



INSTITUTO POLITÉCNICO NACIONAL
ESCUELA SUPERIOR DE FÍSICA Y MATEMÁTICAS

Servicio Social

Angelica Garcia Leon

Tema:

Aplicaciones de las matrices de Toeplitz
al estudio de procesos estocásticos estacionarios

Programa de Servicio Social:

Apoyo al proceso de aprendizaje y/o a la investigación
en Matemáticas en la ESFM

Proyecto de investigación:

IPN-SIP 20140639

Diagonalización de algunas clases de matrices y operadores de Toeplitz

Director del proyecto de investigación:

Dr. Egor Maximenko

MÉXICO, D.F.

Diciembre 2014

Índice

Reporte global	3
1. Algunos conceptos y notaciones generales	5
1.1. Conceptos de álgebra lineal	5
1.2. Sucesiones de matrices asintóticamente equivalentes	7
1.3. Teorema límite de Szegő sobre matrices de Toeplitz	8
1.4. Conceptos de probabilidad	10
2. Artículos o capítulos de libros relacionados con aplicaciones de las matrices de Toeplitz	14
2.1. Función de autocorrelación y espectro de procesos estacionarios	15
2.2. Aplicaciones a series de tiempo estocásticas	22
2.3. Formas bilineales en variables aleatorias gaussianas y matrices Toeplitz	29
2.4. Análisis de componentes principales	33
2.5. Métodos espectrales para análisis de componentes principales de los eventos relacionados con los potenciales cerebrales	40
2.6. Una nota sobre generación, estimación y predicción de procesos estacionarios	44
2.7. Distribución asintótica de valores propios de matrices de bloque Toeplitz y aplicaciones	47
2.8. Propiedades de vectores propios de matrices de Toeplitz y su aplicación al análisis espectral de series de tiempo	52
3. Ejercicios	59

Reporte global

Justificación

La matriz de covarianza es de fundamental importancia en muchos aspectos de estadística incluyendo análisis multivariado, análisis de componentes principales, análisis de discriminantes lineales y modelamiento de gráficas. En el contexto de análisis de series de tiempo estacionarias la matriz de covarianza resulta ser una matriz de Toeplitz.

El presente trabajo apoya al proyecto de investigación IPN-SIP 20140639 presentando algunas aplicaciones de las matrices de Toeplitz y sus propiedades a procesos estocásticos estacionarios mediante resúmenes de artículos o capítulos de libros. Se espera que en este trabajo alumnos o investigadores interesados en dichas aplicaciones encuentren referencias y apoyo para comenzar un estudio más profundo en los temas mencionados en este reporte de servicio social.

Objetivos

- Investigar que aplicaciones de matrices Toeplitz existen en estadística y procesos estocásticos.
- Leer artículos o capítulos de libros sobre el tema y escribir sus resúmenes.
- Brindar información, tal como definiciones y teoremas, que pueden facilitar el estudio y comprensión de los resúmenes presentados.

Marco teórico

Los primeros estudios de la estimación de la minimización de la media al cuadrado en procesos estocásticos fueron hechos por Kolmogorov, Krein y Wiener durante las décadas tercera y cuarta del siglo pasado, Kolmogorov en 1939, Krein en 1945 y Wiener en 1949.

Kolmogorov inspirado por el trabajo de Wold en procesos estacionarios discretos en tiempo, desarrolló un tratamiento comprensivo del problema de predicción lineal para procesos estocásticos discretos en tiempo. Krein notó la relación de los resultados de Kolmogorov con un trabajo por Szegő para polinomios ortogonales (Szegő, 1939; Grenander y Szegő, 1958) y extendió los resultados al tiempo continuo con el uso de una transformación bilineal.

Wiener consideró el problema de filtrado para estimar un proceso contaminado por ruido. La formulación explícita para la estimación óptima requirió de la solución de una ecuación integral conocida como la ecuación Wiener–Hopf (Wiener y Hopf, 1931).

En 1947, Levinson formuló el problema de filtrado de Wiener en tiempo discreto. En el caso de señales de tiempo discreto, la ecuación de Wiener–Hopf en forma matricial es descrita como

$$\mathbf{R}\mathbf{w}_0 = \mathbf{p}, \tag{1}$$

donde \mathbf{w}_0 es el vector de pesos de ajuste del filtro óptimo de Wiener, \mathbf{R} es la matriz de correlación de las entradas, y \mathbf{p} es el vector de correlación cruzada entre las entradas y la respuesta deseada. Para entradas estacionarias, la matriz de correlación \mathbf{R} asume una

estructura especial, a saber, es una *matriz de Toeplitz*. Explotando las propiedades de una matriz de Toeplitz, Levinson derivó un elegante procedimiento recursivo para resolver la forma matricial de la ecuación de Wiener–Hopf. En 1960, Durbin redescubrió el procedimiento recursivo de Levinson como un esquema para el ajuste de modelos autorregresivos a series de tiempo escalares. El problema considerado por Durbin es un caso especial de la ecuación (1), en la cual, el vector columna \mathbf{p} incluye los mismos elementos encontrados en la matriz de correlación \mathbf{R} . En 1963, Whittle mostró que hay una relación cerrada entre el método de recursión de Levinson–Durbin y los polinomios ortogonales de Szegő y también derivó una generalización multivariable de la recursión de Levinson–Durbin.

Desarrollo

Para poder cumplir con los objetivos del trabajo de investigación encontré 8 artículos o capítulos de libros sobre el tema. El trabajo final de servicio social que he escrito lo dividí en tres secciones. En la primera de ellas escribí definiciones y teoremas referentes al álgebra lineal, matrices de Toeplitz y algunas de sus propiedades, conceptos de probabilidad y estadística. La finalidad de esta sección es mostrar los conceptos básicos que se requieren para una comprensión, aunque no muy profunda, de los resúmenes que escribí. Como ya mencioné realicé la búsqueda de 8 artículos o capítulos de libros en los que se presenten aplicaciones de matrices de Toeplitz y algunas de sus propiedades. Los resúmenes los he escrito en la segunda sección de este trabajo. En la sección 3 he escrito ejercicios simples, cuyo fin es orientar a los interesados en dichos temas.

Por último escribí una presentación sobre la forma matricial en la que se pueden escribir los procesos Autorregresivos, así como algunas propiedades que satisfacen los procesos Autorregresivos de orden uno.

Conclusiones

Haber hecho el servicio social participando en el proyecto de investigación me dejó muchas experiencias y enseñanzas muy útiles. Pude desarrollar mi capacidad de investigación y búsqueda de artículos y textos científicos. También desarrollé la capacidad de leer y tratar de comprender el contenido de artículos, incluso esto me ha permitido identificar cuál es su estructura. Al elaborar los resúmenes de los artículos puede percatarme de que no es fácil escribir de manera adecuada textos matemáticos, pero al invertir mucho tiempo pude mejorar mi redacción. Además, aprendí a utilizar varias herramientas de L^AT_EX, no sólo para escribir texto, sino también para escribir presentaciones y hacer gráficas. Hubo muchos conceptos que desconocía así como temas que no comprendía a profundidad, pero con el trabajo de servicio social logré condensar y enlazar algunos conocimientos de álgebra lineal, probabilidad y estadística.

Espero que este proyecto aporte algo al aprendizaje e investigación, no sólo en la ESFM, sino a otros alumnos e investigadores interesados en aplicaciones de matrices de Toeplitz. Este texto está en línea en el acceso gratuito:

http://esfm.egormaximenko.com/Garcia_Leon_2014_social_service_es.pdf

1. Algunos conceptos y notaciones generales

En esta sección se dan definiciones, proposiciones, teoremas y notaciones que se usan en varias partes del texto.

1.1. Conceptos de álgebra lineal

Dada una matriz cuadrada A , usaremos A^\top para denotar la matriz transpuesta, A^* para denotar la matriz transpuesta conjugada y $\text{tr } A$ para denotar la traza de A .

■ Formas bilineales

Sea V un espacio vectorial sobre un campo \mathbb{K} . Una función $f : V \times V \rightarrow \mathbb{K}$, es una *forma bilineal* si es lineal en cada una de sus entradas, es decir, si cumple:

$$f(\alpha u_1 + \beta u_2, v) = \alpha f(u_1, v) + \beta f(u_2, v),$$

$$f(u, \gamma v_1 + \mu v_2) = \gamma f(u, v_1) + \mu f(u, v_2),$$

para todo $u, u_1, u_2, v, v_1, v_2 \in V$ y para todo $\alpha, \beta, \gamma, \mu \in \mathbb{K}$.

■ Valores y vectores propios

Dada una matriz A cuadrada, un número real λ se dice ser un *valor propio* de A si existe un vector no nulo \mathbf{x} tal que $A\mathbf{x} = \lambda\mathbf{x}$, el vector \mathbf{x} se llama *vector propio* asociado al valor propio λ .

■ Conjunto ortonormal

Un conjunto de vectores $\{v_1, \dots, v_n\}$, en un espacio vectorial V , se dice ser *ortogonal* si $\langle v_i, v_j \rangle = 0$ para $i \neq j$ con $i, j \in \{1, \dots, n\}$. Además, si cada vector del conjunto ortogonal $\{v_1, \dots, v_n\}$ cumple que $\|v_i\| = \sqrt{\langle v_i, v_i \rangle} = 1$, $i \in \{1, \dots, n\}$, entonces el conjunto se llama *ortonormal*.

■ Matriz definida positiva

Una matriz A es *definida positiva* si para todo vector $\mathbf{x} \neq 0$, $\mathbf{x}^\top A \mathbf{x} > 0$, donde \mathbf{x}^\top denota el vector transpuesto de \mathbf{x} .

■ Matriz simétrica

Una matriz A es *simétrica* si $A = A^\top$.

■ Matriz Hermitiana

Una matriz A , cuyos elementos son números complejos, es *Hermitiana* si $A = A^*$.

■ Matriz persimétrica

Una matriz A de tamaño $n \times n$, es *persimétrica* si $a_{ij} = a_{n-j+1, n-i+1}$ para todos los $i, j \in \{1, \dots, n\}$.

■ Matriz diagonalizable

Una matriz A de tamaño $n \times n$ se dice que es una *matriz diagonalizable* si existe una matriz \mathbf{P} , de tamaño $n \times n$, invertible tal que $\mathbf{P}^{-1}A\mathbf{P} = \mathbf{D}$, donde \mathbf{D} es una matriz diagonal.

- **Matriz de Toeplitz**

Sea f una función compleja definida en $[0, 2\pi]$ de la clase de Wiener con coeficientes de Fourier t_k . Definimos la *matriz de Toeplitz*

$$T_n(f) = [t_{k,j}]_{j,k=0}^{n-1} = [t_{k-j}]_{j,k=0}^{n-1} = \left[\frac{1}{2\pi} \int_0^{2\pi} f(\lambda) e^{-i(k-j)\lambda} d\lambda \right]_{j,k=0}^{n-1}. \quad (2)$$

- **Matriz circulante**

Una *matriz circulante* es un tipo especial de matriz de Toeplitz en donde cada vector fila se rota un elemento a la derecha con respecto al vector de fila precedente. Así pues una matriz circulante C de tamaño $n \times n$ tiene la siguiente forma

$$C = \begin{bmatrix} c_0 & c_{n-1} & \cdots & c_2 & c_1 \\ c_1 & c_0 & c_{n-1} & & c_2 \\ \vdots & c_1 & c_0 & \ddots & \vdots \\ c_{n-2} & & \ddots & \ddots & c_{n-1} \\ c_{n-1} & c_{n-2} & \cdots & c_1 & c_0 \end{bmatrix}.$$

En otras palabras, cada vector columna se obtiene de la columna anterior por medio de un desplazamiento cíclico hacia abajo.

- **Valores singulares de una matriz**

Para cualquier matriz A de tamaño $m \times n$, con $m \geq n$, la matriz cuadrada A^*A de tamaño $n \times n$ es Hermitiana y definida no negativa. Los valores propios ν_1, \dots, ν_n de A^*A son reales y no negativos. Los números $s_1 = \sqrt{\nu_1}, \dots, s_n = \sqrt{\nu_n}$ son los *valores singulares* de A .

- **Descomposición en valores singulares**

Cualquier matriz A de tamaño $m \times n$, con $m \geq n$, tiene una factorización de la forma

$$A = UDV^\top,$$

donde U es una matriz de tamaño $m \times n$ con columnas ortogonales, V es una matriz ortogonal de tamaño $n \times n$ y D es una matriz diagonal de tamaño $n \times n$ con entradas no negativas. Este resultado se llama *descomposición en valores singulares*. Es fácil ver que las entradas diagonales de la matriz D son los valores singulares de A .

- **Descomposición de Cholesky**

Cada matriz A simétrica definida positiva se puede descomponer como el producto LL^* de una matriz triangular inferior L y su traspuesta conjugada.

1.2. Sucesiones de matrices asintóticamente equivalentes

Definición 1. Sea $A = [a_{j,k}]_{j,k=1}^n$ una matriz y sean s_1, \dots, s_n sus valores singulares. Se define la *norma fuerte* $\|A\|$ por

$$\|A\| = \max_{z: z^*z=1} (z^* A^* A z)^{1/2} = \max_{1 \leq k \leq n} s_k.$$

La *norma de Frobenius* (llamada también la *norma de Hilbert–Schmidt*) se define como la norma euclidiana de la matriz A convertida en un vector de longitud n^2 :

$$\|A\|_F = \left(\sum_{j=1}^n \sum_{k=1}^n |a_{j,k}|^2 \right)^{1/2} = \sqrt{\text{tr}(A^* A)} = \left(\sum_{k=1}^n s_k^2 \right)^{1/2}.$$

En este texto será cómodo usar la *norma débil* definida como la norma de Frobenius dividida entre \sqrt{n} :

$$|A| = \left(\frac{1}{n} \sum_{k=1}^n \sum_{j=1}^n |a_{k,j}|^2 \right)^{1/2} = \sqrt{\frac{1}{n} \text{tr}(A^* A)} = \left(\frac{1}{n} \sum_{k=1}^n s_k^2 \right)^{1/2}.$$

Definición 2 (sucesiones de matrices asintóticamente equivalentes). Dos sucesiones de matrices $(A_n)_{n=1}^\infty$ y $(B_n)_{n=1}^\infty$, de tamaño $n \times n$, se dicen ser *asintóticamente equivalentes* si

- (1) A_n y B_n son uniformemente acotadas en la norma fuerte, esto es, existe un número $M \geq 0$ tal que

$$\|A_n\|, \|B_n\| \leq M < \infty, \quad n = 1, 2, \dots$$

- (2) $A_n - B_n$ tiende a cero en la norma débil cuando $n \rightarrow \infty$:

$$\lim_{n \rightarrow \infty} |A_n - B_n| = 0.$$

La equivalencia asintótica de las sucesiones $(A_n)_{n=1}^\infty$ y $(B_n)_{n=1}^\infty$ se abrevia $A_n \sim B_n$.

Teorema 1 (propiedades de sucesiones de matrices asintóticamente equivalentes). Sean $(A_n)_{n=1}^\infty$, $(B_n)_{n=1}^\infty$, $(C_n)_{n=1}^\infty$ y $(D_n)_{n=1}^\infty$ sucesiones de matrices.

- (1) Si $A_n \sim B_n$, y existe el límite de las normas débiles de A_n o de B_n , entonces existen ambos límites y son iguales:

$$\lim_{n \rightarrow \infty} |A_n| = \lim_{n \rightarrow \infty} |B_n|.$$

- (2) Si $A_n \sim B_n$ y $B_n \sim C_n$, entonces $A_n \sim C_n$.

- (3) Si $A_n \sim B_n$ y $C_n \sim D_n$, entonces $A_n C_n \sim B_n D_n$.

- (4) Si $A_n \sim B_n$ y $\|A_n^{-1}\|, \|B_n^{-1}\| \leq K < \infty$, para cada n , entonces $A_n^{-1} \sim B_n^{-1}$.

- (5) Si $A_n B_n \sim C_n$ y $\|A_n^{-1}\| \leq K < \infty$, para cada n , entonces $B_n \sim A_n^{-1} C_n$.

Observación 1. Si $(A_n)_{n=1}^\infty$ y $(B_n)_{n=1}^\infty$ son dos sucesiones de matrices asintóticamente equivalentes, entonces sus normas son uniformemente acotadas por $M > 0$ y por lo tanto todos sus valores propios están en el intervalo $[-M, M]$.

1.3. Teorema límite de Szegő sobre matrices de Toeplitz

Teorema 2 (de Weierstrass sobre la aproximación uniforme por polinomios). *Si F es una función compleja continua sobre $[a, b]$, entonces existe una sucesión de polinomios p_n tal que*

$$\lim_{n \rightarrow \infty} \sup_{x \in [a, b]} |p_n(x) - F(x)| = 0.$$

Teorema 3. *Sean $(A_n)_{n=1}^{\infty}$ y $(B_n)_{n=1}^{\infty}$ sucesiones de matrices Hermitianas asintóticamente equivalentes. Supongamos que existe un número $M > 0$ tal que $A_n \geq MI_n$ y $B_n \geq MI_n$ para cada n , esto es,*

$$x^* A_n x \geq M \|x\|^2, \quad x^* B_n x \geq M \|x\|^2 \quad (x \in \mathbb{C}^n).$$

Entonces, si existe alguno de los siguientes límites, existe también el otro y son iguales:

$$\lim_{n \rightarrow \infty} (\det A_n)^{1/n} = \lim_{n \rightarrow \infty} (\det B_n)^{1/n}.$$

En los teoremas de esta sección suponemos que $f: [0, 2\pi] \rightarrow \mathbb{C}$ es una función de la clase de Wiener, esto es, sus coeficientes de Fourier t_k forman una serie absolutamente convergente:

$$\sum_{k=-\infty}^{+\infty} |t_k| < +\infty.$$

Denotamos por m_f y M_f al ínfimo (mínimo) y supremo (máximo) de los valores de f :

$$m_f = \inf_{\lambda \in [0, 2\pi]} f(\lambda), \quad M_f = \sup_{\lambda \in [0, 2\pi]} f(\lambda).$$

Para cada $n \in \{1, 2, 3, \dots\}$ denotemos por $\tau_{n,1}, \dots, \tau_{n,n}$ a los valores propios de la matriz $T_n(f)$.

Teorema 4. *Para cualquier s cualquier entero positivo,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \tau_{n,k}^s = \frac{1}{2\pi} \int_0^{2\pi} f(\lambda)^s d\lambda.$$

Además, si f es real, o equivalentemente, las matrices $T_n(f)$ son todas Hermitianas, entonces para cualquier función continua F sobre $[m_f, M_f]$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} F(\tau_{n,k}) = \frac{1}{2\pi} \int_0^{2\pi} F(f(\lambda)) d\lambda.$$

Teorema 5. *Si f es real, entonces para cualquier función continua F sobre $[\phi, \theta] \subset [m_f, M_f]$*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} F(\min\{\phi, \tau_{n,k}, \theta\}) = \frac{1}{2\pi} \int_0^{2\pi} F(\min\{\phi, f(\lambda), \theta\}) d\lambda. \quad (3)$$

Teorema 6. Supongamos que $f \geq 0$ y la igualdad $f(\lambda) = 0$ se tiene en un conjunto de puntos a lo mas numerable. Entonces

a) $T_n(f)$ es no singular

b) Si $f(\lambda) \geq m_f > 0$, entonces

$$T_n(f)^{-1} \sim C_n(f)^{-1},$$

donde $C_n(f)$ es una matriz circulante con fila superior $(c_0^{(n)}, \dots, c_{n-1}^{(n)})$, y

$$c_k^{(n)} = \frac{1}{n} \sum_{m=0}^{n-1} f(2\pi m/n) e^{2\pi i mk/n}.$$

para $k, m = 0, 1, \dots, n-1$. Además si definimos $T_n(f) - C_n(f) = D_n$, entonces $T_n(f)^{-1}$ tiene la expansión

$$\begin{aligned} T_n(f)^{-1} &= [C_n(f) + D_n]^{-1} \\ &= C_n(f)^{-1} [I + D_n C_n(f)^{-1}]^{-1} \\ &= C_n(f)^{-1} [I + D_n C_n(f)^{-1} + (D_n C_n(f)^{-1})^2 + \dots], \end{aligned}$$

la expansión converge (en la norma débil) para n suficientemente grande.

c) Si $f(\lambda) \geq m_f > 0$, entonces

$$T_n(f)^{-1} \sim T_n(1/f) = \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{i(k-j)\lambda}}{f(\lambda)} d\lambda \right]_{k,j=0}^{n-1},$$

esto es, si el espectro es estrictamente positivo, entonces la sucesión de las matrices de Toeplitz inversas es asintóticamente de Toeplitz. Además si $\rho_{n,k}$ son los valores propios de $T_n(f)^{-1}$ y F es una función continua en $[1/M_f, 1/m_f]$, entonces

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} F(\rho_{n,k}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(1/f(\lambda)) d\lambda.$$

d) Supongamos que $m_f = 0$ y que la derivada de $f(\lambda)$ existe y esta acotada para toda λ . Entonces $T_n(f)^{-1}$ no está acotada, $1/f(\lambda)$ no es integrable y por lo tanto $T_n(1/f)$ no está definida y la integral (3) tal vez no exista. Para cualquier θ finito, sin embargo, el siguiente hecho similar es cierto:

Si F es una función continua en $[1/M_f, \theta]$, entonces

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} F(\min\{\rho_{n,k}, \theta\}) = \frac{1}{2\pi} \int_0^{2\pi} F(\min\{1/f(\lambda), \theta\}) d\lambda. \quad (4)$$

Teorema 7. Sean $T_n(f)$ y $T_n(g)$ definidas como en (2), donde f y g son dos funciones en la clase de Wiener. Definimos $C_n(f)$ y $C_n(g)$ como en el Teorema 6 (b), y sean $\rho_{n,k}$ los valores propios de $T_n(f)T_n(g)$. Entonces $T_n(f)T_n(g) \sim C_n(f)C_n(g) = C_n(fg)$ y

$$T_n(f)T_n(g) \sim T_n(g)T_n(f).$$

1.4. Conceptos de probabilidad

■ Esperanza

- a) Sea X una variable aleatoria discreta que toma valores v_k con probabilidades p_k . Se dice que X tiene esperanza finita si $\sum_k p_k |v_k| < \infty$ y se define su *esperanza* como:

$$\mathbb{E}(X) = \sum_k p_k v_k.$$

- b) Sea X una variable aleatoria continua con densidad f . Decimos que X tiene esperanza finita si $\int_{-\infty}^{\infty} |x|f(x) dx < \infty$, en tal caso definimos su *esperanza* por

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

■ Momentos

Sea X una variable aleatoria discreta, y sea $r \geq 0$ un número entero. Decimos que X tiene un *momento* de orden r si X^r tiene esperanza finita. En tal caso definimos el r -ésimo momento de X como $\mathbb{E}(X^r)$.

■ Varianza

Sea X una variable aleatoria, si X tiene segundo momento finito, su *varianza* $\text{Var}(X)$ está dada por

$$\text{Var}(X) = \mathbb{E}((X - \mathbb{E}X)^2) = \mathbb{E}(X^2) - (\mathbb{E}X)^2.$$

■ Covarianza

Sean X y Y variables aleatorias con segundos momentos finitos, su *covarianza* está dada por

$$\text{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))).$$

■ Correlación

Dadas dos variables aleatorias X y Y , su *correlación* está dada por

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)} \sqrt{\text{Var}(Y)}}.$$

■ Función característica

Sea X una variable aleatoria continua con función de densidad f_X , se define la *función característica* de X como sigue:

$$\phi_X : \mathbb{R} \rightarrow \mathbb{C}; \quad \phi_X(t) = \mathbb{E}[e^{itX}] = \int_{-\infty}^{\infty} e^{itx} f_X(x) dx.$$

- **Cumulantes**

Sea X una variable aleatoria con función característica ϕ_X y supongamos que todos sus momentos existen. Puesto que $\phi_X(0) = 1$, en una vecindad de cero, podemos considerar su logaritmo y desarrollarlo en serie de potencias, a la función resultante se le denomina *función generatriz de cumulantes* de X .

$$K_X(t) = \log(\phi_X(t)) = \sum_{v=1}^{\infty} \kappa_v \frac{(it)^v}{v!}.$$

Al coeficiente κ_v se le llama el *cumulante* de orden v de X .

- **Contenido de energía y función de densidad de energía espectral**

Sea f una función. El *contenido de energía* E de f está definido por

$$E = \int_{-\infty}^{\infty} |f(t)|^2 dt.$$

Sea $F = \mathfrak{F}[f]$ la transformada de Fourier de f , tenemos que

$$E = \int_{-\infty}^{\infty} |f(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |F(\omega)|^2 d\omega.$$

Esta ecuación afirma que el contenido de energía de f está dado por $\frac{1}{2\pi}$ multiplicado por el área bajo la curva $|F(\omega)|^2$. Por esta razón la cantidad $|F(\omega)|^2$ se denomina espectro de energía o *función de densidad de energía espectral* de f .

- **Función de correlación**

La función

$$R_{12}(\tau) = \int_{-\infty}^{\infty} f_1(t)f_2(t - \tau) dt.$$

se conoce como *función de correlación* entre las funciones f_1 y f_2 . La función de correlación sumistría una medida de la similitud o interdependencia entre las funciones f_1 y f_2 en función del parámetro τ . Si la función de autocorrelación es cero para todo valor de τ , entonces se dice que las funciones no están correlacionadas.

Si f_1 y f_2 son idénticas, entonces la función de correlación

$$R_{11}(\tau) = \int_{-\infty}^{\infty} f_1(t)f_1(t - \tau) dt.$$

se denomina *función de autocorrelación* de f_1 .

- **Señales no correlacionadas**

Dos señales f_1 y f_2 se dice que *no están correlacionadas* si

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} f_1(t)f_2(t - \tau) dt = \left[\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} f_1(t) dt \right] \left[\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} f_2(t) dt \right].$$

- **Función de correlación promedio**

La función de *correlación promedio* de f_1 y f_2 , denotada por \bar{R}_{12} , está definida por el límite

$$\bar{R}_{12}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} f_1(t) f_2(t - \tau) dt.$$

Análogamente, la función de *autocorrelación promedio* de f , denotada por \bar{R}_{ff} , está definida por el límite

$$\bar{R}_{ff}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} f(t) f(t - \tau) dt.$$

Para funciones periódicas f_1 y f_2 (de periodo T) se tiene que

$$\bar{R}_{12}(\tau) = \frac{1}{T} \int_{-T/2}^{T/2} f_1(t) f_2(t - \tau) dt, \quad \bar{R}_{11}(\tau) = \frac{1}{T} \int_{-T/2}^{T/2} f_1(t) f_1(t - \tau) dt.$$

- **Espectro de potencia o densidad espectral de potencia**

Sea f una función, supongamos que el contenido de energía de f es finito, es decir,

$$\int_{-\infty}^{\infty} (f(t))^2 dt < \infty.$$

Para una función con esta propiedad, la potencia promedio en un intervalo de longitud T se aproxima a cero a medida que T se aproxima a infinito; de esta manera, se tiene

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} (f(t))^2 dt = 0.$$

En relación con los cálculos de ruido, es necesario considerar señales sin contenido finito de energía. En este caso, la potencia promedio de f es la cantidad

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} (f(t))^2 dt.$$

Cuando este límite existe, la cantidad

$$P(\omega) = \lim_{T \rightarrow \infty} \frac{1}{T} \left| \int_{-T/2}^{T/2} f(t) e^{-i\omega t} dt \right|^2 \quad (5)$$

se denomina *espectro de potencia* o *densidad espectral de potencia* de la función f . Aunque generalmente la función f se define como la transformada de Fourier de la

función de correlación promedio de f . De esta manera, se define

$$P(\omega) = \mathfrak{F} [\bar{R}_{ff}(\tau)] = \int_{-\infty}^{\infty} \bar{R}_{ff}(\tau) e^{-i\omega\tau} d\tau,$$

$$\bar{R}_{ff}(\tau) = \mathfrak{F}^{-1} [P(\omega)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} P(\omega) e^{i\omega\tau} d\omega.$$

■ **Proceso estocástico**

Un *proceso estocástico* se define como una colección de variables aleatorias $X(t)$, $t \in T$, definidas sobre un espacio de probabilidad común, donde T es un subconjunto de $(-\infty, \infty)$ y usualmente se considera como un parámetro del tiempo. El proceso es llamado proceso con parámetro continuo si T es un intervalo con longitud positiva y un proceso con parámetro discreto si T es un subconjunto de los enteros. Si $T = \{0, 1, 2, \dots\}$ usualmente denotamos el proceso por X_n , $n \geq 0$.

■ **Proceso estocástico de segundo orden**

Un proceso estocástico $X(t)$, $t \in T$, se llama proceso de *segundo orden* si $\mathbb{E}[X^2(t)] < \infty$ para cada $t \in T$.

■ **Proceso estacionario de segundo orden**

Decimos que un proceso $X(t)$, $-\infty < t < \infty$, es un *proceso estacionario de segundo orden* si para cada número real τ el proceso de segundo orden $Y(t)$, $-\infty < t < \infty$, definido por $Y(t) = X(t + \tau)$, tiene las mismas funciones de media y covarianza que el proceso $X(t)$.

■ **Procesos Gaussianos**

Un proceso estocástico $X(t)$, $t \in T$, se llama *Proceso Gaussiano* si cada combinación lineal finita de las variables aleatorias $X(t)$, $t \in T$, tiene distribución normal. Los procesos Gaussianos son también llamados *procesos normales*, y las variables aleatorias distribuidas normalmente a veces se dice que tienen una distribución gaussiana.

■ **Proceso de Wiener**

Un proceso estocástico $W(t)$, $-\infty < t < \infty$, que satisface las siguientes propiedades se llama *proceso de Wiener* con parámetro σ^2 :

- i) $W(0) = 0$.
- ii) $W(t) - W(s)$ tiene distribución normal con media 0 y varianza $\sigma^2(t - s)$ para $s \leq t$.
- iii) $W(t_2) - W(t_1), W(t_3) - W(t_2), \dots, W(t_n) - W(t_{n-1})$ son independientes para $t_1 \leq t_2 \leq \dots \leq t_n$.

Aquí σ^2 es una constante positiva.

2. Artículos o capítulos de libros relacionados con aplicaciones de las matrices de Toeplitz

En esta sección se presentan resúmenes de artículos o capítulos de libros relacionados con el tema de estudio, es decir, artículos o capítulos que muestran relación entre las matrices de Toeplitz y procesos estacionarios.

1. **Función de autocorrelación y espectro de procesos estacionarios**
es el resumen del Capítulo 2 (Autocorrelation function and spectrum of stationary process) del libro [1] de Box, Jenkins y Reinsel.
2. **Aplicaciones a series de tiempo estocásticas,**
resumen del Capítulo 6 (Applications to Stochastic Time Series) del libro [2] de Gray.
3. **Formas bilineales en variables aleatorias gaussianas y matrices de Toeplitz,**
resumen del artículo [3] de Avram.
4. **Análisis de componentes principales**
es el resumen del artículo [4] de Shlens.
5. **Métodos espectrales para análisis de componentes principales de los eventos relacionados con los potenciales cerebrales,**
resumen del artículo [5] de Rawlings, Rohrbaugh, Begleiter y Eckardt.
6. **Una nota sobre generación, estimación y predicción de procesos estacionarios,**
resumen del artículo [6] de Hauser, Hörmann, Kunst y Lenneis.
7. **Distribución asintótica de valores propios de matrices de Toeplitz por bloques y aplicación a la identificación de canal SIMO ciego,**
resumen del artículo [7] de Gazzah, Regalia, Delmas.
8. **Propiedades de vectores propios de matrices de Toeplitz y su aplicación al análisis espectral de series de tiempo,**
resumen del artículo [8] de Reddi.

2.1. Función de autocorrelación y espectro de procesos estacionarios

Una clase muy especial de procesos estocásticos, se llaman procesos estacionarios, se basan en la suposición de que el proceso está en un estado particular de equilibrio estadístico. Un proceso estocástico se dice ser *estrictamente estacionario* si sus propiedades no se ven afectados por un cambio de origen del tiempo, esto es, si la distribución de probabilidad conjunta asociada con m observaciones $z_{t_1}, z_{t_2}, \dots, z_{t_m}$ realizadas en cualquier conjunto de tiempos t_1, t_2, \dots, t_m , es la misma que la asociada con las m observaciones $z_{t_1+k}, z_{t_2+k}, \dots, z_{t_m+k}$, realizadas en los tiempos $t_1 + k, t_2 + k, \dots, t_m + k$.

Media y varianza de un proceso estacionario. Cuando $m = 1$, la suposición de estacionariedad implica que las variables aleatorias z_t tienen la misma distribución que no depende de t . Suponemos que están definidos los momentos del primer y segundo orden. Entonces el proceso estocástico tiene media constante

$$\mu = \mathbb{E}(z_t) \quad (6)$$

que define el nivel sobre el cual fluctúa, y una varianza constante

$$\sigma_z^2 = \mathbb{E}[(z_t - \mu)^2] \quad (7)$$

que mide su propagación sobre este nivel. Dado que la distribución es la misma para todos los tiempos t , su forma se puede deducir mediante la formación del histograma de las observaciones z_1, z_2, \dots, z_N , que constituyen la serie de tiempo observado. Además, la media μ del proceso estocástico se puede estimar por la media muestral

$$\bar{z} = \frac{1}{N} \sum_{t=1}^N z_t \quad (8)$$

de la serie de tiempo, y la varianza σ_z^2 del proceso estocástico puede estimarse por la varianza de la muestra

$$\hat{\sigma}_z^2 = \frac{1}{N} \sum_{t=1}^N (z_t - \bar{z})^2 \quad (9)$$

de la serie de tiempo.

Coefficientes de autocovarianza y autocorrelación. La suposición que el proceso es estacionario también implica que la distribución de probabilidad conjunta de z_{t_1} y z_{t_2} es la misma para todos los tiempos t_1, t_2 , que están en intervalos distintos. En particular, se deduce que la covarianza entre los valores z_t y z_{t+k} , separados por k intervalos de tiempo, debe ser la misma para todo t bajo la suposición de estacionariedad. Esta covarianza será llamada autocovarianza de orden k y se define por

$$\gamma_k = \text{Cov}[z_k, z_{t+k}] = \mathbb{E}[(z_k - \mu)(z_{t+k} - \mu)].$$

Similarmente, la *autocorrelación* de orden k es

$$\rho_k = \frac{\mathbb{E}[(z_k - \mu)(z_{t+k} - \mu)]}{\sqrt{\mathbb{E}[(z_k - \mu)^2]\mathbb{E}[(z_{t+k} - \mu)^2]}} = \frac{\mathbb{E}[(z_k - \mu)(z_{t+k} - \mu)]}{\sigma_z^2},$$

puesto que, para procesos estacionarios, la varianza $\sigma_z^2 = \gamma_0$ es la misma en el tiempo $t + k$ como en el tiempo t . Así, la autocorrelación de orden k es

$$\rho_k = \frac{\gamma_k}{\gamma_0},$$

en particular $\rho_0 = 1$.

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

$$\Gamma_n = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{n-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{n-3} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \gamma_{n-3} & \cdots & \gamma_0 \end{bmatrix} = \sigma_z^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{n-3} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & 1 \end{bmatrix} = \sigma_z^2 P_n$$

Una matriz de covarianza Γ_n de esta forma, que es simétrica con elementos constantes en cualquier diagonal, se llama *matriz de autocovarianza*, y la correspondiente matriz de correlación P_n se llama *matriz de autocorrelación*.

Funciones de autocovarianza y autocorrelación. La gráfica de γ_k en función k se llama *función de autocovarianza* $\{\gamma_k\}$ del proceso estocástico. Del mismo modo, la gráfica del coeficiente de autocorrelación ρ_k como una función de k se llama *función de autocorrelación* $\{\rho_k\}$ del proceso. Puesto que $\rho_k = \rho_{-k}$, la función de autocorrelación es necesariamente simétrica alrededor de cero, y en la práctica sólo es necesario para trazar la mitad positiva de la función. Cuando hablamos de la función de autocorrelación, a menudo se entenderá la mitad positiva. En el pasado, la función de autocorrelación a veces se llama *correlograma*.

Estimación de funciones de autocovarianza y autocorrelación. Hasta ahora sólo hemos considerado la función de autocorrelación teórica que describe un proceso estocástico conceptual. En la práctica, tenemos una serie de tiempo finita z_1, z_2, \dots, z_N , de N observaciones, de las que sólo podemos obtener estimaciones de la media μ y de las autocorrelaciones.

La media $\mu = \mathbb{E}[z_t]$ se estima con $\bar{z} = \sum_{t=1}^N z_t / N$. Es fácil ver que $\mathbb{E}[\bar{z}] = \mu$, de manera que \bar{z} es un estimador insesgado de μ . Como medida de precisión de \bar{z} como un estimador de μ , nos encontramos con 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 “para una gran muestra” de esta expresión para la varianza está dada por

$$\text{Var}[\bar{z}] = \frac{\gamma_0}{N} \left(1 + 2 \sum_{k=1}^{\infty} \rho_k \right),$$

en el sentido de que $N \text{Var}[\bar{z}] \rightarrow \gamma_0 \left(1 + 2 \sum_{k=1}^{\infty} \rho_k\right)$ cuando $N \rightarrow \infty$, suponiendo que

$$\sum_{k=-\infty}^{\infty} |\rho_k| < \infty.$$

Hay que tener en cuenta que el primer factor en $\text{Var}[\bar{z}]$, γ_0/N , es la expresión familiar para la varianza de \bar{z} obtenida a partir de muestras aleatorias independientes de tamaño N , pero la presencia de autocorrelación entre los valores z_t puede afectar sustancialmente la precisión de \bar{z} . Por ejemplo, un proceso estacionario tiene autocorrelaciones $\rho_k = \phi^{|k|}$, $|\phi| < 1$, la aproximación de una muestra grande para la varianza de \bar{z} se convierte en $\text{Var}[\bar{z}] = (\gamma_0/N)[(1 + \phi)/(1 - \phi)]$, y el segundo factor puede, obviamente, diferir sustancialmente de 1.

Los errores estándares de las estimaciones de autocorrelación. Para identificar un modelo de series de tiempo, es útil tener una comprobación aproximada de si ρ_k es efectivamente cero más allá de un cierto retraso. Para este propósito, se puede hacer uso de la siguiente expresión para la varianza aproximada del coeficiente de autocorrelación estimado de un proceso normal estacionario

$$\text{Var}[r_k] \simeq \frac{1}{N} \sum_{v=-\infty}^{\infty} (\rho_v^2 + \rho_{v+k}\rho_{v-k} - 4\rho_k\rho_v\rho_{v-k} + 2\rho_v^2\rho_k^2). \quad (10)$$

Por ejemplo, si $\rho_k = \phi^{|k|}$, donde $-1 < \phi < 1$, es decir, la función de autocorrelación amortigua exponencialmente, (10) da

$$\text{Var}[r_k] \simeq \frac{1}{N} \left[\frac{(1 + \phi^2)(1 - \phi^{2k})}{1 - \phi^2} - 2k\phi^{2k} \right]$$

y en particular

$$\text{Var}[r_1] \simeq \frac{1}{N}(1 - \phi^2)$$

Para cualquier proceso para el cual todas las autocorrelaciones ρ_v son cero para $v > q$, todos los términos, excepto el primero que aparece en el lado derecho de (10) son cero cuando $k > q$. Por lo tanto, para la varianza de la estimación de la autocorrelación r_k , en el paso k mayor que algún valor q más allá del cual la función de autocorrelación teórica puede ser considerada nula la aproximación está dada por

$$\text{Var}[r_k] \simeq \frac{1}{N} \left(1 + 2 \sum_{v=1}^q \rho_v^2\right) \quad k > q. \quad (11)$$

Para utilizar (11), en la práctica, las autocorrelaciones estimadas r_k ($k = 1, 2, \dots, q$) se sustituyen por las autocorrelaciones teóricas ρ_k , y cuando esto se hace se referirán a la raíz cuadrada de (11), como el *error estándar de paso grande*. En el supuesto de que las autocorrelaciones teóricas ρ_k son todos esencialmente cero más allá de un cierto paso planteado la hipótesis de $k = q$, el error estándar de paso grande se aproxima a la desviación estándar de r_k para pasos adecuadamente grandes ($k > q$).

Periodograma de una serie de tiempo. Otra forma de analizar una serie de tiempo se basa en la suposición de que se compone de ondas seno y coseno con diferentes frecuencias. Un dispositivo que utiliza esta idea, es el periodograma. El periodograma se utilizó originalmente para detectar y estimar la amplitud de un componente sinusoidal, de frecuencia conocida, enterrada en el ruido.

Para ilustrar el cálculo del periodograma, supongamos que el número de observaciones $N = 2q + 1$ es impar. Consideramos el ajuste del modelo en serie de Fourier

$$z_t = \alpha_0 + \sum_{i=1}^q (\alpha_i c_{it} + \beta_i s_{it}) + e_t, \quad (12)$$

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 $1/N$ asociada con la componente de onda sinusoidal i -ésima en (12) con la frecuencia f_i y el período $1/f_i = N/i$. Las estimaciones de mínimos cuadrados de los coeficientes α_0 y (α_i, β_i) serán

$$a_0 = \bar{z} \quad (13)$$

$$a_i = \frac{2}{N} \sum_{t=1}^N z_t c_{it} \quad i = 1, 2, \dots, q \quad (14)$$

$$b_i = \frac{2}{N} \sum_{t=1}^N z_t s_{it} \quad i = 1, 2, \dots, q, \quad (15)$$

puesto que $\sum_{t=1}^N c_{it}^2 = \sum_{t=1}^N s_{it}^2 = N/2$, y todos los terminos en (12) son mutuamente ortogonales sobre $t = 1, \dots, N$. El periodograma consiste entonces en los $q = (N - 1)/2$ valores

$$I(f_i) = \frac{N}{2} (a_i^2 + b_i^2) \quad i = 1, 2, \dots, q, \quad (16)$$

donde $I(f_i)$ se llama intensidad a la frecuencia f_i . Cuando N es par, establecemos $N/2$ en (14), (15), (15) y (16) y aplicamos para $i = 1, 2, \dots, q$, pero

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

$$b_0 = 0$$

y $I(f_q) = I(0.5) = N a_q^2$. Tenga en cuenta que la frecuencia más alta es de 0.5 ciclos por intervalo de tiempo, porque el período más pequeño es de 2 intervalos.

Análisis de la varianza. En un análisis de tabla de varianza asociada con la regresión ajustada (12), cuando N es impar, podemos aislar $(N - 1)/2$ pares de los grados de libertad, después de la eliminación de la media. Estos están asociados con los pares de coeficientes $(a_1, b_1), (a_2, b_2), \dots, (a_q, b_q)$ y por lo tanto con las frecuencias $1/N, 2/N, \dots, q/N$. El periodograma $I(f_i) = (N/2)(a_i^2 + b_i^2)$ se ve que es simplemente la suma de los cuadrados asociados con el par de coeficientes de (a_i, b_i) y por lo tanto con la frecuencia $f_i = i/N$ o el período $p_i = N/i$. Por lo tanto

$$\sum_{t=1}^n (z_t - \bar{z})^2 = \sum_{i=1}^q I(f_i) \quad (17)$$

Cuando N es par, hay $(N - 2)/2$ pares de grados de libertad y un solo grado de libertad futuro asociados con el coeficiente a_q . En la práctica, es poco probable que la frecuencia f de una componente sinusoidal sistemática desconocida coincida exactamente con cualquiera de las frecuencias f_i para las que han sido calculadas intensidades. En este caso el periodograma mostraría un aumento de las intensidades en las inmediaciones f .

Espectro de la muestra. La definición (16) del periodograma supone que las frecuencias $f_i = i/N$ son armónicos de la frecuencia fundamental $1/N$. A modo de introducción al espectro, descansamos este supuesto y permitimos que la frecuencia f varíe continuamente en el ciclo comprendido entre 0 y 0.5. La definición (16) del periodograma puede modificarse a

$$I(f) = \frac{N}{2}(a_f^2 + b_f^2) \quad 0 \leq f \leq \frac{1}{2} \quad (18)$$

y $I(f)$ se denomina entonces como el *espectro de la muestra*. Al igual que el periodograma, que puede ser utilizado para detectar y estimar la amplitud de una componente sinusoidal de frecuencia desconocida f enterrada en ruido y es, de hecho, una herramienta más apropiada para proponer este si se sabe que la frecuencia f no está relacionada armónicamente con la longitud de la serie. Además, proporciona un punto de partida para la teoría de análisis espectral, el espectro de la muestra $I(f)$ y la estimación de la función de autocovarianza c_k están vinculados por la relación

$$I(f) = 2 \left[c_0 + 2 \sum_{k=1}^{N-1} c_k \cos(2\pi f k) \right] \quad 0 \leq f \leq \frac{1}{2} \quad (19)$$

Es decir, el espectro de la muestra es la transformada de Fourier coseno de la estimación de la función de autocovarianza.

Espectro. El periodograma y espectro de la muestra son las herramientas adecuadas para el análisis de series de tiempo formadas por mezclas de ondas seno y coseno, a frecuencias fijas enterradas en el ruido. Sin embargo, las series de tiempo estacionarias se caracterizan por cambios aleatorios de frecuencia, amplitud y fase. Para este tipo de serie, el espectro de la muestra fluctúa violentamente y no es competente cualquier interpretación significativa. Sin embargo, supongamos que el espectro de la muestra se calculó para una serie temporal de N observaciones, que es una realización de un proceso normal estacionario. Un proceso de este tipo no tendría ningún coseno o componentes deterministas sinusoidales, pero podríamos llevar formalmente a través del análisis de Fourier y obtener valores de (a_f, b_f) para cualquier frecuencia dada f . Si se tomaron realizaciones repetidas de N observaciones del proceso estocástico, podríamos construir una población de valores a_f, b_f y $I(f)$. Por lo tanto, se podría calcular el valor medio de $I(f)$ en realizaciones repetidas de tamaño N , a saber,

$$\mathbb{E}[I(f)] = 2 \left[\mathbb{E}[c_0] + s \sum_{k=1}^{N-1} \mathbb{E}[c_k] \cos(2\pi f k) \right] \quad (20)$$

Para N grande se puede demostrar que

$$\lim_{N \rightarrow \infty} \mathbb{E}[c_k] = \gamma_k.$$

Al tomar el límite de (20) cuando N tiende a infinito, el *espectro de potencia* $p(f)$ se define por

$$p(f) = \lim_{N \rightarrow \infty} \mathbb{E}[I(f)] = 2 \left[\gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k \cos(2\pi f k) \right] \quad 0 \leq f \leq \frac{1}{2} \quad (21)$$

Observamos que de

$$\begin{aligned} |p(f)| &\leq 2 \left[|\gamma_0| + 2 \sum_{k=1}^{\infty} |\gamma_k| |\cos(2\pi f k)| \right] \\ &\leq 2 \left(|\gamma_0| + 2 \sum_{k=1}^{\infty} |\gamma_k| \right) \end{aligned} \quad (22)$$

una condición suficiente para que el espectro converja es que γ_k amortigua con suficiente rapidez para que la serie (22) converja. Dado que el espectro de potencia es la transformada de Fourier coseno de la función de autocovarianza, el conocimiento de la función de autocovarianza es matemáticamente equivalente al conocimiento del espectro, y viceversa.

Sobre la integración de (21) entre los límites 0 y $1/2$, la varianza del proceso z_t es

$$\gamma_0 = \sigma_z^2 = \int_0^{1/2} p(f) df \quad (23)$$

Por tanto, en la misma forma que el periodograma $I(f)$ muestra cómo la varianza (17) de una serie, que consiste en mezclas de senos y cosenos, se distribuye entre las diversas frecuencias armónicas distintas, el espectro $p(f)$ muestra cómo se distribuye la varianza de un proceso estocástico entre un rango continuo de frecuencias.

Función de densidad espectral. A veces es más conveniente basar la definición (21) del espectro en las autocorrelaciones ρ_k en lugar de las autocovarianzas γ_k . La función resultante

$$g(f) = \frac{p(f)}{\sigma_z^2} = 2 \left[1 + 2 \sum_{k=1}^{\infty} \rho_k \cos(2\pi f k) \right] \quad 0 \leq f \leq \frac{1}{2}$$

se llama la función de densidad espectral. Usando de (23), se ve que la función de densidad espectral tiene la propiedad

$$\int_0^{1/2} g(f) df = 1.$$

Puesto que $g(f)$ también es positiva, tiene mismas propiedades que una función ordinaria de densidad de probabilidad.

Estimación del espectro. Uno esperaría que una estimación del espectro se obtuvo a partir de (21), mediante la sustitución de las autocovarianzas teóricas γ_k con sus estimaciones c_k . Debido a (19), esto corresponde a tomar el espectro de la muestra como una estimación de $p(f)$. Sin embargo, puede demostrarse que el espectro de la muestra de una serie de tiempo estacionaria fluctúa violentamente sobre el espectro teórico. Una explicación intuitiva de este hecho es que el espectro de la muestra corresponde a la utilización de un intervalo, en el

dominio de la frecuencia, cuya anchura es demasiado pequeño. Esto es análogo al uso de un intervalo demasiado pequeño para el histograma al estimar una distribución de probabilidad normal. Mediante el uso de una estimación modificada o alisada

$$\hat{p}(f) = 2 \left[c_0 + 2 \sum_{k=1}^{N-1} \lambda_k c_k \cos(2\pi f k) \right],$$

donde los pesos λ_k están convenientemente elegidos, es posible aumentar el ancho de banda de la estimación y para obtener una estimación más suave del espectro. Como una forma alternativa de cálculo, también se puede obtener una estimación del espectro más suave que el espectro de la muestra $I(f)$ mediante la formación de un promedio ponderado de un número de valores $I(f_{i+j})$ del periodograma en una vecindad pequeña de frecuencias alrededor de una frecuencia dada f_i . Específicamente, un estimador de un periodograma suavizado de $p(f_i)$ toma la forma

$$\hat{p}(f_i) = \sum_{j=-m}^m W(f_j) I\left(f_i + \frac{j}{N}\right),$$

donde $\sum_{j=-m}^m W(f_j) = 1$, la función de ponderación simétrica $W(f_i)$ se conoce como la *ventana espectral*, y m se elige para que sea mucho menor que $N/2$.

2.2. Aplicaciones a series de tiempo estocásticas

Sea $(X_k)_{k \in I}$ un proceso aleatorio de tiempo discreto. Generalmente tomamos $I = \mathbb{Z}$. Estamos interesados en representaciones vectoriales del proceso, así definimos el vector columna $X^n = (X_0, \dots, X_{n-1})^\top$, la media está dada por $m^n = \mathbb{E}[X^n]$, usualmente asumimos que es cero por conveniencia. La matriz de covarianza, de tamaño $n \times n$, $R_n = [r_{j,k}]_{j,k=1}^n$ está definida por

$$R_n = \mathbb{E}[(X^n - m^n)(X^n - m^n)^*].$$

La matriz de covarianza es Hermitiana ya que

$$R_n^* = \mathbb{E}[(X^n - m^n)(X^n - m^n)^*]^* = \mathbb{E}[(X^n - m^n)(X^n - m^n)^*].$$

Tomando $m = 0$ se obtiene la matriz de autocorrelación.

Si la matriz R_n es de Toeplitz para cada n , digamos $R_n = T_n(f)$, entonces $r_{k,j} = r_{k-j}$ y el proceso se dice ser *débilmente estacionario*. En este caso $f(\lambda) = \sum_{k=-\infty}^{\infty} r_k e^{ik\lambda}$ es la *densidad espectral de potencia* del proceso. Si la matriz R_n no es de Toeplitz pero es asintóticamente de Toeplitz, *i.e.*, $R_n \sim T_n(f)$, entonces decimos que el proceso es asintóticamente débilmente estacionario y $f(\lambda)$ es la densidad espectral de potencia.

Procedemos a estudiar el comportamiento de dos modelos aleatorios comunes de procesos aleatorios. Por simplicidad supondremos que el proceso tiene media cero.

Procesos de media móvil

Consideramos un proceso aleatorio $(U_n)_{n=0}^{\infty}$. Si existe un proceso aleatorio $(W_k)_{k=0}^{\infty}$, independientes e idénticamente distribuidas (iid), con media cero y con varianza σ^2 ($(W_k)_{k=0}^{\infty}$ es un proceso de “ruido blanco” de tiempo discreto), tal que U_n se puede escribir como

$$U_n = \begin{cases} \sum_{k=0}^n b_k W_{n-k} = \sum_{k=0}^n b_{n-k} W_k, & n = 0, 1, 2, \dots; \\ 0, & n < 0, \end{cases} \quad (24)$$

entonces $(U_n)_{n=0}^{\infty}$ se llama *proceso de media móvil*.

Sin pérdida de generalidad suponemos que $b_0 = 1$ pues de lo contrario podemos incorporar b_0 en σ^2 . Se podría ser más general al permitir que el filtro b_k no sea causal y por lo tanto actúe en el futuro de las W_k . También se podría permitir que las W_k y las U_k se extiendan en el pasado infinito en lugar de ser inicializado.

Esto daría lugar a

$$U_n = \sum_{k=-\infty}^{\infty} b_k W_{n-k} = \sum_{k=-\infty}^{\infty} b_{n-k} W_k. \quad (25)$$

Nos limitaremos a los filtros causales por sencillez y mantendremos las condiciones iniciales ya que estamos interesados en el comportamiento del límite.

Puesto que estudiaremos el comportamiento estadístico de U_n cuando n es arbitrariamente grande, suponemos que $\sum_{k=0}^{\infty} |b_k| < \infty$ para asegurar que (24) converge.

La ecuación (24) puede reescribirse como una ecuación matricial mediante la definición de matriz de Toeplitz triangular inferior

$$B_n = \begin{bmatrix} 1 & 0 & & & 0 \\ b_1 & 1 & & & \\ b_2 & b_1 & & & \\ \vdots & b_2 & \ddots & \ddots & \\ b_{n-1} & \dots & & b_2 & b_1 & 1 \end{bmatrix}.$$

así que

$$U^n = B_n W^n.$$

Si el filtro b_n no es causal, entonces B_n no será triangular.

Puesto que la matriz de covarianza de W_k es simplemente $\sigma^2 I_n$, donde I_n es la matriz identidad de tamaño $n \times n$, tenemos que la covarianza de U_n es

$$\begin{aligned} R_U^{(n)} &= \mathbb{E}[U^n (U^n)^*] = \mathbb{E}[B_n W^n (W^n)^* B_n^*] \\ &= \sigma^2 B_n B_n^* \\ &= \left[\sigma^2 \sum_{l=0}^{\min(k,j)} b_{k-l} b_{j-l}^* \right]_{j,k=0}^{n-1}. \end{aligned}$$

La matriz $R_U^{(n)} = [r_{k,j}]_{j,k=0}^{n-1}$ no es Toeplitz. Por ejemplo, la entrada superior izquierda es 1 y la segunda entrada de la diagonal es $1 + b_1 b_1^*$. Sin embargo, como demostraremos a continuación, la sucesión $R_U^{(n)}$, se convierte en asintótica Toeplitz cuando $n \rightarrow \infty$.

Si definimos

$$b(\lambda) = \sum_{k=0}^{\infty} b_k e^{ik\lambda} \quad (26)$$

entonces

$$B_n = T_n(b) \quad (27)$$

así que

$$R_U^{(n)} = \sigma^2 T_n(b) T_n(b)^*. \quad (28)$$

Teorema 8. Sea U_n un proceso de media móvil con matriz de covarianza $R_{U_n}(n)$ dado por (27)-(28). Sean $\rho_{n,k}$ los valores propios de $R_U^{(n)}$. Entonces

$$R_U^{(n)} \sim \sigma^2 T_n(|b|^2) = T_n(\sigma^2 |b|^2), \quad (29)$$

así que U_n es asintóticamente estacionario. Si $m = \text{ess inf } \sigma^2 |b(\lambda)|^2$ y $M = \text{ess sup } \sigma^2 |b(\lambda)|^2$ y F es cualquier función continua sobre $[m, M]$, entonces

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} F(\rho_{n,k}) = \frac{1}{2\pi} \int_0^{2\pi} F(\sigma^2 |b(\lambda)|^2) d\lambda. \quad (30)$$

Si $\sigma^2 |b(\lambda)|^2 \geq m > 0$, entonces

$$R_U^{(n)-1} \sim \sigma^{-2} T_n(1/|b|^2).$$

Demostración. Puesto que $R_U^{(n)}$ es Hermitiana, el resultado se sigue de los teoremas 4, 6 y 2. \square

Si el proceso U_n se hubiese iniciado con su distribución estacionaria entonces habríamos tenido exactamente

$$R_U^{(n)} = \sigma^2 T_n(|b|^2).$$

Notemos que la densidad espectral del proceso de media móvil es $\sigma^2|b(\lambda)|^2$ y que las sumas de funciones de valores propios tienden a una integral de una función de la densidad espectral. En efecto la densidad espectral determina la función de densidad asintótica para los valores propios de $R_U^{(n)}$ y $\sigma^2 T_n(|b|^2)$.

Procesos autorregresivos

Sea W_k como se definió anteriormente, un proceso autorregresivo se define por

$$X_n = \begin{cases} -\sum_{k=1}^n a_k X_{n-k} + W_n & n = 0, 1, 2, \dots \\ 0 & n < 0 \end{cases} \quad (31)$$

Los procesos autorregresivos incluyen procesos no estacionarios, como el proceso de Wiener. La ecuación (31) puede ser reescrita como una ecuación vectorial mediante la definición de matriz triangular inferior

$$A_n = \begin{bmatrix} 1 & 0 & & & 0 \\ a_1 & 1 & & & \\ & a_1 & & & \\ \vdots & & \ddots & \ddots & \\ a_{n-1} & & \dots & a_1 & 1 \end{bmatrix}.$$

así que $A_n X^n = W^n$. Puesto que

$$R_W^{(n)} = A_n R_X^{(n)} A_n^* \quad (32)$$

y $\det(A_n) = 1 \neq 0$, A_n es no singular. Por tanto

$$R_X^{(n)} = \sigma^2 A_n^{-1} A_n^{-1*} \quad (33)$$

o

$$(R_X^{(n)})^{-1} = \sigma^{-2} A_n^* A_n \quad (34)$$

Equivalentemente, si $(R_X^{(n)})^{-1} = \{t_{k,j}\}$ entonces

$$t_{k,j} = \sum_{m=0}^{\min(k,j)} a_{m-k}^* a_{m-j}$$

A diferencia del proceso de media móvil, tenemos que la matriz de covarianza inversa es el producto de matrices Toeplitz inferiores. Definiendo

$$a(\lambda) = \sum_{k=0}^{\infty} a_k e^{ik\lambda} \quad (35)$$

tenemos que

$$\left(R_X^{(n)}\right)^{-1} = \sigma^{-2} T_n(a)^* T_n(a) \quad (36)$$

Observe que $\left(R_X^{(n)}\right)^{-1}$ es Hermitiana.

Teorema 9. *Sea X_n un proceso autorregresivo, con $\{a_k\}$ absolutamente convergente y matriz de covarianza $R_X^{(n)}$ con valores propios $\rho_{n,k}$. Entonces*

$$\left(R_X^{(n)}\right)^{-1} \sim \sigma^{-2} T_n(|a|)^2 \quad (37)$$

Si $m = \text{ess inf } \sigma^{-2}|a(\lambda)|^2$ y $M = \text{ess sup } \sigma^{-2}|a(\lambda)|^2$, entonces para cualquier función $F(x)$, sobre $[m, M]$ tenemos

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} F\left(\frac{1}{\rho_{n,k}}\right) = \frac{1}{2\pi} \int_0^{2\pi} F(\sigma^2|a(\lambda)|^2) d\lambda, \quad (38)$$

donde $1/\rho_{n,k}$ son los valores propios de $\left(R_X^{(n)}\right)^{-1}$. Si $|a(\lambda)|^2 \geq m > 0$, entonces

$$\left(R_X^{(n)}\right) \sim \sigma^2 T_n(1/|a|^2), \quad (39)$$

así que el proceso es asintóticamente estacionario.

Demostración. Demostración. 7. □

Note que si $|a(\lambda)|^2 > 0$, entonces $1/|a(\lambda)|^2$ es la densidad espectral de X_n . Si $|a(\lambda)|^2$ tiene un cero, entonces $R_X^{(n)}$ tal vez no sea asintóticamente Toeplitz y por lo tanto X_n no puede ser asintóticamente estacionario (puesto que $1/|a(\lambda)|^2$ no puede ser integrable). Por lo que, estrictamente hablando X_k no tendrá una densidad espectral. A menudo es conveniente, sin embargo, para definir $\sigma^2/|a(\lambda)|^2$ como la densidad espectral y que es útil para estudiar el valor propio de la distribución de R_n . Se puede relacionar $\sigma^2/|a(\lambda)|^2$ con los valores propios de $R_X^{(n)}$ también en este caso mediante el teorema (6(d)).

Corolario 1. *Dadas las suposiciones del Teorema, entonces para cualquier θ finito y cualquier función $F(x)$ continua sobre $[1/m, \theta]$*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} F\left(\min(\rho_{n,k}, \theta)\right) = \frac{1}{2\pi} \int_0^{2\pi} F\left(\min\left(\frac{1}{|a(\gamma)|^2}, \theta\right)\right) d\lambda \quad (40)$$

Demostración. Teoremas 9 y 5. □

Si consideramos dos modelos de un proceso aleatorio, que son asintóticamente equivalentes, si sus covarianzas son asintóticamente equivalentes, entonces de los teoremas 8 y 9 tenemos el siguiente corolario.

Corolario 2. *Dadas las suposiciones de los teoremas 8 y 9, consideremos el proceso de media móvil definido por:*

$$U^n = T_n(b)W^n,$$

y el proceso autorregresivo definido por:

$$T_n(a)X^n = W^n.$$

Entonces los procesos U_n y X_n son asintóticamente equivalentes si

$$a(\lambda) = \frac{1}{b(\lambda)}.$$

Demostración. Se sigue de los teoremas 6 y 7 y

$$R_X^{(n)} = \sigma^2 T_n(a)^{-1} T_n^{-1}(a)^* \sim \sigma^2 T_n(1/a) T_n(1/a)^* \sim \sigma^2 T_n(1/a)^* T_n(1/a) \quad (41)$$

Comparando (41) con 28 completamos la demostración. \square

Los métodos anteriores también pueden ser fácilmente aplicados para estudiar modelos lineales autorregresivos-media móvil mixtos.

Factorización de matrices de Toeplitz Hermitianas

Considere el problema del comportamiento asintótico de los factores triangulares de una sucesión de matrices de autocovarianza Hermitianas $T_n(f)$ en la clase de Wiener. Es bien sabido que cualquiera de tales matrices puede ser factorizada en el producto de una matriz triangular inferior y su transpuesta conjugada, en particular

$$T_n(f) = \{t_{k,j}\} = B_n B_n^*, \quad (42)$$

donde B_n es una matriz triangular inferior con entradas

$$b_{k,j}^{(n)} = \{\det(T_k) \det(T_{k-1})\}^{-1/2} \gamma(j, k), \quad (43)$$

donde $\gamma(j, k)$ es el determinante de la matriz T_k con la columna de la derecha sustituida por $(t_{j,0}, t_{j,1}, \dots, t_{j,k-1})'$. Note en particular que los elementos de la diagonal están dados por

$$b_{k,k}^{(n)} = \{\det(T_k) \det(T_{k-1})\}^{1/2}. \quad (44)$$

La ecuación (43) es el resultado de una eliminación Gaussiana o un proceso de Gram-Schmidt. La factorización de T_n permite la construcción de un modelo lineal de un proceso aleatorio y es útil en la identificación del sistema y otros procedimientos recursivos. Nuestra pregunta

es cómo se comporta B_n cuando n es grande, en concreto ¿Es B_n asintóticamente Toeplitz? Suponga que $f(\lambda)$ tiene la forma

$$f(\lambda) = \sigma^2 |b(\lambda)|^2,$$

con

$$b^*(\lambda) = b(-\lambda), \quad b^*(\lambda) = \sum_{k=0}^{\infty} b_k e^{ik\lambda}, \quad b_0 = 1.$$

La descomposición de una función no negativa en un producto de esta forma se conoce como *factorización de Wiener-Hopf*. Ya hemos construido funciones de esta forma cuando se consideran los modelos de media móvil y autorregresión. En un resultado clásico una condición necesaria y suficiente para que $\ln f$ tenga tal factorización es que f tenga integral finita. De (42) y el Teorema 6 tenemos

$$B_n B_n^* = T_n(f) \sim T_n(\sigma b) T_n(\sigma b)^* \quad (45)$$

Deseamos demostrar que (45) implica que

$$B_n \sim T_n(\sigma b) \quad (46)$$

Demostración. Puesto que $\det T_n(\sigma b) = \sigma^n \neq 0$, $T_n(\sigma b)$ es invertible. Asimismo, puesto que $\det B_n = [\det T_n(f)]^{1/2}$ tenemos del Teorema 6(a) que $\det T_n(f) \neq 0$, así que B_n es invertible. Luego del teorema 5 y (45) tenemos

$$T_n^{-1} B_n = [B_n^{-1} T_n]^{-1} \sim T_n^* B_n^{*-1} = [B_n^* T_n]^* \quad (47)$$

Dado que B_n y T_n ambas son matrices triangulares inferiores, también lo es B_n^{-1} y por lo tanto $B_n T_n$ y $[B_n^{-1} T_n]^{-1}$. Así (47) establece que una matriz triangular inferior es asintóticamente equivalente a una matriz triangular superior. Esto sólo es posible si ambas matrices son asintóticamente equivalentes a una matriz diagonal, digamos $G_n = \{g_{k,k}^{(n)} \delta_{k,j}\}$. Además a partir de (47) tenemos $G_n \sim G_n^{*-1}$

$$\{|g_{k,k}^{(n)}|^2 \delta_{k,j}\} \sim I_n \quad (48)$$

Puesto que $T_n(\sigma b)$ es triangular inferior con elemento en la diagonal principal σ , $T_n(\sigma b)^{-1}$ es triangular inferior, con todos los elementos de la diagonal principal iguales a $1/\sigma$ a pesar de que la matriz $T_n(\sigma b)^{-1}$ no es Toeplitz. Así $g_{k,k}^{(n)} = b_{k,k}^{(n)}/\sigma$. Puesto que $T_n(f)$ es Hermitiana, $b_{k,k}$ es real, de modo que al tomar la traza en (48) se obtiene

$$\lim_{n \rightarrow \infty} \sigma^{-2} \frac{1}{n} \sum_{k=0}^{n-1} \left(b_{k,k}^{(n)} \right)^2 = 1 \quad (49)$$

De (44) y 3, y el hecho de que $T_n(\sigma b)$ es triangular tenemos que

$$\begin{aligned} \lim_{n \rightarrow \infty} \sigma^{-1} \frac{1}{n} \sum_{k=0}^{n-1} b_{k,k}^{(n)} &= \sigma^{-1} \lim_{n \rightarrow \infty} \{(\det T_n(f))/(\det T_{n-1}(f))\}^{1/2} \\ &= \sigma^{-1} \lim_{n \rightarrow \infty} \{\det T_n(f)\}^{1/2n} \sigma^{-1} \lim_{n \rightarrow \infty} \{\det T_n(\sigma b)\}^{1/n} \\ &= \sigma^{-1} \sigma = 1 \end{aligned} \quad (50)$$

Combinando 49 y (50) tenemos

$$\lim_{n \rightarrow \infty} |B_n^{-1}T_n - I_n| = 0 \quad (51)$$

Aplicando el teorema 1 tenemos (46). □

Dado que los únicos requisitos reales para la prueba fueron la existencia de la factorización de Wiener-Hopf y el comportamiento del límite del determinante, este resultado puede extenderse fácilmente para el caso más general en que $\ln f(\lambda)$ es integrable.

2.3. Formas bilineales en variables aleatorias gaussianas y matrices Toeplitz

Se estudia el comportamiento asintótico de las formas bilineales

$$B_n = \sum_{i,j=1}^n a_{i-j} X_i X_j = \sum_{i=1}^n \sum_{j=1}^n a_{i-j} X_i X_j, \quad (52)$$

donde $(X_i)_{i=0}^{\infty}$ es una sucesión estacionaria gaussiana con media cero.

La covarianza de la sucesión X_n es

$$r_n = \mathbb{E}(X_0 X_n) \quad (53)$$

El hecho clave sobre B_n es que sus cumulantes son:

$$\text{cum}_k(B_n) = 2^{k-1} (k-1)! \text{tr}(A_n R_n)^k,$$

donde $A_n(i, j) = a_{i-j}$, $R_n(i, j) = r_{i-j}$, $i, j = 1, \dots, n$, son matrices Toeplitz $n \times n$. a_{i-j} son los coeficientes de la forma bilineal B_n y r_{i-j} son las covarianzas de la sucesión X_n . Para estudiar B_n , se debe investigar el comportamiento asintótico de la traza de productos de matrices de Toeplitz.

Sean $\hat{f}_k^{(\nu)} = \int_0^1 e^{2\pi i k x} f^{(\nu)}(x) dx$, $\nu = 1, \dots, s$ las sucesiones de los coeficientes de Fourier

de s funciones con valores complejos $f^{(\nu)}(x)$, $\nu = 1, \dots, s$, y sean $T_n(f^{(\nu)})$, $\nu = 1, \dots, s$ las correspondientes matrices de Toeplitz $n \times n$, i.e.:

$$T_n(f^{(\nu)})(i, j) = \hat{f}_{i-j}^{(\nu)} \quad \text{para } i, j = 1, \dots, n.$$

Sea \mathcal{L}_p la cerradura de los polinomios trigonométricos en el espacio $L_p[0, 1]$, para $1 \leq p \leq \infty$.

Teorema 10. *Supongamos que $f^{(\nu)}(x) \in \mathcal{L}_{p_\nu}$, con $1 \leq p_\nu \leq \infty$, para $\nu = 1, \dots, s$,*

a) *Si $\sum_{\nu=1}^s (p_\nu)^{-1} \leq 1$ entonces*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{tr} \left(\prod_{\nu=1}^s T_n \left(f^{(\nu)} \right) \right) = \int_0^1 \prod_{\nu=1}^s \left(f^{(\nu)}(x) \right) dx \quad (54)$$

b) *Si $\alpha > 1$, y $\alpha \geq \sum_{\nu=1}^s (p_\nu)^{-1}$, entonces*

$$\lim_{n \rightarrow \infty} \frac{1}{n^\alpha} \text{tr} \left(\prod_{\nu=1}^s T_n \left(f^{(\nu)} \right) \right) = 0 \quad (55)$$

Nota. La fórmula 54 se obtuvo bajo la suposición de que $f^{(\nu)}(x)$ está acotada.

Definición. Los valores singulares de la matriz A son valores propios de la matriz $(AA^*)^{1/2}$, donde A^* denota la adjunta de A .

Corolario. Sea $f \in L_\infty(0, 1)$ sea

$$M = \sup_x |f(x)|$$

y sea F una función continua, $F : [0, M] \rightarrow R$. Entonces

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n F(s_{j,n}) = \int_0^1 F(|f(x)|) dx,$$

donde $s_{j,n}$ son los valores singulares de $T_n(f)$.

La demostración se sigue del Teorema 10 a, tomando $f^{2\nu-1} = f$, $f^{2\nu} = \bar{f}$.

Teorema 11. Sean a_k y r_k en (52) y (53) los coeficientes de Fourier reales, de funciones pares $a(x)$ y $r(x)$, y supongamos $a(x) \in \mathcal{L}_{p_1}$, $r(x) \in \mathcal{L}_{p_2}$, $1 \leq p_1, p_2 \leq \infty$ y $(p_1)^{-1} + (p_2)^{-1} \leq 2^{-1}$. Entonces,

$$\frac{B_n - E(B_n)}{\sqrt{n}} \xrightarrow{d} N(0, \sigma^2), \quad (56)$$

donde

$$\sigma^2 = \int_0^1 a^2(x)r^2(x)dx.$$

Para demostrar este teorema se utiliza el método de cumulantes:

$$\begin{aligned} cum_k \left(\frac{B_n - EB_n}{\sqrt{n}} \right) &= \begin{cases} 0 & \text{para } k = 1 \\ 2 \operatorname{tr}(A_n R_n)^2 / n & \text{para } k = 2 \\ 2^{k-1} (k-1)! \frac{\operatorname{tr}(A_n R_n)^k}{n^{k/2}} & \text{para } k \geq 3 \end{cases} \\ \xrightarrow{n \rightarrow \infty} &\begin{cases} 0 & \text{para } k = 1 \\ 2 \int_0^1 a^2(x)r^2(x)dx & \text{para } k = 2 \text{ por el Teorema 1a} \\ 0 & \text{para } k \geq 3 \text{ por el Teorema 1b.} \end{cases} \end{aligned}$$

Cuando $a(x)$, $r(x)$ son continuas, la transformada de Legendre del tipo de desviaciones grandes B_n/n también se pueden obtener de forma explícita.

Para la demostración del teorema 10 se necesitan algunas definiciones y propiedades.

Definición. Sea s_j que denota los valores singulares de una matriz A . Para $1 \leq p \leq \infty$, la norma p -Schatten de A se define como:

$$\|A\|_p = \begin{cases} \left[\sum_j (s_j)^p \right]^{1/p} & \text{para } 1 \leq p < \infty \\ \max_j s_j & \text{para } p = \infty \end{cases}$$

Algunas propiedades necesarias de las normas Schatten son:

1. $\|AB\|_1 \leq \|A\|_p \|B\|_q$, donde $\frac{1}{p} + \frac{1}{q} = 1$
2. $|\operatorname{tr} A| \leq \|A\|_1$
3. $\|A\|_2^2 = [\operatorname{tr}(AA^*)] = \sum_{i,j} |A_{i,j}|^2$
4. $\|A\|_\infty = \max_{\|v\|_2=1} \|Av\|_2$ donde $\|v\|_2$ es la norma l_2 de v .

La demostración del teorema 10 se basa en la siguiente desigualdad para las normas Schatten de matrices de Toeplitz:

Lema 1. Para $1 \leq p \leq \infty$, $\|T_n(f)\|_p \leq n^{1/p} \|f\|_p$.

Demostración. Para establecer la desigualdad es suficiente demostrar los casos $p = 1$ y $p = \infty$. Para el caso $p = \infty$, se considera un operador $\bar{T}(f) : L_2 \rightarrow L_2$ tal que $\bar{T}(f)(g) = fg$ y por lo tanto

$$\|\| T(f) \|\| = \|\| \bar{T}(f) \|\| = \|f\|_\infty,$$

entonces

$$\| \|T_n(f)\|_\infty \| \| P_n T(f) P_n \| \leq \|\| T(f) \|\| = \|f\|_\infty.$$

Caso $p = 1$. Aquí se descompone $f \in L_1$ como $f = gh$, de tal forma que $\|f\|_1 = \|g\|_2 \|h\|_2$ y $T(f) = T(g) \cdot T(h)$. Obteniendo así

$$\begin{aligned} \|T_n(f)\|_1 &= \|P_n T(g) T(h) P_n\|_1 \leq \|P_n T(g)\|_2 \|T(h) P_n\|_2 \\ &= n^{1/2} \|g\|_2 \cdot n^{1/2} \|h\|_2 = n \|f\|_1 \quad \square \end{aligned}$$

Demostración del Teorema 10

- a) Se usa inducción sobre m (el número de $f^{(\nu)}$ en (54) que no son polinomios). Para $m = 0$ es suficiente comprobar el caso $f^{(\nu)} = e^{2\pi i k \nu x}$. Supongamos que se tiene (54) cuando tenemos a lo más m no-polinomios. Consideremos un conjunto de $f^{(\nu)}(x)$ que tiene a lo más $m+1$ no-polinomios, sin pérdida de generalidad supongamos que $f^{(1)}(x)$ es un no-polinomio. $f_k^{(1)}(x)$ denota la k -ésima suma de Fejer de $f^{(1)}(x)$ y $f^{(1),k}(x) = f^{(1)}(x) - f_k^{(1)}(x)$ sea el k -ésimo resto.

Entonces

$$\lim_{n \rightarrow \infty} \frac{1}{n} \operatorname{tr} \left[T_n(f_k^{(1)}) \prod_{\nu=2}^s T_n(f^{(\nu)}) \right] = \int_0^1 f_k^{(1)}(x) \prod_{\nu=2}^s f^{(\nu)}(x) dx \quad (57)$$

por la hipótesis de inducción. Para mostrar entonces que (54) se mantiene con un máximo de $m+1$ no-polinomios sólo queda señalar que:

$$\begin{aligned} \lim_{k \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \left| \operatorname{tr} \left[T_n(f^{(1),k}) \prod_{\nu=2}^s T_n(f^{(\nu)}) \right] \right| &\leq \lim_{k \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \left\| T_n(f^{(1),k}) \prod_{\nu=2}^s T_n(f^{(1)}) \right\|_1 \\ &\leq \lim_{k \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} \frac{n^{\left(\sum_{\nu} (p_\nu)^{-1}\right)}}{n} \|f^{(1),k}\|_{p_1} \prod_{\nu=2}^s \|f^{(\nu)}\|_{p_\nu} = 0. \end{aligned}$$

b) Se supone que $\sum_{\nu=1}^s (p_\nu)^{-1} > 1$. (De otra forma el resultado se sigue de **a**).

La prueba es similar al inciso a). Si todas las $f^{(\nu)}$ son polinomios, el límite es 0 pues $\alpha > 1$. Asumimos, sin pérdida de generalidad, que $f^{(1)}(x)$ no es un polinomio, sustituir $f^{(1)}$ por $f^{(1),k} + f_k^{(1)}$ y dividir (55) en dos partes.

El límite doble en k y n de la segunda parte es 0 por la hipótesis de inducción; para la primera parte, sea $\theta = \sum_{\nu=1}^s (p_\nu)^{-1}$ y observe que

$$\begin{aligned} & \frac{1}{n^\alpha} \left| \text{tr} \left[T_n(f^{(1),k}) \prod_{\nu=2}^s T_n(f^{(\nu)}) \right] \right| \leq \frac{1}{n^\alpha} \left\| T_n(f^{(1),k}) \right\|_{\theta p_1} \prod_{\nu=2}^s \left\| T_n(f^{(\nu)}) \right\|_{\theta p_\nu} \\ & \leq \frac{n^{\left(\sum_{\nu} (p_\nu)^{-1}\right)}}{n} \left\| f^{(1),k} \right\|_{p_1} \prod_{\nu=2}^s \left\| f^{(\nu)} \right\|_{p_\nu} \leq \left\| f^{(1),k} \right\|_{p_1} \prod_{\nu=2}^s \left\| f^{(\nu)} \right\|_{p_\nu} \xrightarrow{k \rightarrow \infty} 0. \quad \square \end{aligned}$$

2.4. Análisis de componentes principales

El análisis de componentes principales (ACP) es un método simple no paramétrico que sirve para la extracción de información relevante de conjuntos de datos. El objetivo del análisis de componentes principales es identificar la base más significativa para volver a expresar un conjunto de datos. La esperanza es que esta nueva base filtrará el ruido y revelará la estructura oculta.

Tratamos a cada muestra de tiempo como una muestra individual en nuestro conjunto de datos. Cada \vec{X} de la muestra es un vector m -dimensional, donde m es el número de tipos de medición. Es decir, cada muestra es un vector que se encuentra en un espacio vectorial m -dimensional generado por una base ortonormal. Los vectores son una combinación lineal de este conjunto base de vectores unitarios.

Dado que se esta considerando un espacio m -dimensional hay que tener en cuenta la matriz de identidad $m \times m$

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_m \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix},$$

donde cada fila \mathbf{b}_i es un vector de la base ortonormal con m componetes. Podemos considerar esta base como el punto de partida eficaz. Todos nuestros datos se registran en esta base y por lo tanto fácilmente pueden ser expresados como una combinación lineal de $\{\mathbf{b}_i\}$.

Ahora podemos establecer con mayor precisión lo que ACP se pregunta: *¿Hay alguna otra base, que sea una combinación lineal de la base original, que mejor re-exprese nuestro conjunto de datos?*

El ACP hace una suposición estricta pero de gran alcance: la linealidad.

Sea \mathbf{X} una matriz $m \times n$ que representa el conjunto de datos original, donde cada *columna* es una sola muestra (o momento en el tiempo) de nuestro conjunto de datos (es decir, \vec{X}). Sea \mathbf{Y} otra matriz $m \times n$ relacionada con una transformación lineal \mathbf{P} . \mathbf{Y} es una nueva representación de ese conjunto de datos.

$$\mathbf{PX} = \mathbf{Y}. \tag{58}$$

Definimos las siguientes cantidades, \mathbf{p}_i son las filas de \mathbf{P} , \mathbf{x}_i son las columnas de \mathbf{X} (o \mathbf{X} individual), \mathbf{y}_i son las columnas de \mathbf{Y} .

Podemos pensar que las filas de \mathbf{P} , $\{\mathbf{p}_1, \dots, \mathbf{p}_m\}$, es un conjunto nuevo de vectores base para expresar las columnas de \mathbf{X} . Esta interpretación puede verse escribiendo los productos punto explícitos de \mathbf{PX} .

$$\mathbf{PX} = \begin{bmatrix} \mathbf{p}_1 \\ \vdots \\ \mathbf{p}_m \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 & \dots & \mathbf{x}_n \end{bmatrix},$$

$$\mathbf{Y} = \begin{bmatrix} \mathbf{p}_1 \cdot \mathbf{x}_1 & \dots & \mathbf{p}_1 \cdot \mathbf{x}_n \\ \vdots & \ddots & \vdots \\ \mathbf{p}_m \cdot \mathbf{x}_1 & \dots & \mathbf{p}_m \cdot \mathbf{x}_n \end{bmatrix}.$$

Podemos notar la forma de cada columna de \mathbf{Y}

$$\mathbf{y}_i = [\mathbf{p}_1 \cdot \mathbf{x}_i \cdots \mathbf{p}_m \cdot \mathbf{x}_i]^\top.$$

Cada coeficiente de \mathbf{y}_i es un producto punto de \mathbf{x}_i con la fila correspondiente de \mathbf{P} . Es decir, el coeficiente j -ésimo de \mathbf{y}_i es una proyección sobre la j -ésima fila de \mathbf{P} . Esto es, de hecho, la propia forma de una ecuación donde \mathbf{y}_i es una proyección en la base de $\{\mathbf{p}_1, \dots, \mathbf{p}_m\}$. Por lo tanto, las filas de \mathbf{P} son un nuevo conjunto de vectores base para representar de las columnas de \mathbf{X} .

Al asumir la linealidad, el problema se reduce a encontrar el cambio adecuado de la base. Los vectores fila $\{\mathbf{p}_1, \dots, \mathbf{p}_m\}$ en esta transformación se convertirán en las componentes principales de \mathbf{X} .

¿Cuál es la mejor manera de expresar los datos? Lo que a continuación se expone nos dará una respuesta intuitiva a esta pregunta.

Ruido y Rotación. La medición del ruido en cualquier conjunto de datos debe ser baja o de lo contrario, no importa la técnica de análisis, no hay información que se pueda extraer acerca de una señal. No existe ninguna escala absoluta de ruido, pero en lugar de ruido todo se cuantifica en relación con la intensidad de la señal. Una medida común es la *razón señal-ruido (RSR)*, o una razón de varianzas σ^2 .

$$RSR = \frac{\sigma_{\text{señal}}^2}{\sigma_{\text{ruido}}^2}.$$

Una alta $RSR (\gg 1)$ indica una medición de alta precisión, mientras que una RSR baja indica datos con mucho ruido. La dinámica de interés existe a lo largo de las direcciones con mayor varianza y la más alta, presumiblemente RSR . Maximizar la varianza (y por supuesto RSR) corresponde a la búsqueda de la rotación apropiada de la base $\{\mathbf{b}_1, \dots, \mathbf{b}_m\}$.

Matriz de Covarianza. Consideremos dos conjuntos de mediciones con media cero

$$A = \{a_1, \dots, a_n\}, \quad B = \{b_1, \dots, b_n\},$$

donde el subíndice n denota el tamaño de la muestra. La varianza de A y B se define individualmente como,

$$\sigma_A^2 = \frac{1}{n} \sum_i a_i^2, \quad \sigma_B^2 = \frac{1}{n} \sum_i b_i^2.$$

La covarianza entre A y B es una generalización sencilla:

$$\text{covarianza de } A \text{ y } B \equiv \sigma_{AB}^2 = \frac{1}{n} \sum_i a_i b_i.$$

La covarianza mide el grado de la relación lineal entre dos variables. Un valor positivo grande indica que los datos están correlacionados positivamente. Del mismo modo, un valor negativo grande indica que los datos están correlacionados negativamente. La magnitud absoluta de la covarianza mide el grado de redundancia.

Algunos datos adicionales acerca de la covarianza:

- σ_{AB} es cero si y sólo si A y B no están correlacionadas.
- $\sigma_{AB}^2 = \sigma_A^2$ si $A = B$.

Podemos convertir a A y a B en vectores fila, $\mathbf{a} = [a_1 \ a_2 \ \dots \ a_n]$ y $\mathbf{b} = [b_1 \ b_2 \ \dots \ b_n]$, para expresar la covarianza como un producto punto

$$\sigma_{\mathbf{ab}}^2 \equiv \frac{1}{n} \mathbf{ab}^\top \quad (59)$$

Por último, podemos generalizar a partir de dos vectores a un número arbitrario. Cambiamos el nombre de los vectores fila \mathbf{a} y \mathbf{b} , por \mathbf{x}_1 y \mathbf{x}_2 , respectivamente, y consideramos otros vectores fila indexados $\mathbf{x}_3, \dots, \mathbf{x}_m$. Definimos una nueva matriz \mathbf{X} de tamaño $m \times n$:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_m \end{bmatrix}$$

Una interpretación de \mathbf{X} es la siguiente: Cada *fila* de \mathbf{X} corresponde a todas las mediciones de un tipo particular. Cada *columna* de \mathbf{X} corresponde a un conjunto de mediciones de un ensayo en particular. Llegamos a una definición de la *matriz de covarianza* $\mathbf{C}_\mathbf{X}$.

$$\mathbf{C}_\mathbf{X} \equiv \frac{1}{n} \mathbf{X}\mathbf{X}^\top.$$

En $\mathbf{C}_\mathbf{X}$ el ij -ésimo elemento de $\mathbf{C}_\mathbf{X}$ es el producto punto entre el vector del i -ésimo tipo de medición con el vector del j -ésimo tipo de medición. Podemos resumir varias propiedades de $\mathbf{C}_\mathbf{X}$:

- $\mathbf{C}_\mathbf{X}$ es una matriz cuadrada simétrica de tamaño $m \times m$.
- Los términos de la diagonal de $\mathbf{C}_\mathbf{X}$ son la *varianza* de los tipos de medición particulares.
- Los términos fuera de la diagonal de $\mathbf{C}_\mathbf{X}$ son la *covarianza* entre los tipos de medición.

$\mathbf{C}_\mathbf{X}$ captura la covarianza entre todos los posibles pares de mediciones. Los valores de covarianza reflejan el ruido y la redundancia en nuestras mediciones.

- En los términos diagonales, por supuesto, los valores grandes corresponden a la estructura interesante.
- Los términos fuera de la diagonal con grandes magnitudes corresponden a alta redundancia.

Imaginemos que tenemos la opción de manipular $\mathbf{C}_\mathbf{X}$. Sugestivamente definiremos nuestra matriz de covarianza manipulada $\mathbf{C}_\mathbf{Y}$. Esto podría ser realizado por un simple algoritmo:

1. Seleccionamos una dirección normalizada en el espacio m -dimensional a lo largo del cual se maximiza la varianza de \mathbf{X} . Guardamos este vector como \mathbf{p}_1 .

2. Buscamos otra dirección a lo largo de la cual se maximiza la varianza, sin embargo, debido a la condición ortonormalidad, restringimos la búsqueda de todas las direcciones ortogonales a todas las direcciones seleccionadas anteriores. Guardar este vector como \mathbf{p}_i .
3. Repetimos este procedimiento hasta que m -vectores sean seleccionados.

El conjunto ordenado de las \mathbf{p} 's que resulta, son las componentes principales. En principio, este simple algoritmo funciona, sin embargo, la verdadera razón es por que la suposición de ortonormalidad es acertada. El verdadero beneficio de este supuesto es que existe una solución eficiente, analítica a este problema.

Tenga en cuenta lo que hemos ganado con la condición de rango ordenado por la varianza. Tenemos un método para valorar la importancia de la dirección principal. Es decir, las varianzas asociadas a cada dirección \mathbf{p}_i cuantifican cada *dirección principal* por el rango ordenado de cada vector \mathbf{p}_i de la base de acuerdo con las variaciones correspondientes.

Hemos discutido todos los aspectos de la derivación del *ACP* lo que queda son las soluciones de álgebra lineal. La primera solución es un tanto sencilla, mientras que la segunda solución implica la comprensión de una descomposición algebraica importante.

Solución de *ACP* mediante descomposición de vectores propios

Como antes, el conjunto de datos es una matriz \mathbf{X} de tamaño $m \times n$, donde m es el número de tipos de medición y n es el número de muestras. Comenzamos reescribiendo \mathbf{C}_Y en términos de la variable desconocida:

$$\begin{aligned}\mathbf{C}_Y &= \frac{1}{n}\mathbf{Y}\mathbf{Y}^\top = \frac{1}{n}(\mathbf{P}\mathbf{X})(\mathbf{P}\mathbf{X})^\top \\ &= \frac{1}{n}\mathbf{P}\mathbf{X}\mathbf{X}^\top\mathbf{P}^\top = \mathbf{P}\left(\frac{1}{n}\mathbf{X}\mathbf{X}^\top\right)\mathbf{P}^\top. \\ &= \mathbf{P}\mathbf{C}_X\mathbf{P}^\top.\end{aligned}$$

Tenga en cuenta que hemos identificado la matriz de covarianza de \mathbf{X} en la última línea.

Nuestro plan es reconocer que cualquier matriz simétrica \mathbf{A} se puede diagonalizar mediante una matriz ortogonal de sus vectores propios. Para una matriz simétrica \mathbf{A} , $\mathbf{A} = \mathbf{E}\mathbf{D}\mathbf{E}^\top$, donde \mathbf{D} es una matriz diagonal y \mathbf{E} es una matriz de vectores propios de \mathbf{A} dispuestos en forma de columnas.¹

Ahora viene el truco. *Seleccionamos la matriz \mathbf{P} que es una matriz donde cada fila \mathbf{p}_i es un vector propio de $\frac{1}{n}\mathbf{X}\mathbf{X}^\top$.* En esta selección, $\mathbf{P} \equiv \mathbf{E}^\top$. Entonces $(\mathbf{P}^{-1} = \mathbf{P}^\top)$ se puede terminar la evaluación de \mathbf{C}_Y .

¹La matriz \mathbf{A} puede tener $r \leq m$ vectores propios ortonormales donde r es el rango de la matriz. Cuando el rango de \mathbf{A} es menor que m , \mathbf{A} está degenerada o todos los datos ocupan un subespacio de dimensión $r \leq m$. Manteniendo la restricción de ortogonalidad, se puede remediar esta situación mediante la selección de $(m - r)$ vectores adicionales ortonormales para "llenar" la matriz \mathbf{E} . Estos vectores adicionales no afectan a la solución final, ya que las varianzas asociadas a estas direcciones son iguales a cero.

$$\begin{aligned}
\mathbf{C}_Y &= \mathbf{P}\mathbf{C}_X\mathbf{P}^\top = \mathbf{P}(\mathbf{E}\mathbf{D}\mathbf{E}^\top)\mathbf{P}^\top = \mathbf{P}(\mathbf{P}^\top\mathbf{D}\mathbf{P})\mathbf{P}^\top \\
&= (\mathbf{P}\mathbf{P}^\top)\mathbf{D}(\mathbf{P}\mathbf{P}^\top) = (\mathbf{P}\mathbf{P}^{-1})\mathbf{D}(\mathbf{P}\mathbf{P}^{-1}) \\
\mathbf{C}_Y &= \mathbf{D}
\end{aligned}$$

Es evidente que la elección de \mathbf{P} diagonaliza \mathbf{C}_Y . Este es el objetivo de *ACP*. Podemos resumir los resultados del *ACP* en las matrices \mathbf{P} y \mathbf{C}_Y .

- Las componentes principales de \mathbf{X} son los vectores propios de $\mathbf{C}_X = \frac{1}{n}\mathbf{X}\mathbf{X}^\top$.
- El *i*-ésimo valor de la diagonal de \mathbf{C}_Y es la varianza de \mathbf{X} a lo largo de \mathbf{p}_i .

Una solución más general usando DVS

Sea \mathbf{X} una matriz arbitraria de tamaño $n \times m$. Entonces $\mathbf{X}^\top\mathbf{X}$ es una matriz cuadrada simétrica de tamaño $m \times m$. Denotemos su rango por r . Definimos las siguientes cantidades de interés:

- $\{\hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2, \dots, \hat{\mathbf{v}}_r\}$ es el conjunto *ortonormal* de vectores propios de tamaño $m \times 1$ asociados con los valores propios $\{\lambda_1, \lambda_2, \dots, \lambda_r\}$ de la matriz simétrica $\mathbf{X}^\top\mathbf{X}$.

$$(\mathbf{X}^\top\mathbf{X})\hat{\mathbf{v}}_i = \lambda_i\hat{\mathbf{v}}_i$$

- $\sigma_i \equiv \sqrt{\lambda_i}$ son reales positivos y se llaman *valores singulares*.
- $\{\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, \dots, \hat{\mathbf{u}}_r\}$ es el conjunto de vectores de tamaño $n \times 1$ definidos por $\hat{\mathbf{u}}_i \equiv \frac{1}{\sigma_i}\mathbf{X}\hat{\mathbf{v}}_i$.

La definición final incluye dos propiedades nuevas e inesperadas.

- $\hat{\mathbf{u}}_i \cdot \hat{\mathbf{u}}_j = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{en otro caso} \end{cases}$
- $\|\mathbf{X}\hat{\mathbf{v}}_i\| = \sigma_i$

La versión escalar de descomposición de valor singular es sólo una reformulación de la tercera definición

$$\mathbf{X}\hat{\mathbf{v}}_i = \sigma_i\hat{\mathbf{u}}_i. \quad (60)$$

Este resultado dice bastante. \mathbf{X} multiplicado por un vector propio de $\mathbf{X}^\top\mathbf{X}$ es igual a varias veces un escalar por otro vector. El conjunto de vectores propios $\{\hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2, \dots, \hat{\mathbf{v}}_r\}$ y el conjunto de vectores $\{\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, \dots, \hat{\mathbf{u}}_r\}$ son ambos conjuntos ortonormales o bases de un

Podemos definir una cantidad similar en el *espacio de fila*:

$$\begin{aligned}\mathbf{XV} &= \mathbf{\Sigma U}, \\ (\mathbf{XV})^\top &= (\mathbf{\Sigma U})^\top, \\ \mathbf{V}^\top \mathbf{X}^\top &= \mathbf{U}^\top \mathbf{\Sigma}, \\ \mathbf{V}^\top \mathbf{X}^\top &= \mathbf{Z},\end{aligned}$$

donde hemos definido $\mathbf{Z} \equiv \mathbf{U}^\top \mathbf{\Sigma}$. De nuevo las filas de \mathbf{V}^\top (o las columnas de \mathbf{V}) son una base ortonormal para la transformación de \mathbf{X}^\top en \mathbf{Z} . Debido a la transposición sobre \mathbf{X} resulta que \mathbf{V} es una base ortonormal que abarca el *espacio fila* de \mathbf{X} . El espacio fila del mismo modo formaliza la noción de lo que son posiblemente las “entradas” en una matriz arbitraria.

DVS y ACP. Es evidente que *ACP* y *DVS* están íntimamente relacionados. Volvamos a la matriz original de datos \mathbf{X} de tamaño $m \times n$. Podemos definir una nueva matriz \mathbf{Y} como una matriz de tamaño $n \times m$.

$$\mathbf{Y} \equiv \frac{1}{\sqrt{n}} \mathbf{X}^\top,$$

donde cada columna de \mathbf{Y} tiene media cero. La elección de \mathbf{Y} se hace evidente por el análisis de $\mathbf{Y}^\top \mathbf{Y}$:

$$\mathbf{Y}^\top \mathbf{Y} = \left(\frac{1}{\sqrt{n}} \mathbf{X}^\top \right)^\top \left(\frac{1}{\sqrt{n}} \mathbf{X}^\top \right) = \frac{1}{n} \mathbf{X} \mathbf{X}^\top = \mathbf{C}_\mathbf{X}.$$

Por construcción $\mathbf{Y}^\top \mathbf{Y}$ es igual a la matriz de covarianza de \mathbf{X} . Sabemos que las componentes principales de \mathbf{X} son los vectores propios de $\mathbf{C}_\mathbf{X}$. Si calculamos la DVS de \mathbf{Y} , las columnas de la matriz \mathbf{V} contienen los vectores propios de $\mathbf{Y}^\top \mathbf{Y} = \mathbf{C}_\mathbf{X}$. *Por lo tanto, las columnas de \mathbf{V} son las principales componentes de \mathbf{X} .*

¿Qué significa esto? \mathbf{V} se extiende por el espacio fila de $\mathbf{Y} \equiv \frac{1}{\sqrt{n}} \mathbf{X}^\top$. Por lo tanto, \mathbf{V} también deben extenderse a el espacio columna de $\frac{1}{\sqrt{n}} \mathbf{X}$. Podemos concluir que encontrando las cantidades de las componentes principales encontramos una base ortonormal que abarca todo el *espacio columna* de \mathbf{X} .

Discusión. El análisis de componentes principales (ACP) tiene amplias aplicaciones, ya que revela simples estructuras subyacentes en conjuntos de datos complejos con soluciones analíticas de álgebra lineal. La ventaja principal del ACP surge de la cuantificación de la importancia de cada dimensión para describir la variabilidad de un conjunto de datos. En particular, la medición de la variación a lo largo de cada componente principal proporciona un medio para comparar la importancia relativa de cada dimensión. Una esperanza implícita detrás de emplear este método es que la variación a lo largo de un pequeño número de componentes principales (es decir, menor que el número de tipos de medición) proporciona una caracterización razonable del conjunto completo de datos. Esta declaración es la intuición exacta detrás de cualquier método de *reducción dimensional*, un amplio campo de investigación activa.

2.5. Métodos espectrales para análisis de componentes principales de los eventos relacionados con los potenciales cerebrales

El método de Análisis de Componentes Principales (ACP) ha sido ampliamente utilizado para el análisis de eventos relacionados con los potenciales eléctricos del cerebro (ERPs). ERPs consisten en cambios característicos en el voltaje a través del tiempo con respecto a un estímulo específico. En la investigación de ERP, se considera que el ERP es una señal determinista, mientras que el proceso de la actividad espontánea de EEG (electroencefalografía) se considera que es un proceso estocástico. Nos atenemos a la designación de este proceso estocástico como el ruido. El término ruido se utiliza para enfatizar la desviación de la señal determinista ERP.

Uno de los objetivos del ACP es la descomposición de la forma compleja de la onda ERP en un número relativamente reducido, y por lo tanto más manejable de “componentes”.

Uno de los objetivos de este artículo es analizar las dificultades inherentes a la utilización de ACP, como una técnica analítica en estudios de múltiples grupos. Otro objetivo es demostrar la utilidad del análisis espectral y su equivalencia con el ACP cuando la señal incrustada en el modelo de ruido estacionario (SSN) se utiliza.

Análisis de componentes principales

Sea $x(t) = \mu(t) + \eta(t)$ un proceso estocástico en el intervalo $[0, T]$ con $\mathbb{E}(X(t)) = \mu(t)$, \mathbb{E} denota la esperanza matemática. El término $\eta(t)$ se considera un proceso de ruido tal que $\mathbb{E}(\eta(t)) = 0$. Sólo el caso para tiempo discreto $t = t_1, \dots, t_n$ será discutido. El proceso estocástico es un vector n -dimensional $x = (x(t_1), \dots, x(t_n))^T$ donde \top denota la transpuesta del vector. La matriz de covarianza es $\Sigma = \mathbb{E}((x - \mu)(x - \mu)^T)$, sus valores y vectores propios son $\{(\lambda_i, \Phi_i); i = 1, \dots, n\}$. El proceso aleatorio $\eta(t) = x(t) - \mu(t)$ se puede expresar como la expansión

$$\sum_{i=1}^n a_i \Phi_i,$$

donde las componentes principales están dadas por $a_i = \eta^T \Phi_i$. Notemos que $\{\Phi_i : i = 1, \dots, n\}$ son determinísticas mientras que las a_i son variables aleatorias con $\mathbb{E}(a_i) = 0$. El proceso η puede aproximarse mediante la expansión

$$\boldsymbol{\eta} = \sum_{i=1}^k a_i \Phi_i,$$

y el error cuadrático medio será

$$\mathbb{E}((\boldsymbol{\eta} - \boldsymbol{\eta})^T (\boldsymbol{\eta} - \boldsymbol{\eta})) = \sum_{i=k+1}^n \lambda_i.$$

Puesto que

$$\boldsymbol{\mu} = \sum_{i=1}^n (\boldsymbol{\mu}^T \Phi_i) \Phi_i = \sum_{i=1}^k (\boldsymbol{\mu}^T \Phi_i) \Phi_i + \sum_{i=k+1}^n (\boldsymbol{\mu}^T \Phi_i) \Phi_i,$$

se ve que mediante la utilización de sólo las primeras k componentes principales, μ se puede aproximar con un error de $\sum_{i=k+1}^n (\mu^\top \Phi_i) \Phi_i$ (i.e., se obtiene la proyección del vector μ sobre el subespacio generado por los primeros k vectores propios). De las propiedades de vectores propios, y definiendo el vector aleatorio $F' = (f_1, \dots, f_n)$, donde $F = \Lambda^{-1/2} a = \Lambda^{-1/2} \Phi^\top \eta$ con las propiedades $\mathbb{E}(FF^\top) = 1$ y $\mathbb{E}(F) = 0$, la expresión $\Sigma = \Phi \Lambda \Phi^\top$ puede escribirse como $\Sigma = \mathbb{E}(\Phi \Lambda^{1/2} F)(\Phi \Lambda^{1/2} F)^\top = \mathbb{E}(\eta \eta^\top)$. Los valores de F se denominan *puntuaciones factoriales*. η puede ser expresado como $\Phi \Lambda^{1/2} F$ y la matriz $\Phi \Lambda^{1/2}$ se conoce como la matriz de factor de carga. Bajo una rotación en la matriz de factor de carga se obtiene como resultado puntuaciones factoriales para factores de rotación. El error cuadrático medio y el error de la señal siguen siendo los mismos bajo rotaciones ortogonales.

Es habitual en la investigación de ERP limitar el análisis a los vectores propios con valor propio más grande que corresponde a un valor elegido. Mientras que este procedimiento tendrá como resultado un error cuadrático medio aceptable para el ruido, no necesariamente resultará un error aceptable para la señal. Si deseamos estudiar K grupos en términos de componentes principales comunes, entonces se requiere que $\Sigma_1 = \dots = \Sigma_K = \Sigma$, de modo que un conjunto de n componentes principales comunes se puedan obtener. Pero si suponemos que $\Sigma_1 = \dots = \Sigma_K = \Sigma$, entonces es posible obtener un único conjunto de valores y vectores propios para todos los conductores. Sin embargo, esto requiere la suposición de que los procesos de ruido para cada conductor son el mismo. En consecuencia, el ACP no proporciona una metodología que prácticamente pueda manejar múltiples electrodos con procesos de ruido diferentes.

Propiedades estadísticas de ACP

Ya se ha hablado del modelo de probabilidad de un proceso estocástico η . Dado que, en general, los parámetros de población no se conocen, los estimadores estadísticos se deben determinar. Para el problema de la población solo los parámetros $\mu, (\lambda_i, \Phi_i), i = 1, \dots, n$, deben estimarse a partir de observar una muestra de N vectores x_1, \dots, x_N . Si a partir de la observación de los vectores x_j ($j = 1, \dots, N$) se obtiene una distribución normal multivariante entonces los estimadores de máxima verosimilitud son

$$\hat{\mu} = \frac{1}{N} \sum_{j=1}^N x_j \quad \text{y} \quad \frac{1}{(N-1)} \sum_{j=1}^N (x_j - \hat{\mu})(x_j - \hat{\mu})^\top = \hat{\Sigma}$$

Los estimadores $\{(\hat{\lambda}_i, \hat{\Phi}_i), i = 1, \dots, n\}$ se obtienen de $\hat{\Sigma}$.

Debido a la gran cantidad de puntos en el tiempo y el pequeño número de sujetos en el típico estudio de ERP, la proporción de N/n no suele ser lo suficientemente grande como para obtener estimaciones estables de la matriz de factor de carga. Esto implica que la interpretación de las componentes no sea muy fiable. Debido al tamaño limitado de la muestra, un subconjunto de los vectores propios es todo lo que es estimable resultando en el mejor de los casos, un subespacio S generado por este subconjunto de vectores propios. Por consiguiente, sólo la proyección de la señal en S puede ser estudiada. Ya que no se ha asumido una relación entre μ y η , no hay ninguna razón particular para suponer que μ se encuentran en S .

En el problema de K -grupos podemos estimar los parámetros para cada grupo de una manera similar a la anterior. Si los vectores observados $\{x_{ij} : i = 1, \dots, K; j = 1, \dots, N_i\}$

se obtienen a partir de K distribuciones normales multivariadas entonces los estimadores de máxima verosimilitud habituales son

$$\hat{\mu}_i = (1/N_i) \sum_{j=1}^{N_i} x_{ij}, \quad i = 1, \dots, K,$$

y

$$\hat{\Sigma} = (1/(N_1 + \dots + N_K - K)) \sum_{i=1}^K \sum_{j=1}^{N_i} (x_{ij} - \hat{\mu}_i)(x_{ij} - \hat{\mu}_i)^\top.$$

Los estimadores $\{\hat{\lambda}_l, \hat{\Phi}_l; l = 1, \dots, n\}$ se obtienen de $\hat{\Sigma}$.

Otro aspecto de la descomposición de la señal con ACP se debe discutir. Si se supone que una señal en particular consiste en la suma de ondas, entonces el ACP no necesariamente proporciona estimaciones de las ondas componentes. Los vectores propios se derivan de la matriz de covarianza de ruido, y no hemos asumido ninguna relación entre la señal y ruido. Una de las razones para la utilización de metodología de ACP en los estudios de ERP es la obtención de las variables ortogonales en lugar de que las variables correlacionadas de modo que la inferencia estadística más fiable puede obtenerse a partir de los vectores multivariados de grandes dimensiones. Sin embargo, debido a los pequeños tamaños de muestra que se encuentran normalmente, no es posible el uso de todos los vectores propios. Con el fin de resolver el problema de forma fiable, la estimación del gran número de parámetros, existe otro método que consiste en restringir la clase de los procesos de ruido a aquellos con matrices de covarianza especialmente estructuradas que contienen menor número de parámetros. Un modelo de matriz de covarianza de uso común es la matriz de Toeplitz que surge al considerar la serie de tiempo estacionaria. A continuación se muestra cómo el análisis de componentes principales de las matrices de Toeplitz se relaciona con el análisis espectral de series de tiempo estacionarias.

Análisis espectral

Consideremos $\mathbf{x}_{jl}(t) = \boldsymbol{\mu}_j(t) + \boldsymbol{\eta}_{jl}(t)$ series de tiempo multivariadas de dimensión p donde $t = 0, 1, \dots, T-1$; $j = 1, \dots, q$; $l = 1, \dots, N_j$, tal que q es el número de grupos y N_j denota el número de series de tiempo independientes en el j -ésimo grupo. $\boldsymbol{\eta}_{jl}(t)$ son procesos estacionarios Gaussianos con media cero y con matriz de correlación cruzada $R_j(t-u) = [\gamma_j^{mn}(t-u); m, n = 1, \dots, p]$, $\boldsymbol{\mu}_j(t)$ denota el potencial de eventos relacionados para grupo j . La matriz de densidad espectral $F_j(\lambda)$ para el grupo de j está definida por la representación $R_j(t-u) = \int_{-\pi}^{\pi} e^{i\lambda(t-u)} F_j(\lambda) d\lambda$. La transformada de Fourier Finita multivariada (TFF) de $\mathbf{x}(t)$ se define por $\hat{\mathbf{x}}(k) = \sum_{t=0}^{T-1} \mathbf{x}(t) e^{-i\lambda_k t}$ donde $\lambda_k = 2\pi k/T$ para $k = 0, \dots, T-1$. Cuando la TFF se aplica al modelo de series de tiempo anterior, se obtiene $\hat{\mathbf{x}}_{jl}(t) = \hat{\boldsymbol{\mu}}_j(t) + \hat{\boldsymbol{\eta}}_{jl}(t)$. $\hat{\boldsymbol{\eta}}_{jl}(k)$ tiene una distribución normal compleja multivariada, $N(0, 2\pi T F_j(\lambda))$ y para T grande los vectores $\hat{\boldsymbol{\eta}}_{ji}(k)$ y $\hat{\boldsymbol{\eta}}_{jl}(k')$ son independientes para $k \neq k'$. Realizando la TFF inversa, la siguiente representación de $\boldsymbol{\mu}(t) = 1/\sum_{k=0}^{T-1} \hat{\boldsymbol{\mu}}(k) e^{i\lambda_k t}$ puede obtenerse en términos de los términos trigonométricos complejos $e^{i\lambda_k t}$.

La representación anterior de $\boldsymbol{\mu}(t)$ para el análisis de componentes principales de la matriz de Toeplitz para el proceso del ruido $\boldsymbol{\eta}(t)$ se discute para el caso univariado. Los vectores propios de la matriz de Toeplitz son aproximadamente

$$\left\{ \frac{1}{\sqrt{T}} e^{i\lambda_k 0}, \frac{1}{\sqrt{T}} e^{i\lambda_k 1}, \dots, \frac{1}{\sqrt{T}} e^{i\lambda_k (T-1)} \right\} = \frac{1}{\sqrt{T}} \psi_k \quad k = 0, \dots, T-1$$

Las componentes principales están dadas por $a_k = (1/\sqrt{T})\psi_k^* \eta$ donde

$$\eta = (\eta(0), \dots, \eta(t-1))^T \quad \text{y} \quad \sqrt{T}(a_0, \dots, a_{T-1})^T = \sqrt{T}a$$

es la TFF para η . Una aproximación para η en terminos de componentes principales de la matriz de Toeplitz es $1/\sqrt{T} \sum_{k=0}^{T-1} a_k \psi_k = (1/\sqrt{T})\psi a$.

En el caso ruido estacionario no es necesario calcular los valores y vectores propios por los métodos usuales de componentes principales ya que estas estimaciones pueden obtenerse mediante la utilización de métodos más sencillos como TFF. En cualquier caso, los métodos habituales de ACP no serán correctos a menos de que la forma de la matriz de Toeplitz de la matriz de covarianza se tenga en cuenta. Una estimación de la matriz de covarianza puede obtenerse mediante la utilización de la TFF de la función de densidad espectral del ruido. Existe considerable evidencia de que la investigación de estacionariedad en ERP de la actividad del EEG varía con el tiempo como una función de estado cerebral. El resultado de esto es que las matrices de densidad espectral estarán en diferentes estados cerebrales. Si se utiliza un proceso de ruido estacionario, se ha demostrado que el análisis de componentes principales y el análisis espectral serán equivalentes.

2.6. Una nota sobre generación, estimación y predicción de procesos estacionarios

1. Introducción Algunos procesos estacionarios como los procesos fraccionalmente integrados no pueden ser descritos por modelos autoregresivos o por modelos de promedio móvil. Se ofrece un enfoque basado en la descomposición de Cholesky de la matriz de covarianza que hace resolubles a estos problemas. Partimos de procesos estacionarios con una representación de Wold de la forma

$$y_t - \mu = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} \quad (62)$$

donde ϵ_t es el ruido correlacionado con media cero. Las ψ_i convergen cuadráticamente y la varianza del ruido, σ_ϵ^2 , es más grande que cero. Se supone que $\mu = 0$. Y_T denota el vector $(y_1, \dots, y_T)^\top$ y $E_T = (\epsilon_1, \dots, \epsilon_T)^\top$. La matriz de covarianza de Y_T , Σ_T , es definida positiva, simétrica y Toeplitz y así persimétrica. Puede ser factorizada de acuerdo a la descomposición de Cholesky

$$\Sigma_T = L_T L_T^\top \quad (63)$$

L_T es una matriz triangular inferior.

Una posibilidad para generar la muestra de tamaño T de un proceso dado con la misma estructura de covarianza, es utilizar la relación

$$Y_T = L_T E_T \quad (64)$$

Suponiendo que la estimación del ruido con distribución normal se puede llevar a cabo mediante la maximización de la verosimilitud Gaussiana

$$f(Y_T; \mu, \Sigma_T) = (2\pi)^{-T/2} |\Sigma_T|^{-1/2} \exp[-(Y_T - \mu)' \Sigma_T^{-1} (Y_T - \mu) / 2]. \quad (65)$$

Para modelos ARMA existen presentaciones más simples de verosimilitud pero para modelos fraccionalmente integrados esta es la única solución exacta.

El vector de ruido implícito se puede obtener por

$$E_T = L_T^{-1} Y_T \quad (66)$$

La predicción lineal de un paso a los τ -pasos se puede hacer mediante la extensión de la ecuación anterior a $T + \tau$ y reemplazando los ruidos futuros por la esperanza que es cero. Esto es

$$Y_{T+\tau} = \begin{bmatrix} Y_T \\ Y_\tau \end{bmatrix} = L_{T+\tau} E_{T+\tau} = \begin{bmatrix} L_T & 0 \\ L_{\tau T} & L_{\tau\tau} \end{bmatrix} \begin{bmatrix} E_T \\ E_\tau \end{bmatrix}$$

y

$$E[Y_\tau | Y_T] = \begin{bmatrix} L_{\tau T} & L_{\tau\tau} \end{bmatrix} \begin{bmatrix} E_T \\ 0 \end{bmatrix} = L_{\tau T} E_T \quad (67)$$

La varianza de la predicción lineal Y_τ , dada Y_T , E_T , respectivamente, esta dada por la matriz de covarianza $\Sigma_{T+\tau}$, con $\Sigma_{T+\tau} = L_{T+\tau}L_{T+\tau}^\top$,

$$V[Y_\tau|Y_T] = E[(Y_\tau - E[Y_\tau|Y_T])(Y_\tau - E[Y_\tau|Y_T])^\top | Y_T] = L_{\tau\tau}L_{\tau\tau}^\top \quad (68)$$

Si ϵ_t es correlacionado y normal con varianzas no-constantes, que dependen del pasado, el proceso de verosimilitud está dado por (65) mediante la sustitución de la matriz de covarianza Σ_T por una matriz de covarianza de un proceso dependiente

$$\Sigma_T = L_T H_T L_T^\top \quad (69)$$

donde H_T es diagonal y contiene la varianza condicional de ϵ_t normalizado.

Generación y predicción lineal son análogas para el caso homoscedástico una vez que se dan las innovaciones heterocedásticos. La varianza del predictor lineal es, sin embargo,

$$V[Y_\tau|Y_T] = L_{\tau\tau} H_{\tau|Y_T} L_{\tau\tau}^\top \quad (70)$$

con $H_{T+\tau} = \begin{bmatrix} H_T & 0 \\ 0 & H_\tau \end{bmatrix}$ y $H_{\tau|Y} = E[H_\tau|Y_T]$.

2. La multiplicación del factor de Cholesky con un vector arbitrario.

Notación

$$\Sigma_{T+1} = \begin{bmatrix} \Sigma_T & \Sigma_{1T}^\top \\ \Sigma_{1T} & \Sigma_{11} \end{bmatrix}, L_{T+1} = \begin{bmatrix} L_T & 0 \\ L_{1T} & L_{11} \end{bmatrix}, R_{T+1} = \begin{bmatrix} R_T & Er \\ (Er)^\top & 1 \end{bmatrix}$$

E es una matriz cuadrada con unos en la segunda diagonal y ceros en las demás. R_T es la matriz de correlación.

Lema 2. *El mejor predictor lineal por delante de un paso \hat{y}_{T+1} , de y_{T+1} , en terminos de Y_T y su error cuadrático medio es*

$$\hat{y}_{T+1} = \Sigma_{1T}\Sigma_T^{-1}Y_T, \quad v_T = \Sigma_{11} - \Sigma_{1T}\Sigma_T^{-1}\Sigma_{1T}^\top. \quad (71)$$

Proposición 12. El mejor predictor lineal por delante de un paso \hat{y}_{T+1} de y_{T+1} en terminos de los factores de Cholesky y el vector pasado de innovación E_T y su error cuadrático medio es

$$\hat{y}_{T+1} = L_{1T}E_T, \quad v_T = L_{11}L_{11} \quad (72)$$

Para la generación de muestras de un proceso con cierta matriz de covarianza, el mejor predictor lineal se puede utilizar de forma recursiva de la siguiente manera a partir de $T = 0$ con $v_0 = \sigma_y^2$

$$y_{T+1} = \Sigma_{1T}\Sigma_T^{-1}Y_T + \sqrt{v_T}\epsilon_{T+1} \quad (73)$$

donde ϵ_t es una secuencia de innovación. En notación de la matriz de Cholesky esto equivale a

$$y_{T+1} = L_{1T}E_T + L_{11}\epsilon_{T+1} \quad (74)$$

Esta es la multiplicación de la última línea de L_{T+1} con E_{T+1} . Un algoritmo eficiente para calcular el mejor predictor lineal y su error cuadrático medio, es el algoritmo Durbin-Levinson.

Este algoritmo realiza la multiplicación de la matriz de Cholesky de una matriz de Toeplitz simétrica con un vector arbitrario. Esto es notable, ya que no existe un procedimiento conocido para las matrices de Toeplitz de estructura sencilla para calcular la matriz de Cholesky con menos de $O(T^3)$ fracasos y $O(T^2)$ de almacenamiento.

Derivación del algoritmo. Reformulamos el problema en correlaciones en lugar de covarianzas, lo que implica que $\sigma_y^2 = 1, \Sigma_T = \sigma_y^2 R_T$ respectivamente. Los dos primeros momentos de y_{T+1} simplifican $(Er)^\top R_T^{-1} Y_T$ y $1 - r^\top R_T^{-1} r$, usando la notación de antes

$$R_{T+1}^{-1} = \begin{bmatrix} R_T & Er \\ (Er)^\top & 1 \end{bmatrix}^{-1} = \begin{bmatrix} \beta & v \\ v^\top & \gamma \end{bmatrix}$$

Esto implica que

$$\begin{bmatrix} R_T & Er \\ (Er)^\top & 1 \end{bmatrix} \begin{bmatrix} v \\ \gamma \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Resolviendo para v y γ tenemos $R_T v = -\gamma Er$ para la primer ecuación. Así v puede ser expresada a través de la solución y de las ecuaciones de Yule-Walker, $R_T y = -r$, $y = -R_T^{-1} r$ y $v = \gamma E y$. Reemplazando v en la segunda ecuación γ puede ser expresada como $\gamma = 1/(1 + r'y) = 1/(1 - r'R_T^{-1} r)$. Los dos primeros momentos de y_{T+1} pueden ser expresados en terminos de y . Esto es: $(Er)' R_T^{-1} Y_T = -(Ey)' Y_T$ y $1 - r'R_T^{-1} r = 1/\gamma = 1 + r'y$.

3. El factor inverso de Cholesky.

Proposición 13. La matriz de Cholesky inversa está relacionada con la matriz de Cholesky de la inversa por transposición con respecto a la diagonal secundaria.

Cálculos. La generación de muestras del proceso (y_t) pueden ser obtenidas de manera eficiente en los requerimientos lineales de almacenamiento, se toma una vez la matriz de autocorrelación. Si la estimación se realiza através de la función de verosimilitud el algoritmo de Levinson se puede utilizar para calcular la descomposición de Cholesky y por lo tanto el determinate requerido. La innovaciones resultantes pueden ser calculadas usando la descomposición de Cholesky de la última iteración del procedimiento de optimización, proposición 13 y (66). La predicción lineal del paso 1 al paso τ (pronóstico) del vector dado Y_T se puede calcular a través del vector residual (estimado) y (67) usando la función de autocovarianza(estimada). Especialmente en el caso del cálculo de la varianza del predictor lineal, (68), la proposición 12 es muy útil dado que τ es típicamente pequeña. Multiplicando $L_{T+\tau}$ por un vector con ceros y un 1 en la posición $(T + j)$, escoge exactamente la $(T + j)$ -ésima columna que es la columna j en $L_{\tau\tau}$. Sin almacenar los resultados intermedios de la multiplicación de $L_{T+\tau}$ con los primeros T ceros el número de fracasos es $O(\tau T^2)$. El almacenamiento es lineal si sólo son necesarios los elementos de la diagonal. El procedimiento puede ser fácilmente generalizado para innovaciones heterocedásticas. Los 1's se tienen que reemplazar por la raíz cuadrada de las varianzas condicionales.

2.7. Distribución asintótica de valores propios de matrices de bloque Toeplitz y aplicaciones

Se sugiere una prueba sencilla para la extensión más importante del teorema de Szegő para matrices de Toeplitz en bloques con bloques no Toeplitz en donde el número de bloques tiende a infinito. Nos centramos, en una clase especial de matrices de Toeplitz en bloques, que con frecuencia encontrados en el procesamiento de la señal, para dar una forma más simple del teorema Szegő y deducir resultados sobre el valor propio más bajo distinto de cero, que expresa el acondicionamiento con respecto a la inversión de tales matrices.

II. Resultados previos

Se presenta una definición, algunos lemas, teoremas y por supuesto el teorema de Szegő que será extendido.

Definición 1. Equivalencia asintótica. Dos sucesiones de matrices (A_n) y (B_n) , $n = 1, 2, \dots$ se dicen ser asintóticamente equivalentes y se denota $(A_n) \sim (B_n)$ si

$$\exists M < \infty \text{ tal que } \forall n, \|A_n\| \leq M \text{ y } \|B_n\| \leq M \quad (75)$$

$$\lim_{n \rightarrow \infty} \|A_n - B_n\| = 0. \quad (76)$$

Lema 3. Si $(A_n) \sim (B_n)$ y si

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \lambda_k^s(A_n)$$

existe y es finito para cualquier entero positivo s , entonces

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \lambda_k^s(A_n) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \lambda_k^s(B_n)$$

Lema 4. Para todas las sucesiones absolutamente convergentes $(t_k)_{k=\dots,-1,0,1,\dots}$ existe una sucesión de matrices circulantes $\mathbf{C}_n(t)$ asintóticamente equivalentes a $\mathbf{T}_n(t)$ y están dadas por $\mathbf{C}_n(t) = \mathbf{U}_n^H \mathbf{D}_n(t) \mathbf{U}_n$ donde $\mathbf{D}_n(t)$ es una matriz diagonal donde la k -ésima entrada está dada por

$$(\mathbf{D}_n(t))_{k,k} = t\left(\frac{2\pi(k-1)}{n}\right)$$

y \mathbf{U}_n es la matriz de la transformada de Fourier discreta unitaria

$$(U_n)_{k,l} = \frac{1}{\sqrt{n}} e^{-i2\pi \frac{(k-1)(l-1)}{n}}$$

Teorema 14. Teorema de Szegő. Para todas las sucesiones absolutamente convergentes $(t_k)_{k=\dots,-1,0,1,\dots}$ si $\mathbf{T}_n(t)$ es Hermitiana, entonces para todas las funciones continuas F sobre $[\min_w t(w), \max_w t(w)]$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n F(\lambda_k(\mathbf{T}_n(t))) = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(t(w)) dw.$$

Teorema 15. *Para sucesiones absolutamente convergentes $(t_k)_{k=\dots,-1,0,1,\dots}$, si $\mathbf{T}_n(t)$ es Hermitiana, entonces, para cualquier l , los valores propios bajos (resp., más altos) l de $\mathbf{T}_n(t)$ decrecen (resp., crecen) con n y convergen a $\min_w t(w)$ (resp., $\max_w t(w)$).*

III. Matrices de Toeplitz por bloques

A. Definiciones

Para extender los resultados, definimos la matriz de Toeplitz por bloques

$$T'_n(\{t^{u,v}\}) = \begin{bmatrix} \mathbf{T}_0 & \mathbf{T}_{-1} & \cdots & \mathbf{T}_{-(n-1)} \\ \mathbf{T}_1 & \ddots & \ddots & \mathbf{T}_{-(n-2)} \\ \vdots & \ddots & & \vdots \\ \mathbf{T}_{n-1} & \mathbf{T}_{n-2} & \cdots & \mathbf{T}_0 \end{bmatrix},$$

donde $(\mathbf{T}_k)_{k=-(n-1),\dots,n-1}$ son matrices $c \times c$ (no necesariamente de Toeplitz) de entradas $t_k^{u,v} \hat{=} (\mathbf{T}_n)_{u,v}$, $u, v = 1, \dots, c$. Consideramos la matriz asociada

$$T_n(\{t^{u,v}\}) \hat{=} \begin{bmatrix} \mathbf{T}_n(t^{1,1}) & \mathbf{T}_n(t^{1,2}) & \cdots & \mathbf{T}_n(t^{1,c}) \\ \mathbf{T}_n(t^{2,1}) & \mathbf{T}_n(t^{2,2}) & \cdots & \mathbf{T}_n(t^{2,c}) \\ \vdots & & & \vdots \\ \mathbf{T}_n(t^{c,1}) & \mathbf{T}_n(t^{c,2}) & \cdots & \mathbf{T}_n(t^{c,c}) \end{bmatrix}$$

donde $\mathbf{T}_n(t^{u,v})$, $u, v = 1, \dots, c$ son matrices de Toeplitz $n \times n$. Suponemos que $\{t^{u,v}\}$, $u, v = 1, \dots, c$ son conjuntos finitos de sucesiones absolutamente convergentes. Así, la transformada de Fourier

$$t^{u,v}(w) \hat{=} \sum_k t_k^{u,v} e^{-ikw}$$

puede asociarse con cada sucesión. Usando la matriz de permutación $\mathbf{K}_{r,s}$, tenemos

$$T'_n(\{t^{u,v}\}) = \mathbf{K}_{n,c} T_n(\{t^{u,v}\}) \mathbf{K}_{c,n} = \mathbf{K}_{n,c} T_n(\{t^{u,v}\}) \mathbf{K}_{n,c}^{-1}$$

Así que las matrices $T'_n(\{t^{u,v}\})$ y $T_n(\{t^{u,v}\})$ son similares y por lo tanto, equivalentes desde el punto de vista de los valores propios. $T_n(\{t^{u,v}\})$ es Hermitiana si y sólo si $\mathbf{T}_n(t^{u,v}) = \mathbf{T}_n^H(t^{u,v})$, $u, v = 1, \dots, c$.

B. Distribución asintótica de valores propios El Lema 4 se extiende para matrices de Toeplitz por bloques como sigue.

Lema 5. *Para sucesiones absolutamente convergentes $(t_k)_{k=\dots,-1,0,1,\dots}$ existe una sucesión de matrices $(C_n(\{t^{u,v}\}))$ asintóticamente equivalentes a $(T_n(\{t^{u,v}\}))$ y dadas por $C_n(\{t^{u,v}\}) = U_n^H D_n(\{t^{u,v}\}) U_n$, donde U_n es una matriz unitaria independiente de $T_n(\{t^{u,v}\})$ y $D_n(\{t^{u,v}\})$ es la siguiente matriz*

$$D_n(\{t^{u,v}\}) \hat{=} \begin{bmatrix} \mathbf{D}_n(t^{1,1}) & \mathbf{D}_n(t^{1,2}) & \cdots & \mathbf{D}_n(t^{1,c}) \\ \mathbf{D}_n(t^{2,1}) & \mathbf{D}_n(t^{2,2}) & \cdots & \mathbf{D}_n(t^{2,c}) \\ \vdots & & & \vdots \\ \mathbf{D}_n(t^{c,1}) & \mathbf{D}_n(t^{c,2}) & \cdots & \mathbf{D}_n(t^{c,c}) \end{bmatrix} \quad (77)$$

$\mathbf{D}_n(\{t^{u,v}\})$ es la matriz definida en el Lema 4.

Lema 6. Para todos los enteros $s \geq 1$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^{cn} \lambda_k^s(T_n(\{t^{u,v}\})) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{1 \leq k_1, \dots, k_s \leq c} t^{k_1, k_2}(w) t^{k_2, k_3}(w) \dots t^{k_s, k_1}(w) dw \quad (78)$$

Teorema 16. Suponemos que $T_n(\{t^{u,v}\})$ es Hermitiana; entonces para todas las funciones continuas F

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^{cn} F(\lambda_k(T_n(\{t^{u,v}\}))) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{u=1}^c F[\lambda_u(\mathbf{T}(w))] dw.$$

Añadido al hecho de que, para todos los n , los valores propios de $T_n(\{t^{u,v}\})$ se encuentran en $[\min_w \lambda_c(\mathbf{T}(w)), \max_w \lambda_1(\mathbf{T}(w))]$, el Teorema 16 implica que para cualquier entero l , los valores propios más bajos (resp., grandes) de l son convergentes en n y

$$\lim_{n \rightarrow \infty} \lambda_{cn-l+1}(T_n(\{t^{u,v}\})) = \min_w \lambda_c(\mathbf{T}(w)) \quad (79)$$

$$\lim_{n \rightarrow \infty} \lambda_1(T_n(\{t^{u,v}\})) = \max_w \lambda_1(\mathbf{T}(w)). \quad (80)$$

C. Una clase de matrices Toeplitz en bloques

Hipótesis H1: $\mathbf{T}(w)$ tiene rango 1, para toda w . Esto equivale a tener para toda w

$$\mathbf{T}(w) = [t^1(w), \dots, t^c(w)]^\top |\text{Tr}(\mathbf{T}(w))| [t^1(w), \dots, t^c(w)]$$

donde $[t^1(w), \dots, t^c(w)]^\top$ y $[t^1(w), \dots, t^c(w)]^\top$ tienen norma unitaria y son vectores singulares izquierdo y derecho, respectivamente, están asociados con un unico valor singular no cero de $\mathbf{T}(w)$. Como este valor singular es de multiplicidad uno, que es, así como los vectores singulares asociados, una función continua de $\mathbf{T}(w)$, que, a su vez, es función continua de w por la construcción. Por lo tanto, mediante la redefinición de $t^u(w)$ y $t'^u(w)$ a ser, respectivamente, $\sqrt{|\text{Tr}(\mathbf{T}(w))|} t^u(w)$ y $\sqrt{|\text{Tr}(\mathbf{T}(w))|} t'^u(w)$. La hipótesis *H1* equivale a

$$T_n(\{t^{u,v}\}) = \begin{bmatrix} \mathbf{T}_{(n)}(t^1) \\ \vdots \\ \mathbf{T}_{(n)}(t^c) \end{bmatrix} \begin{bmatrix} \mathbf{T}_{(n)}(t'^1) \\ \vdots \\ \mathbf{T}_{(n)}(t'^c) \end{bmatrix}^\top \quad (81)$$

donde $\mathbf{T}_{(n)}(t^u)$ (resp., $\mathbf{T}_{(n)}(t'^u)$), $u = 1, \dots, c$, denota las n filas de la matriz Toeplitz de la primera fila $[\dots, t_{-1}^u, t_0^u, t_1^u, \dots]$ (resp., $[\dots, (t'_{-1})^u, (t'_0)^u, (t'_1)^u, \dots]$). Por otra parte, si $T_n(\{t^{u,v}\})$ es Hermitiana y semi-definida positiva, *H1* se cumple si y sólo si $(t'_k)^u = (t_{-k}^u)^*$, $u = 1, \dots, c$ o equivalente $\mathbf{T}_{(n)}((t')^u) = \mathbf{T}_{(n)}^*(t^u)$. i.e.,

$$T_n(\{t^{u,v}\}) = \begin{bmatrix} \mathbf{T}_{(n)}(t^1) \\ \vdots \\ \mathbf{T}_{(n)}(t^c) \end{bmatrix} \begin{bmatrix} \mathbf{T}_{(n)}(t^1) \\ \vdots \\ \mathbf{T}_{(n)}(t^c) \end{bmatrix}^H \quad (82)$$

La condición anterior se encuentra con frecuencia en aplicaciones de procesamiento de señales, porque (82) representa la matriz de covarianza de un proceso estacionario c -variable

obtenida mediante el filtrado de un proceso estacionario escalar blanco. Sin embargo, observamos que esta factorización y así $H1$ no satisfacen las matrices de covarianza de procesos estacionarios c -variados más generales. De la misma manera, (81) representa la matriz de covarianza cruzada de dos procesos estacionarios c -variados obtenidos mediante el filtrado de un proceso estacionario escalar blanco de valor complejo. Con la hipótesis $H1$, el siguiente teorema queda probado.

Teorema 17. *Supongamos que $T_n(\{t^{u,v}\})$ es Hermitiana y cumple $H1$, entonces para todas las funciones continuas F*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^{cn} F(\lambda_k(T_n(\{t^{u,v}\}))) = (c-1)F(0) + \frac{1}{2\pi} \int_{-\pi}^{\pi} F\left(\sum_{u=1}^c t^{u,u}(w)\right) dw.$$

IV. Aplicación a la identificación ciega de canales SIMO

A. *Resultados para el Filtrado de Matriz de Canales SIMO.* Un canal SIMO de orden m , es un conjunto de c filtros $\mathbf{h}^i \triangleq [h_0^i, \dots, h_m^i]^\top$, $i = 1, \dots, c$ impulsado por una entrada escalar común $s(k)$, relacionada con el vector de salida c -dimensional $\mathbf{x}(k)$ por

$$\mathbf{x}(k) \triangleq [x^1(k), \dots, x^c(k)]^\top = \mathbf{G}(\mathbf{h})\mathbf{s}_{m+1}(k)$$

con $\mathbf{s}_{m+1}(k) \triangleq [s(k), \dots, s(k-m)]^\top$ y $\mathbf{G}(\mathbf{h}) \triangleq [\mathbf{h}(0) \dots \mathbf{h}(m)]$ donde $\mathbf{h}(k) \triangleq [h_k^1 \dots h_k^c]^\top$. Los canales SIMO de orden m se definen como el orden máximo entre los de los diferentes filtros $\mathbf{h}^1 \dots \mathbf{h}^c$. n observaciones de salidas sucesivas se apilan, de tiempo en tiempo, en

$$\mathbf{x}'_n(k) \triangleq [x^\top(k) \dots x^\top(k-(n-1))]^\top$$

y la matriz de covarianza se define como $\mathbf{R}'_n \triangleq E[\mathbf{x}'_n(k)(\mathbf{x}'_n(k))^H]$. Si la entrada $s(k)$ tiene media cero y blanca con varianza σ_s^2 entonces $\mathbf{R}'_n = \sigma_s^2 \mathbf{G}_n(\mathbf{h})\mathbf{G}_n^H(\mathbf{h})$, donde

$$\mathbf{G}_n(\mathbf{h}) \triangleq \begin{bmatrix} \mathbf{G}(\mathbf{h}) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{G}(\mathbf{h}) & \dots & \mathbf{0} \\ & \vdots & & \\ \mathbf{0} & \dots & & \mathbf{G}(\mathbf{h}) \end{bmatrix}$$

es la matriz de filtrado $cn \times (n+m)$ y $\mathbf{0}$ es un vector nulo c -dimensional. La aplicación del teorema 17 a \mathbf{R}'_n implica que para todas las funciones continuas F , se tiene

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^{cn} F(\lambda_k(\mathbf{R}'_n)) = (c-1)F(0) + \frac{1}{2\pi} \int_{-\pi}^{\pi} F\left(\sigma_s^2 \sum_{u=1}^c |h^u(w)|^2\right) dw. \quad (83)$$

Apartir de la igualdad anterior se puede demostrar el siguiente teorema

Teorema 18. *Para cualquier función continua F*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^{n+m} F(\sigma_k^{(n)}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} F\left(\sigma_s \sqrt{\sum_{u=1}^c |h^u(w)|^2}\right) dw. \quad (84)$$

Es interesante estudiar el comportamiento asintótico del valor singular más pequeño $\sigma_{n+m}^{(n)}$. Sin embargo, no se puede escribir como $\lambda_l(\mathbf{R}_n)$ o λ_{nc-l+1} , para algún l fijo y por lo tanto no se puede aplicar ni (79) ni (80).

B. Implicaciones de la identificación ciega de canales SIMO

La matriz de covarianza \mathbf{R}_n contiene información sobre la fase del canal y se utiliza para deducir el canal de coeficientes $\mathbf{h}(k)$, es llamado problema de identificación, para una gran variedad de algoritmos de segundo orden. El valor propio más pequeño distinto de cero es fundamental para el rendimiento del algoritmo de identificación ciega. Los resultados demostraron asintóticas en el mismo se estableció para matrices de Toeplitz de bloques con bloques de Toeplitz donde tanto el tamaño y el número de los bloques tienden al infinito, mientras que en esta correspondencia, sólo el tamaño de los bloques de n tiende a infinito. Cuando se observa sobre intervalos de tiempo finitos y en la presencia de ruido, lo anterior es insuficiente y el canal tiene que exhibir una diversidad suficiente para permitir la estimación de la respuesta precisa.

C. Canales de banda limitada fraccionalmente espaciados Ahora nos centramos en los canales de banda limitada fraccionalmente espaciados. Si los subcanales $h^k(w)$, $k = 1, \dots, c$ se emiten desde el sobremuestreo en forma de onda $h(t)$, entonces

$$h^k(w) = \sum_l h(w - 2l\pi) e^{-j(w-2l\pi)\frac{k-1}{c}}$$

donde

$$h(w) \hat{=} \int h(t) e^{-j\omega t} dt.$$

Cuando $h(t)$ es una banda limitada (para $[-\frac{1}{T}, \frac{1}{T}]$), entonces

$$h^k(w) = h(w) e^{-jw\frac{k-1}{c}} + h(w - 2\pi) e^{-j(w-2\pi)\frac{k-1}{c}}$$

para $w \in [0, 2\pi]$. Más comúnmente, $h(w)$ es un respuesta del filtro de conformación (una onda de coseno elevado con mayor frecuencia) se propaga a través de una selectiva frecuencia en canal multitrayecto. Debido a la selectividad severa de banda limitada, algunos componentes de frecuencia pueden ser significativamente atenuados que llevan al límite superior anterior a ser débil. Esto justifica el bajo rendimiento de los algoritmos de ciegos en la identificación de los canales de comunicación con los receptores de fraccionales.

2.8. Propiedades de vectores propios de matrices de Toeplitz y su aplicación al análisis espectral de series de tiempo

Ha habido un interés reciente en ciertas propiedades de las matrices de Toeplitz relativas a los ceros de polinomios que se forman a partir de los vectores propios con valores propios máximos o mínimos. Esto se debe a la conexión de las propiedades de análisis espectral de máxima entropía de Burg y el análisis de series de tiempo Auto-Regresivas (AR) estacionarias usando ecuaciones de Yule–Walker. Estas propiedades se han demostrado como un teorema de Carathéodory por Grenander y Szegő y por Robinson utilizando diferentes argumentos. Los conceptos desarrollados para la prueba se aplican luego al análisis espectral de series de tiempo y se hace una comparación entre los esquemas de Burg y de Pisarenko.

Una propiedad de matrices de Toeplitz

Sea \mathbf{B}_q una matriz de Toeplitz Hermitiana de tamaño $(q + 1) \times (q + 1)$:

$$\mathbf{B}_q = \begin{bmatrix} c_0 & c_1 & \cdots & c_q \\ c_1^* & c_0 & \cdots & c_{q-1} \\ \vdots & \vdots & \cdots & \vdots \\ c_q^* & c_{q-1}^* & \cdots & c_0 \end{bmatrix}$$

Indicamos los vectores columna como $[., \dots, ., ., .]$ y los vectores fila como $(., \dots, ., ., .)$. Para cada vector propio $f = [f_0, f_1, f_2, \dots, f_q]$ de \mathbf{B}_q , definimos el polinomio:

$$f(z) = f_0 + f_1 z + f_2 z^2 + \cdots + f_q z^q.$$

Nos referimos a este polinomio como el polinomio propio $f(z)$ correspondiente al vector propio f . El polinomio propio que corresponde a un vector propio con el máximo (mínimo) valor propio será llamado polinomio propio de valor máximo (polinomio propio de valor mínimo) para esa matriz particular.

Teorema 19. *Sea \mathbf{B}_q una matriz de Toeplitz hermitiana semidefinida positiva de tamaño $(q + 1) \times (q + 1)$ con un único valor propio cero. El polinomio propio de valor mínimo en z (correspondiente al valor propio cero) posee diferentes raíces en el círculo unitario $|z| = 1$.*

De esto tenemos el siguiente corolario.

Corolario 3. *Sea \mathbf{B}_q una matriz semidefinida positiva con un único valor máximo (mínimo) λ_{\max} (λ_{\min}). El polinomio propio de valor máximo (de valor mínimo) tiene ceros en el círculo unitario.*

Definimos la sucesión de matrices $\mathbf{B}_0, \mathbf{B}_1, \dots, \mathbf{B}_q$:

$$\mathbf{B}_i = \begin{bmatrix} c_0 & c_1 & \cdots & c_i \\ c_1^* & c_0 & \cdots & c_{i-1} \\ \vdots & \vdots & \cdots & \vdots \\ c_i^* & c_{i-1}^* & \cdots & c_0 \end{bmatrix}, \quad i = 0, 1, \dots, q.$$

Lema 7. $\mathbf{B}_0, \mathbf{B}_1, \dots, \mathbf{B}_{q-1}$ son definidas positivas y por lo tanto no singulares.

Definimos un vector ortogonal \mathbf{v}_i diferente de cero para las primeras i filas de \mathbf{B}_i y \mathbf{l}_i para la última fila de \mathbf{B}_i .

Lema 8. $\langle \mathbf{l}_i, \mathbf{v}_i \rangle \neq 0$ para $i = 0, 1, \dots, q-1$ y $\langle \mathbf{l}_q, \mathbf{v}_q \rangle = 0$.

Aquí $\langle \cdot, \cdot \rangle$ representa el producto interno entre dos vectores.

Lema 9. EL polinomio propio de valor mínimo para \mathbf{B}_q está dado por $cv_q(z)$ donde c es una constante diferente de cero y $v_q(z)$ es el polinomio en z formado por el vector $\mathbf{v}_q = [v_0, v_1, \dots, v_q]$:

$$v_q(z) = v_0 + v_1z + v_2z^2 + \dots + v_qz^q.$$

En vista del Lema 9, para probar el teorema 19 es suficiente demostrar que los ceros de $v_q(z)$ están sobre el círculo unitario. \mathbf{v}_i puede ser expresada como:

$$\mathbf{v}_i = [\mathbf{B}_{i-1}^{-1} \mathbf{a}_{i-1} | -1], \quad i = 1, 2, \dots, q, \quad (85)$$

donde $\mathbf{a}_{i-1} = [c_i c_{i-1} \dots c_1]$. Definimos

$$\begin{aligned} \mathbf{b}_{i-1} &= [c_1^* c_2^* \dots c_i^*], & \mathbf{a}_{i-1} &= [c_i c_{i-1} \dots c_1], \\ \mathbf{t}_i &= \mathbf{B}_i^{-1} \mathbf{a}_i, & \mathbf{s}_i &= \mathbf{B}_i^{-1} \mathbf{b}_i, \\ X_i &= \mathbf{a}_i^\top \mathbf{t}_i = \mathbf{a}_i^\top \mathbf{B}_i^{-1} \mathbf{a}_i, & Y_i &= \mathbf{b}_i^\top \mathbf{s}_i = \mathbf{b}_i^\top \mathbf{B}_i^{-1} \mathbf{b}_i, \\ Z_i &= \mathbf{b}_i^\top \mathbf{t}_i = \mathbf{b}_i^\top \mathbf{B}_i^{-1} \mathbf{a}_i, & Z_i^* &= \mathbf{a}_i^\top \mathbf{s}_i = \mathbf{a}_i^\top \mathbf{B}_i^{-1} \mathbf{b}_i, \\ \alpha_i &= (c_{i+1} - Z_{i-1}) / (c_0 - Y_{i-1}), & \beta_i &= (c_{i+1}^* - Z_{i-1}^*) / (c_0 - X_{i-1}). \end{aligned}$$

Usando las propiedades de las matrices con particiones, se puede demostrar:

$$\begin{aligned} \mathbf{t}_i &= [\alpha_i | \mathbf{t}_i - \alpha_i \mathbf{s}_{i-1}], \\ \mathbf{s}_i &= [\mathbf{s}_i - \beta_i \mathbf{t}_{i-1} | \beta_i]. \end{aligned}$$

Estas ecuaciones constituyen el algoritmo de Durbin. De

$$v_i = [\mathbf{t}_i | -1] \quad (86)$$

podemos establecer relaciones entre $v_i(z)$. Dado un vector

$$\mathbf{h} = (h_0, h_1, \dots, h_n),$$

definimos el *vector palindromo* de \mathbf{h} como

$$\mathbf{h}^\dagger = (h_n^*, h_{n-1}^*, \dots, h_1^*, h_0^*).$$

Lema 10. $\mathbf{t}_i^\dagger = \mathbf{s}_i$ y $t_i(z) = z^i s_i^*(1/z)$ donde $t_i(z)$ y $s_i(z)$ son polinomios formados por \mathbf{t}_i y \mathbf{s}_i .

Definimos el polinomio palindromo $f^\dagger(z)$ de $f(z) = f_0 + f_1z + f_2z^2 + \dots + f_nz^n$ como $f^\dagger(z) = f_n^* + f_{n-1}^*z + \dots + f_1^*z^{n-1} + f_0^*z^n$.

Lema 11. $v_i(z) = zv_{i-1}(z) - \alpha_{i-1}v_{i-1}^\dagger(z)$. Si $v_i(z) = u_0 + u_1z + u_2z^2 + \dots + u_iz^i$, entonces $u_0 = \alpha_{i-1}$ y $u_i = -1$.

Lema 12. $|\alpha_i| < 1$ para $i = 0, 1, \dots, q-2$ y $|\alpha_i| = 1$ para $i = q-1$.

Para demostrar que los ceros de $v_q(z)$ se encuentran en el círculo unitario, la prueba se hace en dos partes. Primero se demuestra que los ceros de $v_i(z)$ están dentro del círculo unitario para $i = 1, 2, \dots, q-2$ y luego que los ceros de $v_q(z)$ se encuentran sobre el círculo unitario. Después se demuestra que las raíces son distintas para completar la demostración del Teorema 1.

Teoremas de representación para matrices de Toeplitz

Procedemos a demostrar la importancia del Teorema 1 en análisis de series de tiempo, esto requiere los siguientes resultados.

Teorema 20. La matriz de Toeplitz \mathbf{B}_q del Teorema 19 se puede representar como sigue:

$$\mathbf{B}_q = \Theta^\top \mathbf{A} \Theta,$$

donde

$$\Theta = \begin{bmatrix} 1 & e^{j\theta_1} & e^{2j\theta_1} & \dots & e^{qj\theta_1} \\ 1 & e^{j\theta_2} & e^{2j\theta_2} & \dots & e^{qj\theta_2} \\ 1 & e^{j\theta_3} & e^{2j\theta_3} & \dots & e^{qj\theta_3} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & e^{j\theta_q} & e^{2j\theta_q} & \dots & e^{qj\theta_q} \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} a_1 & 0 & 0 & \dots & 0 \\ 0 & a_2 & 0 & \dots & 0 \\ 0 & 0 & a_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & a_q \end{bmatrix},$$

y a_i son números reales positivos.

Teorema 21. Sea \mathbf{B} una matriz de Toeplitz arbitraria semidefinida positiva con valor propio mínimo $\lambda_{\min} \geq 0$. Existen valores para $p, a_1, a_2, \dots, a_p, \theta_1, \theta_2, \dots, \theta_p$ en función de la estructura propia de \mathbf{B} tal que

$$\mathbf{B} = \Theta^\top \mathbf{A} \Theta + \lambda_{\min} \mathbf{I}, \quad (87)$$

donde

$$\Theta = \begin{bmatrix} 1 & e^{j\theta_1} & e^{2j\theta_1} & \dots & e^{(n-1)j\theta_1} \\ 1 & e^{j\theta_2} & e^{2j\theta_2} & \dots & e^{(n-1)j\theta_2} \\ 1 & e^{j\theta_3} & e^{2j\theta_3} & \dots & e^{(n-1)j\theta_3} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & e^{j\theta_p} & e^{2j\theta_p} & \dots & e^{(n-1)j\theta_p} \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} a_1 & 0 & 0 & \dots & 0 \\ 0 & a_2 & 0 & \dots & 0 \\ 0 & 0 & a_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & a_p \end{bmatrix},$$

Aquí n es la dimensión de \mathbf{B} .

Denotamos con p en el Teorema 21 el índice(\mathbf{B}) y se refiere a a'_i s y θ'_i s como las amplitudes propias y fases propias de \mathbf{B} . El vector propio v_p y el polinomio $v_p(z)$ se denotarán por $\text{minvec}(\mathbf{B})$ y $\text{minpol}(\mathbf{B}, z)$.

Corolario 4. Los elementos $b_{i,j}$ de una matriz de Toeplitz \mathbf{B} semidefinida positiva con $\text{índice}(\mathbf{B})=p < \infty$, un valor propio mínimo λ_{\min} , las amplitudes propias a_i , $i = 1, \dots, p$ y las fases propias θ_i , $i = 1, \dots, p$ se pueden representar como:

$$b_{i,j} = c_k = \sum_{i=1}^p a e^{kj\theta_i} + \lambda_{\min}\delta(k), \quad \text{donde } k = j - i. \quad (88)$$

Además las cantidades c_k obedecen la siguiente relación:

$$\sum_{i=0}^p f_i c_{k-p+i} = 0, \quad \text{para } k > p, \quad (89)$$

donde f_i están dados por:

$$f(z) = \sum_{i=0}^p f_i z^i = \prod_{i=0}^p [z - e^{j\theta_i}]. \quad (90)$$

Por supuesto, el vector (f_0, f_1, \dots, f_p) es $\text{minvec}(\mathbf{B})$ y $f(z) = \text{minpol}(\mathbf{B}, z)$.

Aplicación al análisis espectral de series de tiempo

Hay muchas implicaciones interesantes del teorema 21 en el análisis espectral. Si uno conoce los coeficientes de covarianza de una serie de tiempo dada hasta de n elementos, es posible extender estos coeficientes de una manera tal como para mantener el valor propio mínimo de \mathbf{B}_n . Así si el valor propio mínimo de \mathbf{B}_n es λ_{\min} , es posible calcular los nuevos coeficientes $c_{n+1}, c_{n+2}, \dots, c_{n+m}$ tales que la matriz de Toeplitz \mathbf{B}_{n+m} creada para esos elementos tiene valor propio mínimo λ_{\min} . El presente enfoque de análisis espectral extiende los coeficientes de una manera que no introduzca ruido adicional. En la práctica se puede suponer que λ_{\min} es el ruido blanco debido a los errores de observación y el ruido del sistema, y se presume que el cálculo de los coeficientes de covarianza se detienen cuando se alcanza este límite. También mediante la ampliación de la matriz de covarianza no quisiéramos disminuir el valor propio mínimo de la matriz de covarianza resultante esto significa que el error de predicción se reduce para la matriz extendida.

También es interesante observar que el análisis de máxima entropía defendida por Burg hace una extensión similar (pero no la misma) de los coeficientes de covarianza, y cada extensión de la matriz de covarianza maximiza la entropía (suponiendo proceso de Gauss) al maximizar el determinante. En este punto, es conveniente establecer la conexión entre nuestro enfoque y la técnica de máxima entropía de Burg. Se resuelve la ecuación siguiente:

$$\begin{bmatrix} c_0 & c_1 & c_2 & c_3 & \cdots & c_m \\ c_1^* & c_0 & c_1 & c_2 & \cdots & c_{m-1} \\ c_2^* & c_1^* & c_0 & c_1 & \cdots & c_{m-2} \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ c_m^* & c_{m-1}^* & \cdots & & & c_0 \end{bmatrix} \begin{bmatrix} 1 \\ g_1 \\ g_2 \\ \vdots \\ g_m \end{bmatrix} = \begin{bmatrix} P_m \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

que es equivalente a:

$$\begin{bmatrix} c_0 & c_1 & c_2 & \cdots & c_{m-1} \\ c_1^* & c_0 & c_1 & \cdots & c_{m-2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ c_{m-1}^* & c_{m-2}^* & \cdots & & c_0 \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_m \end{bmatrix} = \begin{bmatrix} c_1^* \\ c_2^* \\ \vdots \\ c_m^* \end{bmatrix},$$

y $P_m = c_0 + c_1g_1 + c_2g_2 + \dots + c_mg_m$.

Podemos escribir $\mathbf{B}_{m-1}\mathbf{h}_{m-1} = -\mathbf{b}_{m-1}$ donde $\mathbf{h}_{m-1} = [g_1g_2 \dots g_m]$. Por tanto $\mathbf{h}_{m-1} = -\mathbf{s}_{m-1} = \mathbf{t}_{m-1}^\dagger$. Así $g_m = [1|\mathbf{h}_{m-1}] = -v_m^\dagger$. En cada paso el espectro de potencia se calcula como:

$$P_{AR}(\lambda) = P_m/|Y_m(e^{-j\lambda})|^2,$$

donde $Y_m(z)$ es un polinomio formado por los coeficientes de g_m .

Por lo tanto nuestro desarrollo mediante la estructura propia de las matrices Toeplitz es aplicable al análisis de máxima entropía y se analiza una perspectiva adicional. También se relaciona el enfoque de Burg con el de Pisarenko. El espectro de Pisarenko se puede calcular usando la ecuación (88).

Para una serie de tiempo determinada $z(t)$ con matriz de covarianza \mathbf{B} cuyas propiedades se establecen en el corolario 4, el coeficiente de correlación $B(\tau) = \mathbb{E}[z(t+\tau)z^*(t)]$ puede ser tomado como:

$$B(\tau) = \sum_{i=1}^p a_i e^{-j\tau\theta_i} + \lambda_{\min}\delta(\tau). \quad (91)$$

El espectro de Khinchin-Wiener para $z(t)$ de (91) será:

$$P_p(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} B(\tau) e^{-j\lambda\tau} d\tau = \frac{1}{2\pi} \left(\sum_{i=0}^p \alpha_i \delta(\lambda + \theta_i) + \lambda_{\min} \right). \quad (92)$$

Para deducir el espectro de potencia en (92) de \mathbf{B} , encontramos el valor propio mínimo λ_{\min} , de la matriz $\mathbf{C} = \mathbf{B} - \lambda_{\min}\mathbf{I}$, y encontramos θ_i mediante la resolución de la polinomio propio de valor mínimo $f(z)$ que corresponde a la primer componente principal de \mathbf{C} con un valor propio único de cero. Los coeficientes a_i pueden resolverse mediante la formación de \mathbf{A} y $\mathbf{\Theta}$ y el calculando:

$$\mathbf{A} = (\mathbf{\Theta}\mathbf{\Theta}^\top)^{-1}\mathbf{\Theta}(\mathbf{B} - \lambda_{\min}\mathbf{I})\mathbf{\Theta}^\top(\mathbf{\Theta}\mathbf{\Theta}^\top)^{-1}. \quad (93)$$

En general los espectros calculados por análisis de máxima entropía Burg y la técnica Pisarenko no son lo mismo y última es siempre discreta. A continuación vamos a tratar de resolver este dilema e indicar cuál es la técnica que en general puede ser mejor.

Máxima entropía de Burg

Paso 1. Calcule los coeficientes de correlación para el serie de tiempo $z(t)$

$$B(\tau) = \mathbb{E}[z(t+\tau)z^*(t)]$$

Sea $c_k = B(-k)$ y $c_{-k} = c_k^*$. De la matriz \mathbf{B} cuyo (i, j) -ésimo elemento $b_{i,j}$ está dada por:

$$b_{i,j} = c_k \quad \text{donde} \quad k = j - i.$$

Set m , el parámetro de iteración, para 1.

Paso 2. Calcule $v_m(z)$ el polinomio correspondiente a v_m definido en (85) y (86), de una manera recursiva utilizando lema 11. Obtener su palíndromo $v_m^\dagger(z)$ y hacer

$$Y_m(z) = -v_m^\dagger(z) = 1 + g_1z + g_2z^2 + \dots + g_mz^m. \quad (94)$$

El espectro de potencia se calcula como:

$$P_{AR}(\lambda) = P_m/|Y_m(z)|^2$$

donde $P_m = c_0 + c_1g_1 + \dots + c_mg_m$.

Paso 3. Utilice un criterio para la determinación del orden del proceso autorregresivo que mejor aproxime a $z(t)$ y terminar si se alcanza este orden.

Técnica de Pisarenko

Paso 1. Igual que la etapa 1 del método de Burg.

Paso 2. Calcule λ_{\min}^m para \mathbf{B}_m , la $(m+1) \times (m+1)$ menor componente principal de \mathbf{B} . A partir de $\mathbf{C}_m = \mathbf{B}_m - \lambda_{\min}^m \mathbf{I}$. Si $\lambda_{\min}^{m-1} = \lambda_{\min}^m$, parar. Calcular $\text{minpol}(\mathbf{C}_m)$, $\text{minvec}(\mathbf{C}_m)$, las fases propias θ_i^m , $i = 1, \dots, m$ y las amplitudes propias a_i^m , $i = 1, \dots, m$ de \mathbf{C}_m . Formar el espectro:

$$P_P^m = (1/2\pi) \left(\sum_{i=1}^m \alpha_i^m \delta(\lambda + \theta_i^m) + \lambda_{\min}^m \right).$$

Paso 3. Detención del criterio. Sea $\text{minvec}(\mathbf{C}_m) = (f_0, f_1, \dots, f_m)$. Calcule:

$$\delta_m = f_0c_1 + f_1c_2 + f_2c_3 + \dots + f_m c_{m+1}.$$

Si $\delta_m < \varepsilon$, una cantidad pequeña preseleccionado, terminar. De lo contrario incrementar m y vaya al paso 2.

El criterio de parada para la técnica de Pisarenko se basa en (89). Una de las diferencias entre Burg y de las técnicas de Pisarenko es que esta última calcula el valor propio mínimo en cada paso y esto se suma a la carga computacional sustancialmente. Básicamente Burg intenta modelar la serie de tiempo dada $z(t)$ con un proceso AR utilizando los parámetros $1, g_1, g_2, \dots, g_m$ en (94). Pisarenko intenta aproximar la matriz de covarianza como en (87) y derivar el espectro de potencia de los elementos de esta matriz utilizando relación (88), (91) y (92). La complejidad del procedimiento de Burg es esencialmente de $o(m^2)$ donde M es el orden del proceso autorregresivo que mejor se adapte la serie de tiempo dada. El procedimiento de Pisarenko es de mayor complejidad y esto es debido a *i*) el cálculo del valor propio mínimo en cada paso, *ii*) cálculo de vector propio que corresponde al valor propio cero (que es de $o(m^2)$ en la m -ésima iteración) y *iii*) después de la resolución de (93) se alcanza el valor correcto de m . El cálculo de valor propio mínimo puede reducirse observando el valor propio mínimo de las formas iteración anterior un límite superior para la presente iteración. A pesar de su enorme complejidad conviene preferir el esquema de Pisarenko sobre el esquema de Burg en el análisis espectral general debido a los siguientes argumentos.

Considere la serie de tiempo:

$$z(t) = \sum_{i=0}^p \xi_i e^{-j\theta_i t} + \eta, \quad (95)$$

donde ξ_i and η son variables aleatorias independientes con media cero y $\mathbb{E}(|\xi_i|^2) = a_i$ y $\mathbb{E}(|\eta|^2) = \lambda$. Suponga que los coeficientes de correlación se calculan con exactitud y no se introducen errores debido a los intervalos de observación finitos. Para esta serie de tiempo el esquema de Pisarenko obtiene el espectro correcto mientras que el procedimiento de Burg diverge ya que el error mínimo del polinomio de predicción tiene ceros en el círculo unitario. El esquema de Burg también diverge cada vez que la serie de tiempo es singular. Existe otra razón por la cual el esquema de Pisarenko puede ser preferible. Se sabe que cualquier serie de tiempo estacionaria $z(t)$ puede ser aproximada para cualquier t en el intervalo $-T \leq t \leq T$ por una suma de componentes armónicos con amplitudes aleatorias (como en (95))tales que:

$$\mathbb{E} \left| z(t) - \sum_{i=0}^p \xi_i e^{-j\theta_i t} \right|^2 < \varepsilon$$

para cualquier $\varepsilon > 0$ dado. Por lo tanto, puede verse que el esquema de Pisarenko eventualmente converge al espectro verdadero.

3. Ejercicios

En esta sección se presentan varios ejercicios resueltos, que ayudarán a comprender algunos temas de los resúmenes presentados en la sección anterior.

Problema 1. Considera el proceso autorregresivo de orden 3 definido por

$$X_n = \begin{cases} \frac{1}{2}X_{n-1} + \frac{1}{9}X_{n-2} - \frac{1}{18}X_{n-3} + W_n & \text{si } n = 0, 1, \dots; \\ 0 & \text{si } n < 0. \end{cases}$$

Aquí $(W_n)_{n=0}^{\infty}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas, tales que $W_n \sim N(0, 1)$ para cada $n \in \mathbb{N}$.

- Escribe el proceso $AR(3)$ para los primeros siete tiempos.
- Escribe la forma matricial de las ecuaciones que se obtuvieron en el inciso anterior.
- Hay que notar que la matriz del inciso anterior es una matriz de Toeplitz triangular inferior, escribe el símbolo generador de dicha matriz.
- Un proceso $AR(3)$ es asintóticamente estacionario si y sólo si los módulos de las raíces del polinomio autorregresivo $a(t)$ son mayores que 1. Verifica que el proceso autorregresivo es asintóticamente estacionario.

Solución 1. a) Las ecuaciones que describen el proceso para los primeros siete tiempos son las siguientes:

$$\begin{aligned} X_0 &= W_0, \\ X_1 &= \frac{1}{2}X_0 + W_1, \\ X_2 &= \frac{1}{2}X_1 + \frac{1}{9}X_0 + W_2, \\ X_3 &= \frac{1}{2}X_2 + \frac{1}{9}X_1 - \frac{1}{18}X_0 + W_3, \\ X_4 &= \frac{1}{2}X_3 + \frac{1}{9}X_2 - \frac{1}{18}X_1 + W_4, \\ X_5 &= \frac{1}{2}X_4 + \frac{1}{9}X_3 - \frac{1}{18}X_2 + W_5, \\ X_6 &= \frac{1}{2}X_5 + \frac{1}{9}X_4 - \frac{1}{18}X_3 + W_6. \end{aligned}$$

Pasamos X_0, \dots, X_6 al lado izquierdo:

$$\begin{aligned}
 X_0 &= W_0, \\
 -\frac{1}{2}X_0 + X_1 &= W_1, \\
 -\frac{1}{9}X_0 - \frac{1}{2}X_1 + X_2 &= W_2, \\
 \frac{1}{18}X_0 - \frac{1}{9}X_1 - \frac{1}{2}X_2 + X_3 &= W_3, \\
 \frac{1}{18}X_1 - \frac{1}{9}X_2 - \frac{1}{2}X_3 + X_4 &= W_4, \\
 \frac{1}{18}X_2 - \frac{1}{9}X_3 - \frac{1}{2}X_4 + X_5 &= W_5, \\
 \frac{1}{18}X_3 - \frac{1}{9}X_4 - \frac{1}{2}X_5 + X_6 &= W_6.
 \end{aligned}$$

b) El sistema de ecuaciones del inciso anterior se puede escribir en forma matricial:

$$\begin{bmatrix}
 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 -\frac{1}{2} & 1 & 0 & 0 & 0 & 0 & 0 \\
 -\frac{1}{9} & -\frac{1}{2} & 1 & 0 & 0 & 0 & 0 \\
 \frac{1}{18} & -\frac{1}{9} & -\frac{1}{2} & 1 & 0 & 0 & 0 \\
 0 & \frac{1}{18} & -\frac{1}{9} & -\frac{1}{2} & 1 & 0 & 0 \\
 0 & 0 & \frac{1}{18} & -\frac{1}{9} & -\frac{1}{2} & 1 & 0 \\
 0 & 0 & 0 & \frac{1}{18} & -\frac{1}{9} & -\frac{1}{2} & 1
 \end{bmatrix}
 \begin{bmatrix}
 X_0 \\
 X_1 \\
 X_2 \\
 X_3 \\
 X_4 \\
 X_5 \\
 X_6
 \end{bmatrix}
 =
 \begin{bmatrix}
 W_0 \\
 W_1 \\
 W_2 \\
 W_3 \\
 W_4 \\
 W_5 \\
 W_6
 \end{bmatrix}.$$

c) La matriz de tamaño 7×7 es una matriz de Toeplitz triangular inferior, cuyo símbolo generador es:

$$a(t) = 1 - \frac{1}{2}x - \frac{1}{9}x^2 + \frac{1}{18}x^3.$$

Este polinomio también se conoce como el *polinomio autorregresivo*.

d) Para verificar que el proceso es asintóticamente estacionario, basta ver que el módulo de las raíces del polinomio autorregresivo es mayor que 1. El polinomio a se puede factorizar:

$$a(t) = 1 - \frac{1}{2}x - \frac{1}{9}x^2 + \frac{1}{18}x^3 = \frac{1}{18}(x-2)(x-3)(x+3).$$

De aquí fácilmente se verifica que las raíces del polinomio a son:

$$x_1 = 2, \quad x_2 = 3, \quad \text{y} \quad x_3 = -3.$$

Es evidente que $|x_1| > 1$, $|x_2| > 1$ y $|x_3| > 1$, por tanto el proceso es asintóticamente estacionario.

Problema 2. Sean s una señal útil y n una señal que presenta el ruido. La señal de ruido se presenta en la práctica, es una señal cuya amplitud varía al azar. Supongamos que el ruido tiene un valor promedio de cero, es decir,

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} n(t) dt = 0.$$

Demostrar que si s y n no están correlacionadas, entonces

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} s(t)n(t - \tau) dt = 0, \quad \text{para cada } \tau.$$

Solución 2. Como s y n no están correlacionadas, entonces se tiene que

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} s(t)n(t - \tau) dt = \left[\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} s(t) dt \right] \left[\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} n(t) dt \right].$$

Por la hipótesis, el segundo factor es cero, así que el producto es cero.

Problema 3. Demostrar que la función de autocorrelación promedio de la suma de la señal y del ruido es la suma de las funciones individuales de autocorrelacion de la señal y del ruido respectivamente. Suponga que la señal y el ruido no están correlacionados.

Solución 3. Sea $f(t) = s(t) + n(t)$, donde s es la señal útil y n es el ruido, entonces tenemos

$$\begin{aligned} \bar{R}_{ff}(\tau) &= \lim_{T \rightarrow \infty} \int_{-T/2}^{T/2} f(t)f(t - \tau) dt \\ &= \lim_{T \rightarrow \infty} \int_{-T/2}^{T/2} (s(t) + n(t))(s(t - \tau) + n(t - \tau)) dt \\ &= \lim_{T \rightarrow \infty} \int_{-T/2}^{T/2} [s(t)s(t - \tau) + s(t)n(t - \tau) + n(t)s(t - \tau) + n(t)n(t - \tau)] dt \\ &= \bar{R}_{ss}(\tau) + \bar{R}_{sn}(\tau) + \bar{R}_{ns}(\tau) + \bar{R}_{nn}(\tau). \end{aligned}$$

Dado que la señal y el ruido no están correlacionadas se tiene que $\bar{R}_{sn}(\tau) = 0 = \bar{R}_{ns}(\tau)$, por tanto,

$$\bar{R}_{ff}(\tau) = \bar{R}_{ss}(\tau) + \bar{R}_{nn}(\tau).$$

Problema 4. Sea f una señal transmitida y g la señal recibida. Demostrar que la función de correlación promedio entre la señal transmitida y la señal recibida, en la misma función de correlación promedio entre la señal útil y la señal recibida. Suponga que la señal útil no está correlacionada con la señal recibida.

Solución 4. Sea $f(t) = s(t) + n(t)$ la señal transmitida, en donde s es la señal útil y n es el ruido, la correlación promedio de la señal transmitida y la señal útil es:

$$\begin{aligned}\bar{R}_{fg}(\tau) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} f(t)g(t - \tau) dt \\ &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} (s(t) + n(t))g(t - \tau) dt \\ &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} s(t)g(t - \tau) dt + \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} n(t)g(t - \tau) dt \\ &= \bar{R}_{sg}(\tau) + \bar{R}_{ng}.\end{aligned}$$

Pero, n y g no están correlacionadas, así que $\bar{R}_{ng}(\tau) = 0$ para cada τ . Por lo tanto,

$$\bar{R}_{fg}(\tau) = \bar{R}_{sg}(\tau) + 0 = \bar{R}_{sg}(\tau).$$

Problema 5. Sea f una función periódica, con período T_* . Demostrar que

$$\bar{R}_{ff}(\tau) = \frac{1}{T_*} \int_{-T_*/2}^{T_*/2} f(t)f(t - \tau) dt.$$

Solución 5. Sea $(T_n)_{n=1}^{\infty}$ la sucesión definida como $T_n := nT_*$, notemos que

$$\bar{R}_{ff}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} f(t)f(t - \tau) dt = \lim_{n \rightarrow \infty} \frac{1}{T_n} \int_{-T_n/2}^{T_n/2} f(t)f(t - \tau) dt.$$

Por otro lado,

$$\int_{-T_n/2}^{T_n/2} f(t)f(t - \tau) dt = n \int_{-T_*/2}^{T_*/2} f(t)f(t - \tau) dt,$$

entonces

$$\bar{R}_{ff}(\tau) = \lim_{n \rightarrow \infty} \frac{n}{nT_*} \int_{-T_*/2}^{T_*/2} f(t)f(t - \tau) dt = \frac{1}{T_*} \int_{-T_*/2}^{T_*/2} f(t)f(t - \tau) dt.$$

Problema 6. Demostrar que si f y g son funciones reales y periódicas de la clase $C^1(\mathbb{R})$, ambas con el mismo periodo T , entonces

$$\bar{R}_{fg}(\tau) = \sum_{n=-\infty}^{\infty} (a_n^* b_n) e^{-in\omega_0\tau},$$

donde $\omega_0 = \frac{2\pi}{T}$ y a_n, b_n son los coeficientes de Fourier de f y g respectivamente, y a_n^* denota el complejo conjugado de a_n .

Solución 6. Puesto que f y g son funciones periódicas se tiene que

$$\bar{R}_{fg}(\tau) = \frac{1}{T} \int_{-T/2}^{T/2} f(t)g(t-\tau) dt. \quad (96)$$

La condición $f, g \in C^1(\mathbb{R})$ garantiza que las siguientes series de Fourier de f y g convergen absolutamente:

$$f(t) = \sum_{n=-\infty}^{\infty} a_n e^{in\omega_0 t}, \quad g(t) = \sum_{n=-\infty}^{\infty} b_n e^{in\omega_0 t},$$

donde

$$a_n = \frac{1}{T} \int_{-T/2}^{T/2} f(t) e^{-in\omega_0 t} dt \quad y \quad b_n = \frac{1}{T} \int_{-T/2}^{T/2} g(t) e^{-in\omega_0 t} dt.$$

Sustituyendo estas series de Fourier en (96) se tiene

$$\bar{R}_{fg} = \frac{1}{T} \int_{-T/2}^{T/2} f(t) \left[\sum_{n=-\infty}^{\infty} b_n e^{in\omega_0(t-\tau)} \right] dt.$$

La convergencia absoluta permite intercambiar el orden de la sumatoria y de la integral:

$$\bar{R}_{fg}(\tau) = \sum_{n=-\infty}^{\infty} b_n e^{-in\omega_0\tau} \left[\frac{1}{T} \int_{-T/2}^{T/2} f(t) e^{in\omega_0 t} dt \right].$$

La expresión dentro del corchete es el complejo conjugado de a_n . Por tanto,

$$\bar{R}_{fg}(\tau) = \sum_{n=-\infty}^{\infty} (a_n^* b_n) e^{-in\omega_0\tau}.$$

Problema 7. Demostrar que si h es una función real y periódica con período T de clase $C^1(\mathbb{R})$, entonces

$$\bar{R}_{hh}(\tau) = \sum_{n=-\infty}^{\infty} |c_n|^2 e^{in\omega_0\tau},$$

donde $\omega_0 = \frac{2\pi}{T}$ y c_n son los coeficientes complejos de Fourier de h .

Solución 7. Si hacemos $f = g = h$ en el Problema 6, se obtiene

$$\bar{R}_{hh}(\tau) = \sum_{n=-\infty}^{\infty} c_n^* c_n e^{-in\omega_0\tau} = \sum_{n=-\infty}^{\infty} |c_n|^2 e^{-in\omega_0\tau} = \sum_{n=-\infty}^{\infty} |c_n|^2 e^{in\omega_0\tau},$$

puesto que $|c_{-n}|^2 = |c_n|^2$.

Problema 8. Hallar la densidad espectral de potencia de una función T -periódica f de clase $C^1(\mathbb{R})$.

Solución 8. Supongamos que la serie de Fourier de la función f , es

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{in\omega_0 t}, \quad \text{donde } \omega_0 = \frac{2\pi}{T}.$$

En el Problema 7, se demostró que la función de autocorrelación promedio de f está dada por

$$\bar{R}_{ff}(\tau) = \sum_{n=-\infty}^{\infty} |c_n|^2 e^{in\omega_0\tau}.$$

La transformada de Fourier de la función $\tau \mapsto e^{in\omega_0\tau}$ es $\delta(\omega - n\omega_0)$, por eso

$$P(\omega) = \mathfrak{F}[\bar{R}_{ff}](\omega) = \sum_{n=-\infty}^{\infty} 2\pi |c_n|^2 \delta(\omega - n\omega_0).$$

Referencias

- [1] BOX, G. E. P.; JENKINS, G. M.; REINSEL G. C. (2008): *Time Series Analysis, Forecasting and Control*, 4th edition, Wiley.
ISBN: 978-0-470-27284-8.
De este libro se utilizó el Capítulo 2 (pp 21–46).
- [2] GRAY, R. M. (2006): *Toeplitz and Circulant Matrices: A Review*. Foundations and Trends in Communications and Information Theory, Vol. 2, Issue 3, pp 155–239.
<http://ee.stanford.edu/~gray/toeplitz.html>
De este texto se utilizó el Capítulo 6.
- [3] AVRAM, F. (1988): On bilinear forms in Gaussian random variables and Toeplitz matrices. *Probability Theory and Related Fields*, Vol. 79, Issue 1, pp 37–45.
<http://dx.doi.org/10.1007/BF00319101>
- [4] SHLENS, J. (2014): A tutorial on principal component analysis. arXiv:1404.1100 [cs.LG].
<http://arxiv.org/pdf/1404.1100.pdf>
- [5] RAWLINGS, R.; ROHRBAUGH, J. W.; BEGLEITER, H.; ECKARDT, M. J. (1986): Spectral methods for principal components analysis of event related brain potentials. *Computers and Biomedical Research*, Vol. 19, Issue 6, pp 497–507.
[http://dx.doi.org/10.1016/0010-4809\(86\)90024-8](http://dx.doi.org/10.1016/0010-4809(86)90024-8)
- [6] HAUSER, M. A.; HÖRMANN, W.; KUNST, R. M.; LENNEIS, J. (1994): A note on generation, estimation and prediction of stationary processes. *CompStat 1994, Proceedings in computational statistics*, ed. R. Dutter and W. Grossmann, Heidelberg: Physica, pp 323–328.
http://dx.doi.org/10.1007/978-3-642-52463-9_36
<http://epub.wu.ac.at/142/>
- [7] GAZZAH, H.; REGALIA, P. A.; DELMAS, J.-P. (2001): Asymptotic Eigenvalue Distribution of Block Toeplitz Matrices and Application to Blind SIMO Channel Identification. *Information Theory, IEEE Transactions*, 2001, Vol. 47, Issue 3, pp 1243–1251.
<http://dx.doi.org/10.1109/18.915697>
- [8] REDDI, S. S. (2001): Eigenvector properties of Toeplitz matrices and their application to spectral analysis of time series. *Signal Processing*, Vol. 7, Issue 1, pp 45–56.
[http://dx.doi.org/10.1016/0165-1684\(84\)90023-9](http://dx.doi.org/10.1016/0165-1684(84)90023-9)
- [9] HOEL, P. G.; PORT, S. C.; STONE, C. J. (1986): *Introduction to Stochastic Processes*. Waveland Pr Inc. ISBN: 978-0-881-33267-4.
De este libro se utilizaron las páginas 111, 119, 122.
- [10] HSU, H. P. *Análisis de Fourier*. Addison-Wesley Iberoamericana, S.A., 1987.
ISBN: 978-0-201-02942-0.
De este libro se utilizaron las páginas 74, 94–95, 166–173.