamentales del proceso de estacionarización de una serie temporal, el estudio del periodograma y un método de estimación del espectro.

El material ha sido organizado de tal forma que su desarrollo corresponda lo mejor posible, con el orden que usualmente se sigue cuando se realiza un análisis espectral de series temporales.

Utilizamos una serie de datos reales para completar la exposición de los métodos de cálculo. Acudimos también a la representación gráfica para ilustrar diversos conceptos y facilitar las interpretaciones.

Naturalmente, hay varios temas relacionados con esta materia que no fueron abordados en el presente texto. Algunos de ellos son, por ejemplo, filtros, desestacionalización, uso del "taper" y el análisis de otros tipos de tendencias. Por su interés, serán objeto de un trabajo posterior.

ANEXO I: DEFINICIONES

Este anexo contiene las definiciones que hemos considerado más cercanamente relacionadas con el contenido del texto.

Algunas definiciones muy conocidas fueron, por su relevancia en análisis espectral, incluidas en este anexo. Nos referimos por ejemplo a la media, varianza, covarianza, número complejo y su conjugado, etc.

Otras definiciones menos importantes en relación con los aspectos operativos del análisis espectral, no se incluyen, a fin de acortar la extensión de este anexo. Pensamos por ejemplo en las definiciones de variable aleatoria, función continua, función derivable, espacio de probabilidad, densidad de probabilidad, etc.

1. Número complejo: se llama número complejo a la expresión $a + b \cdot i$, donde $i^2 = -1$. Los números reales a y b se llaman respectivamente, parte real y parte imaginaria del número complejo.

2. Conjugado de un número complejo: el conjugado de un número complejo $a + b \cdot i$ es $a - b \cdot i$. Usualmente se escribe:

$$\overline{a + b \cdot i} = a - b \cdot i.$$

3. Módulo de un número complejo: el módulo de un número complejo $a + b \cdot i$ es $(a^2 + b^2)^{1/2}$. Usualmente se denota por $|a + b \cdot i|$.

La relación $\overline{(a + b \cdot i)} \cdot (a + b \cdot i) = |a + b \cdot i|^2$ es de uso común.

4. Media y varianza: la media de una variable aleatoria X con densidad de probabilidad f , se denota con $E[X]$ o μ , y se define por

$$E[X] = \mu = \int_{-\infty}^{\infty} x f(x) dx.$$

La varianza de X se denota con σ^2 y se define por

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx.$$

Para el caso particular en que la variable X asume solamente los valores x_1, \dots, x_r , las definiciones anteriores se transforman en las siguientes sumas:

$$\mu = p_1 x_1 + \dots + p_r x_r \quad \text{y} \quad \sigma^2 = p_1 (x_1 - \mu)^2 + \dots + p_r (x_r - \mu)^2$$

donde p_i es la probabilidad de que la variable X asuma el valor x_i . El valor de μ expresa el valor promedio de X , mientras que la varianza es una medida de la dispersión de los valores de X con respecto a la media.

5. Covarianza y correlación: la covarianza entre dos variables aleatorias X y Y , se define como $E[(X - \mu_X)(Y - \mu_Y)]$ donde μ_X y μ_Y son las medias de X y Y respectivamente. Este valor expresa el "grado" de dependencia lineal entre las variables. Sin embargo, para investigar esta forma de relación entre las variables, se prefiere usar un coeficiente cuyo valor se encuentra entre -1 y 1 . Este es el coeficiente de correlación que se define como

$$\frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

donde $(\sigma_X)^2$ y $(\sigma_Y)^2$ son las varianzas de X y Y . Si la correlación entre las variables es próxima a 1 ó -1 entonces las variables son correlacionadas. Por otra parte, si su valor es cercano a cero entonces no son correlacionadas.

6. Proceso estocástico: un proceso estocástico es una secuencia de variables aleatorias $\{X_t\}$ de valores reales, definidas sobre un mismo espacio de probabilidad, donde t varía en un subconjunto no vacío, de los números enteros. Cuando no hay ambigüedad diremos simplemente que $\{X_t\}$ es un proceso.

7. Función de medias y de varianzas de un proceso: dado el proceso $\{X_t\}$, queda determinada una función que a cada valor de t le asocia $\mu_t = E[X_t]$. Esta función se llama función de medias del proceso $\{X_t\}$.

Análogamente, si a cada t se le hace corresponder la varianza $(\sigma_t)^2$, de la variable X_t , obtenemos la función de varianzas del proceso $\{X_t\}$.

8. Función de autocovarianzas de un proceso: sea $\{X_t\}$ un proceso y $\Gamma(h,t)$ la covarianza entre las variables X_t y X_{t+h} . Si a cada par (h,t) se le asocia el valor $\Gamma(h,t)$, obtenemos la función de covarianzas del proceso $\{X_t\}$. Observemos que $\Gamma(0,t)$ es la función de varianzas.

9. Función de autocorrelaciones de un proceso: sea $\{X_t\}$ un proceso estocástico y $\delta(h,t)$ la correlación (o autocorrelación) entre las variables X_t y X_{t+h} .

La función de autocorrelaciones del proceso es la que asocia a cada par (h,t) la autocorrelación $\delta(h,t)$.

10. Proceso estrictamente estacionario: un proceso $\{X_t\}$ es estrictamente estacionario si la densidad conjunta de cualquier grupo de variables X_t, \dots, X_{t+k} , no cambia cuando los índices se desplazan r unidades. Es decir, la densidad conjunta de X_t, \dots, X_{t+k} , es igual a la del grupo de variables $X_{t+r}, \dots, X_{t+r+k}$.

Si p denota la densidad conjunta de un grupo de r variables, entonces la propiedad de estacionariedad se escribe así:

$$p(x_t, \dots, x_{t+k}) = p(x_{t+r}, \dots, x_{t+r+k})$$

11. Proceso estacionario: un proceso estocástico es estacionario si su función de medias y de varianzas son constantes, mientras que la función de autocovarianzas $\Gamma(h,t)$, sólo depende de la separación "h" entre las variables X_t y X_{t+h} . La misma propiedad vale para las autocorrelaciones. Escribimos $\delta(h,t) = \delta(h)$.

Se muestra fácilmente que todo proceso estrictamente estacionario es un proceso estacionario.

12. Serie de tiempo: una serie de tiempo es una colección de números reales x_1, \dots, x_n . Para hacer un análisis estadístico de una serie, se asocia ésta a un proceso estocástico, de modo que cada valor x_i corresponda a una realización de la variable aleatoria X_i del proceso $\{X_t\}$.

13. Transformada discreta de Fourier: sea $g(t)$ una función definida sobre los números enteros. La transformada discreta de Fourier de la función g , es $H(w)$, definida por:

$$H(w) = \sum_{t=-\infty}^{+\infty} g(t) e^{-iwt} ; \quad w \in (-\pi, \pi)$$

14. Periodograma: el periodograma de la serie x_1, \dots, x_n , es la función definida por:

$$I_n(w_k) = (1/n\pi) \left| \sum_{t=1}^n x_t e^{-iw_k t} \right|^2 ; \quad k = 1, \dots, (n/2)-1 \quad \text{y} \quad w_k = 2\pi k/n.$$

15. Espectro: sea $\Gamma(h)$ la función de autocovarianza del proceso estacionario $\{X_t\}$. La función f continua y no negativa, definida por:

$$f(w) = (1/\pi) \left[\Gamma(0) + 2 \sum_{h=1}^{+\infty} \Gamma(h) \cos(hw) \right], \quad \text{con } w \in (0, \pi)$$

se llama espectro del proceso $\{X_t\}$. El resultado 1. del anexo II establece condiciones bajo las cuales esta función existe.

16. Estimador asintóticamente insesgado : Sea A un número real y (X_n) una sucesión de variables aleatorias; X_n es un estimador asintóticamente insesgado de A , si $E[X_n] \rightarrow A$, cuando $n \rightarrow +\infty$.

17. Estimador consistente: sea X_n una sucesión de variables aleatorias y A un número real; X_n es un estimador consistente de A , si $E[(X_n - A)_2] \rightarrow 0$, cuando $n \rightarrow +\infty$.

ANEXO II: RESULTADOS

El propósito de este anexo es adjuntar alguna información sobre las bases teóricas del análisis espectral.

Se enuncia a continuación algunos resultados en relación a las propiedades del periodograma, y del periodograma ponderado. Para ampliar más sobre el tema se sugiere consultar Fuller (1976).

1. Sea $\Gamma(h)$ y $\delta(h)$ las funciones de autocovarianza y de autocorrelación respectivamente, de un proceso estacionario. Supongamos además que:

$$\sum_{h=-\infty}^{+\infty} |\Gamma(h)| < +\infty$$

Entonces existe una función f , llamada espectro del proceso $\{X_t\}$, con las siguientes propiedades:

- (a) f es continua
- (b) $f(w) \geq 0$ para todo $w \in (0, \pi)$
- (c) $f(w) = \left[\Gamma(0) + 2\Gamma(0) \sum_{h=1}^{+\infty} \delta(h) \cos(wh) \right] (1/\pi)$
- (d) $\Gamma(h) = \int_0^{\pi} f(w) \cos(wh) dw.$

Observación

Como $\Gamma(0)$ es la varianza del proceso, sustituyendo h por cero, en (d), se tiene que:

$$\text{varianza del proceso} = \Gamma(0) = \int_0^{\pi} f(w) dw.$$

Por lo tanto $f(w)$ se puede interpretar como el aporte a la varianza del proceso, en un intervalo "infinitesimal" alrededor de w . Por otra parte, si comparamos la propiedad (c) de $f(w)$, con

$$I_n(w) = S^2/\pi + (2S^2/\pi) \sum_{r=1}^{n-1} \text{Aut}(r) \cos(wr)$$

vemos que el periodograma es un estimador "natural" del espectro. No obstante, tiene el defecto de que su varianza no decrece a cero cuando la longitud de la serie crece (resultado 3.b de este anexo). Surge así la necesidad de definir el periodograma ponderado afín de lograr mejores propiedades estadísticas (ver resultado 5., más abajo).

2. Sea $\{X_t\}$ un proceso estacionario. Sea además,

$$\mu = E[X_t] \quad \text{y} \quad \sum_{h=-\infty}^{+\infty} |\Gamma(h)| < +\infty$$

Entonces:

$$E[I_n(w)] \rightarrow f(w), \text{ cuando } n \rightarrow +\infty.$$

Observación

Este resultado dice que el periodograma es un estimador del espectro, asintóticamente insesgado.

3. Sea el proceso estocástico $\{X_t\}$ definido por

$$X_t = \sum_{j=0}^{+\infty} a_j e_{t-j}$$

donde

$\{e_t\}$ es ruido blanco con varianza σ^2

$$E[e_t^4] = \beta \sigma^4$$

$$\sum_{j=1}^{+\infty} j^{1/2} |a_j| < +\infty$$

Entonces

a) covarianza($I_n(w_k), I_n(w_r)$) = a_n si $w_k \neq w_r$
 donde $a_n \rightarrow 0$ si $n \rightarrow +\infty$ y $w_k, w_r \in (0, \pi)$.

b) var($I_n(w_k)$) = $(f(w_k))^2 + b_n$
 donde $b_n \rightarrow 0$ si $n \rightarrow +\infty$

Observación

El resultado a) expresa que bajo las hipótesis enunciadas los valores del periodograma son variables aproximadamente no correlacionadas, siempre y cuando n sea suficientemente grande. El resultado b) dice que $I_n(w_k)$ es un estimador inconsistente del espectro (su varianza no tiende a cero cuando $n \rightarrow +\infty$).

4. (propiedades distribucionales del periodograma)

Sea $\{X_t\}$ un proceso tal que:

$$X_t = \sum_{j=0}^{+\infty} a_j e_{t-j}$$

$$\sum_{j=0}^{+\infty} |a_j| < +\infty$$

$\{e_t\}$ es ruido blanco con varianza σ^2

$f(w)$ es el espectro del proceso, no negativo para todo w . Sea $F_n(x, w)$ la distribución de la variable aleatoria $2I_n(w)/f(w)$ con $w \in (0, \pi)$.

Entonces

(a) $F_n(x, w) \rightarrow F(x, w)$ si $n \rightarrow +\infty$, donde $F(x, w)$ es la distribución de una variable Y_w , chi-cuadrado con dos grados de libertad.

(b) Y_w y Y_q son independientes si $w \neq q$; $w, q \in (0, \pi)$.

5. (sobre el periodograma ponderado). Supongamos que el proceso $\{X_t\}$ tiene las siguientes propiedades:

$$(a) \quad X_t = \sum_{j=0}^{+\infty} a_j e_{t \cdot j}$$

$$(b) \quad \sum_{j=1}^{+\infty} j^{1/2} |a_j| < +\infty$$

(c) $\{e_t\}$ es ruido blanco con varianza σ^2 y $E[x^4] = 3\sigma^4$ finito. Sean además las sucesiones de números reales d_n y $L(j, n)$ tales que:

$$d_n \rightarrow +\infty, \quad d_n/n \rightarrow 0 \quad \text{y} \quad \sum_{j=-d_n}^{d_n} L(j, n)^2 \rightarrow 0; \quad \text{siempre que } n \rightarrow +\infty$$

Se define para cada $w_k = 2\pi k/n \in (0, \pi)$, el periodograma suavizado por la fórmula:

$$\phi(w_k, n) = \sum_{j=-d_n}^{d_n} L(j, n) I_n(w_{k+j})$$

Bajo las condiciones anteriores se tienen los siguientes resultados:

a) $E[\phi(w_k, n)] \rightarrow f(w_k)$, cuando $n \rightarrow +\infty$

b) $[\sum_{j=-d_n}^{d_n} L(j, n)^2]^{-1} \text{var}(\phi(w_k, n)) \rightarrow f(w_k)$, cuando $n \rightarrow +\infty$

Observación

Este resultado garantiza que bajo las hipótesis enunciadas, el periodograma suavizado $\phi(w_k, n)$, es un estimador del espectro, consistente y asintóticamente insesgado.

ANEXO 3: TABLAS

Tabla N° 1. Bautizos mensuales de Cartago (1800 - 1814)

Año	N° obs	N° de bautizos	Año	N° obs	N° de bautizos	Año	N° obs	N° de bautizos	Año	N° obs	N° de bautizos	Año	N° obs	N° de bautizos
1800	1	22	1803	37	23	1806	73	23	1809	109	19	1812	145	35
	2	21		38	25		74	26		110	12		146	18
	3	38		39	24		75	22		111	34		147	21
	4	26		40	25		76	32		112	43		148	14
	5	25		41	31		77	29		113	42		149	57
	6	33		42	29		78	28		114	42		150	47
	7	38		43	34		79	36		115	34		151	33
	8	36		44	35		80	36		116	30		152	30
	9	38		45	41		81	39		117	35		153	36
	10	54		46	41		82	36		118	30		154	32
	11	15		47	31		83	32		119	30		155	28
	12	21		48	17		84	28		120	24		156	35
1801	13	35	1804	49	34	1807	85	22	1810	121	35	1813	157	30
	14	29		50	35		86	22		122	33		158	22
	15	43		51	40		87	29		123	26		159	41
	16	54		52	39		88	24		124	42		160	26
	17	58		53	45		89	47		125	33		161	36
	18	38		54	43		90	33		126	33		162	38
	19	35		55	36		91	26		127	35		163	28
	20	27		56	38		92	34		128	29		164	24
	21	43		57	33		93	32		129	25		165	36
	22	30		58	29		94	35		130	26		166	41
	23	24		59	41		95	20		131	26		167	28
	24	25		60	19		96	22		132	23		168	32
1802	25	26	1805	61	29	1808	97	38	1811	133	19	1814	169	47
	26	22		62	22		98	22		134	34		170	40
	27	33		63	31		99	43		135	35		171	41
	28	27		64	30		100	45		136	32		172	44
	29	43		65	43		101	26		137	29		173	47
	30	39		66	21		102	27		138	39		174	45
	31	29		67	15		103	37		139	32		175	56
	32	25		68	16		104	37		140	30		176	32
	33	24		69	27		105	28		141	36		177	34
	34	36		70	26		106	32		142	33		178	45
	35	24		71	47		107	33		143	23		179	34
	36	18		72	28		108	18		144	23		180	33

Tabla N° 2. Serie de bautizos anuales de Cartago (1770 - 1900)

Año	Bautizos	Año	Bautizos	Año	Bautizos	Año	Bautizos
1770	274	1805	335	1840	586	1875	805
1771	358	1806	367	1841	594	1876	677
1772	266	1807	346	1842	565	1877	765
1773	260	1808	377	1843	628	1878	694
1774	292	1809	375	1844	590	1879	738
1775	332	1810	366	1845	611	1880	680
1776	316	1811	365	1846	559	1881	691
1777	339	1812	386	1847	738	1882	761
1778	391	1813	382	1848	630	1883	706
1779	364	1814	498	1849	728	1884	709
1780	370	1815	455	1850	784	1885	786
1781	364	1816	449	1851	653	1886	763
1782	296	1817	489	1852	722	1887	805
1783	451	1818	482	1853	719	1888	844
1784	378	1819	516	1854	776	1889	823
1785	381	1820	439	1855	703	1890	785
1786	346	1821	497	1856	661	1891	894
1787	429	1822	435	1857	530	1892	885
1788	397	1823	451	1858	637	1893	873
1789	382	1824	482	1859	743	1894	863
1790	436	1825	473	1860	718	1895	910
1791	399	1826	424	1861	737	1896	866
1792	421	1827	411	1862	584	1897	903
1793	408	1828	448	1863	573	1898	938
1794	411	1829	425	1864	637	1899	946
1795	440	1830	465	1865	671	1900	910
1796	423	1831	555	1866	611		
1797	379	1832	534	1867	708		
1798	480	1833	539	1868	657		
1799	397	1834	501	1869	696		
1800	367	1835	565	1870	697		
1801	441	1836	569	1871	693		
1802	346	1837	502	1872	700		
1803	356	1838	620	1873	767		
1804	432	1839	448	1874	698		

Tabla N° 3. Autocorrelaciones (AUT_t) de la serie de bautizos anuales de Cartago (ver Tabla N° 2)

t	AUT_t	t	AUT_t	t	AUT_t
1	1.00	11	0.65	21	0.48
2	0.91	12	0.63	22	0.48
3	0.89	13	0.62	23	0.45
4	0.86	14	0.58	24	0.45
5	0.82	15	0.58	25	0.43
6	0.80	16	0.55	26	0.41
7	0.77	17	0.54	27	0.39
8	0.74	18	0.53	28	0.38
9	0.70	19	0.52	29	0.35
10	0.68	20	0.50	30	0.34

Tabla N° 4. Valores del periodograma ($I_{180}(w_k)$) para la serie de la tabla No. 1

k	$I_{180}(w_k)$	k	$I_{180}(w_k)$	k	$I_{180}(w_k)$
1	73.60	31	13.35	61	18.16
2	29.19	32	27.08	62	21.41
3	43.10	33	0.06	63	2.42
4	7.97	34	9.59	64	10.43
5	59.83	35	5.06	65	0.56
6	90.70	36	39.21	66	12.34
7	41.50	37	0.74	67	11.20
8	22.88	38	35.12	68	11.51
9	68.49	39	12.03	69	11.01
10	45.32	40	8.77	70	2.46
11	0.01	41	11.52	71	29.05
12	1.01	42	24.59	72	36.31
13	27.21	43	10.85	73	8.64
14	5.98	44	20.55	74	14.04
15	279.99	45	42.56	75	29.12
16	2.39	46	16.53	76	15.13
17	34.46	47	6.46	77	7.92
18	8.17	48	15.33	78	3.14
19	41.24	49	17.09	79	7.76
20	72.22	50	19.70	80	0.86
21	16.27	51	4.98	81	6.28
22	10.77	52	81.42	82	8.51
23	15.29	53	13.89	83	13.26
24	41.64	54	11.79	84	13.35
25	6.37	55	0.18	85	1.17
26	61.37	56	0.86	86	17.57
27	11.51	57	25.05	87	1.45
28	11.96	58	45.54	88	11.92
29	14.54	59	28.08	89	3.82
30	99.46	60	16.11	90	134.49

Tabla N° 5. Espectro estimado de la serie de la tabla N° 1. Se usó una ventana triangular con ancho 5.

k	$I_{180}(w_k)$	k	$I_{180}(w_k)$	k	$I_{180}(w_k)$
1	60.34	31	34.19	61	17.78
2	44.72	32	24.12	62	14.66
3	37.45	33	10.21	63	9.96
4	38.85	34	11.70	64	7.89
5	51.26	35	12.62	65	6.76
6	56.17	36	19.33	66	9.16
7	53.32	37	18.66	67	10.32
8	47.18	38	19.87	68	10.42
9	42.60	39	15.12	69	11.25
10	32.98	40	14.79	70	15.04
11	20.93	41	13.80	71	20.48
12	12.10	42	16.43	72	22.31
13	41.73	43	19.66	73	20.53
14	70.64	44	23.29	74	18.78
15	102.00	45	24.35	75	18.03
16	72.20	46	20.39	76	15.18
17	49.52	47	15.86	77	10.80
18	27.83	48	14.37	78	6.31
19	37.25	49	14.75	79	5.05
20	38.96	50	22.22	80	4.70
21	30.15	51	27.57	81	6.51
22	23.25	52	34.83	82	8.76
23	19.26	53	25.92	83	10.11
24	26.70	54	16.20	84	10.56
25	27.99	55	7.20	85	8.90
26	30.38	56	12.26	86	9.25
27	22.45	57	21.80	87	7.59
28	27.64	58	28.87	88	20.08
29	32.40	59	27.86	89	34.39
30	43.69	60	23.08	90	49.17

BIBLIOGRAFIA BASICA

1. Bartlett, M.S. (1956). An introduction to stochastic processes with special reference to methods and applicatios. Cambridge Univ. Press.

Es una obra consagrada al estudio de los procesos estocásticos. Trata muy brevemente el problema del análisis del periodograma. El libro se puede catalogar como de carácter teórico, aunque contiene muchos ejemplos ilustrativos. Su lectura presenta un nivel intermedio de dificultad.

2. Box, G.E.P. y Jenkins, G.M. (1976). Time series analysis forecasting and control. Holden - Day, San Francisco.

El libro es orientado a la teoría de modelos lineales de series temporales. Solamente de manera esporádica se relaciona con el análisis espectral.

3. Chatfield, C. (1975). The analysis of time series : theory and practice. London, First published by Chapman and Hall Ltd.

Es introductorio al análisis de series de tiempo. Aborda problemas teóricos y prácticos e incluye ejemplos con fines ilustrativos. Está orientado a los usuarios de las series temporales, con los conocimientos básicos en estadística.

4. Chou, Ya-Lun (1977). Análisis estadístico. Editorial Interamericana, S.A.

Contiene dos capítulos sobre series de tiempo. En uno trata brevemente algunas clases de tendencias y en el otro, el problema de la pronosticación. Es un libro de muy fácil lectura.

5. Doan, Thomas A y Litterman, Robert B. (1986). RATS, versión 2.00.

6. Fuller, W.A. (1976). Introduction to statistical time series. N.Y. : John Wiley and Sons.

Se trata de un enfoque teórico y matemáticamente riguroso, del análisis de series temporales. La lectura de este libro es difícil (se requiere buena formación en matemática y en estadística).

6. Granger, G.W.J. (1980). Forecasting in Business and Economics. Academic Press, N.Y.

Contiene una exposición muy clara sobre algunas clases de tendencias y procedimientos para removerla. Es un libro de muy fácil lectura, destinado a una gama muy amplia de lectores.

7. Granger, C.W., y Hatanaka, M. (1964). Spectral Analysis of Economic Time Series. Princeton University Press, Princeton, N.J.

Contiene un estudio de las técnicas del análisis espectral con referencia constante a series económicas (precios, producción, etc.). Específicamente, concede especial importancia al fenómeno de la no estacionaridad de una serie y al análisis de las tendencias. Su lectura presenta un grado de dificultad intermedio.

8. Jenkins, G.M., y Watts, D.G. (1968). Spectral Analysis and its Aplicatios. Holden - Day, San Francisco.

Su contenido es casi por completo, una lista de fórmulas y resultados del análisis espectral. Contiene pocos ejemplos y pocas aplicaciones.

9. Priestley, M.B. (1981). Spectral Analysis and Time Series, vol 1. Academic Press.

Es un enfoque teórico - matemático del análisis de series de tiempo univariadas. Es de difícil lectura.

10. Statistical graphics corporation (1986). STATGRAPHICS (Stat. graphics system). USER'S GUIDE. Stsc, Inc., USA.

11. Vandaele, W. (1983). Applied Time Series and Box - Models. Academic Press. INC. N.Y.

El tema del libro es los modelos del tipo Box-Jenkins. Hace una breve reseña sobre algunas formas de remover la tendencia y estabilizar la varianza. Por la forma de escritura poco matemática y la abundancia de ejemplos y aplicaciones, la obra está dirigida a una amplia gama de usuarios de las series temporales.

12. Wonnacott, Thomas y Wonnacott, Ronald (1981). Fundamentos de Estadística para administración y Economía. Editorial Limusa, México.

Contiene un tratamiento a nivel muy básico, sobre algunos temas tales como estacionaridad, tendencias, ciclos, análisis espectral, etc. Es de muy fácil lectura.

APLICACIONES DEL ANALISIS ESPECTRAL

1. Adelman, I., " Long Cycles - Fact or artifact? ". American Economic Review, Vol. 55, 1965.

2. Bengtsson, T. y Jorberg, L., " Market Integration in Sweden during the 18th and 19th centuries ". Economy and History, Vol. 28, 1975.

3. Eagly, R.V. y Smith, V.K., " Domestic and International Integration of the London Money Market 1731 - 1789 ". Journal of Economic History, Vol. 36, N° 1, March 1976.

4. Harkness, J.P., " A Spectral Analytic test of the Long - Swing hypothesis in Canada ". Review of Economics and Statistics, Vol. 2, 1968.

5. Hatanaka, M. y Howrey, E.P., "Low frequency variations in economic time series ". Kyclos, Vol. 22, 1969.

6. Klotz, B.P. y Neal, L., " Spectral and Cross - Spectral Analysis of the Long - Swing Hypothesis ". Review of Ecomics and Statistics, Vol. LV, 1973.

Se terminó de imprimir en la Oficina de Publicaciones de la Universidad de Costa Rica en el mes de junio de 1990. Su edición consta de 500 ejemplares.

Ciudad Universitaria Rodrigo Facio
San José, Costa Rica.

