



Análisis de Señales en Geofísica

13° Clase

Estimación del Espectro de Potencia

Prof. Ricardo C. Rebollo



Facultad de Ciencias Astronómicas y Geofísicas,
Universidad Nacional de La Plata, Argentina





Introducción:

La estimación del espectro de potencia de una señal, con su amplia variedad de aplicaciones, puede ser comprendida fácilmente en términos del teorema de corrimiento de fase de la transformada de Fourier.

Para entender porqué esto es así, veamos este problema de estimación desde un punto de vista práctico. Analicemos cuáles son las consecuencias de promediar en tiempo distintas realizaciones sincronizadas de una misma señal determinística (*signal like signal*) que está contaminada con ruido aleatorio no correlacionable (*noise like signal*). Y luego veamos qué solución podemos darle al problema en el caso en que la componente coherente de las realizaciones no está sincronizada en tiempo.



Relación Señal-Ruido S/R:

Imaginemos que estamos realizando un experimento controlado. El experimento es confiable, es repetible pero la componente determinística o coherente que queremos obtener está contaminada por una componente de ruido aleatorio. El cociente entre la amplitud de la componente coherente que nos interesa y la amplitud de la componente aleatoria no deseada, recibe el nombre de relación señal-ruido (*S/N signal-to-noise ratio*).

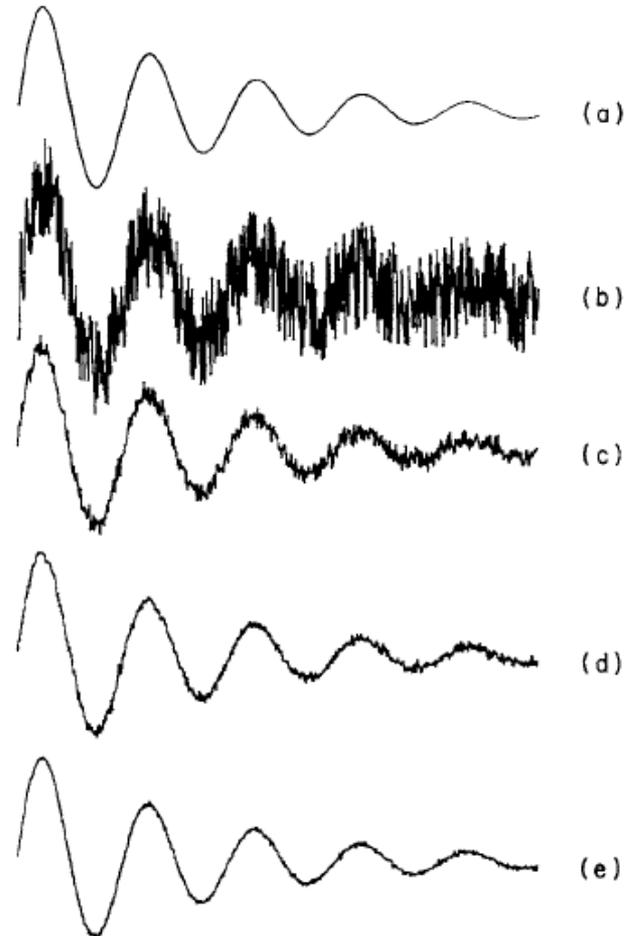
Supongamos que las repetidas realizaciones tienen una mala relación señal ruido. La componente determinística es repetible y está sincronizada en cada experimento, sin embargo el ruido es aleatorio y cambia de un experimento al siguiente. Como resultado de cada realización del experimento obtenemos una secuencia que contiene una misma señal determinística más ruido aleatorio cambiante. Una forma sencilla de mejorar la relación S/R es repitiendo varias veces el experimento y promediarlos. Si la señal determinística está sincronizada en todos los experimentos, al promediarlos, el nivel de la señal se mantendrá constante, mientras que, como sabemos de estadística aplicada, la amplitud del ruido disminuirá en proporción a $1/\sqrt{N}$.

Mejora de La Relación S/R:

Veamos un ejemplo de lo que sucede cuando promediamos una señal que posee una componente coherente más otra componente de ruido aleatorio:

- a) Componente coherente de la señal.
- b) Señal contaminada con ruido aleatorio.
- c) Promedio de $N=100$ observaciones.
- d) Promedio de $N=2500$ observaciones.
- e) Promedio de $N=10000$ observaciones.

A medida que aumenta el número de observaciones que se promedian, el nivel de la componente coherente se mantiene constante, mientras que el nivel del ruido se reduce en $1/\sqrt{N}$. Es decir se produce una mejora de la relación S/N en \sqrt{N} .





Experimentos no sincronizados:

Qué sucedería si la componente determinística no estuviera sincronizada en los distintos experimentos. Obviamente ya no podríamos promediar los experimentos para mejorar la relación S/R. El teorema del corrimiento de fase nos dice que el espectro de amplitud de una señal es insensible a los corrimientos arbitrarios de tiempo:

$$f(t) \Leftrightarrow F(\Omega) = |F(\Omega)|e^{i\phi(\Omega)}$$

$$f(t - t_0) \Leftrightarrow F(\Omega) = |F(\Omega)|e^{i[\phi(\Omega) - \Omega t_0]}$$

Es decir, ante un corrimiento en tiempo, la fase de la señal cambia linealmente con la frecuencia pero el espectro de amplitud permanece invariante.

Por lo tanto si la señal coherente no está sincronizada en todos los experimentos y no podemos promediarlos, no todo está perdido, todavía podemos promediar los espectros de amplitud. Es decir que la estimación que hagamos del espectro de amplitud irá mejorando a medida que apilemos más y más experimentos. El precio que debemos pagar por esta falta de sincronía entre los experimentos es la pérdida de la información de fase. Lo cual no es tan malo ya que en muchas aplicaciones el espectro de amplitud nos provee con información sumamente útil.



El Problema de la estimación de la Densidad del Espectro de Potencia:

Tratamos con dos tipos de señales con características sumamente diferentes: las señales determinísticas y el ruido aleatorio.

Si el ruido innovativo ε_n es un proceso aleatorio, estacionario y ergódico, del cual disponemos una única realización de N muestras, en consecuencia sólo podremos calcular un estimador de su autocovarianza:

$$\hat{p}_k = \frac{1}{N} \sum_{n=0}^{N-k-1} \varepsilon_n^* \varepsilon_{n+k}$$

Existen dos problemas fundamentales con este estimador de la autocovarianza. El primero es que a medida que el *lag* k aumenta, la porción de la secuencia que se solapa con otra porción de la secuencia donde no se disponen datos y se tratan incorrectamente como si fueran nulos o ceros, es cada vez mayor. Esto tiene como consecuencia un deterioro de la calidad del estimador cada vez más importante a medida que el *lag* k aumenta.

El segundo problema está relacionado con el comportamiento de la densidad del espectro de potencia calculado a partir de este estimador de la autocovarianza aplicándole la transformada de Fourier.



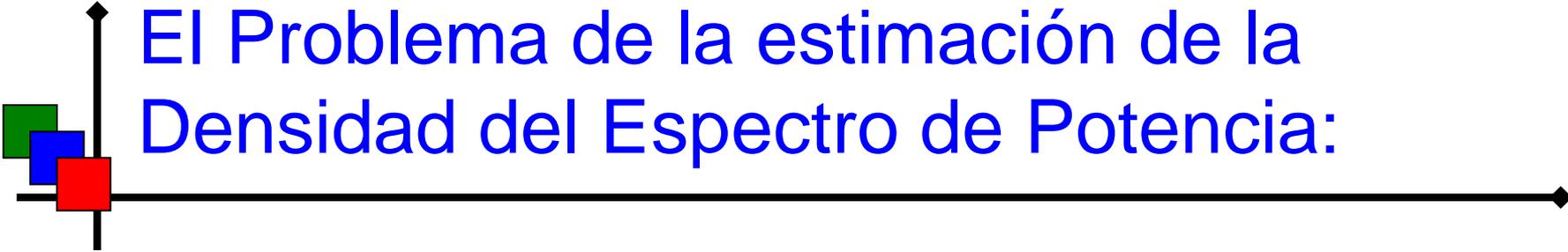
El Problema de la estimación de la Densidad del Espectro de Potencia:

A primera vista parecería posible que un estimador razonable de la densidad del espectro de potencia del proceso aleatorio pudiera obtenerse calculando la transformada de Fourier del estimador de la autocovarianza y que la calidad de este estimador, mejorará a medida que aumente la longitud de la muestra.

Sin embargo sorprendentemente esta mejora no se produce. Para ver el porqué de esto consideremos un proceso aleatorio estacionario ε_n con valor medio μ_ε nulo y varianza σ_ε^2 constante:

$$\sigma_\varepsilon^2 = \lim_{M \rightarrow \infty} \frac{1}{(2M + 1)} \sum_{n=-M}^M (\varepsilon_n - \mu_\varepsilon)^2$$

Consideremos además que el proceso es no correlacionable, es decir que las variables aleatorias que conforman el proceso son estadísticamente independientes.



El Problema de la estimación de la Densidad del Espectro de Potencia:

Como sólo tenemos una realización de longitud N de este proceso aleatorio, el estimador de la autocovarianza estará dado por:

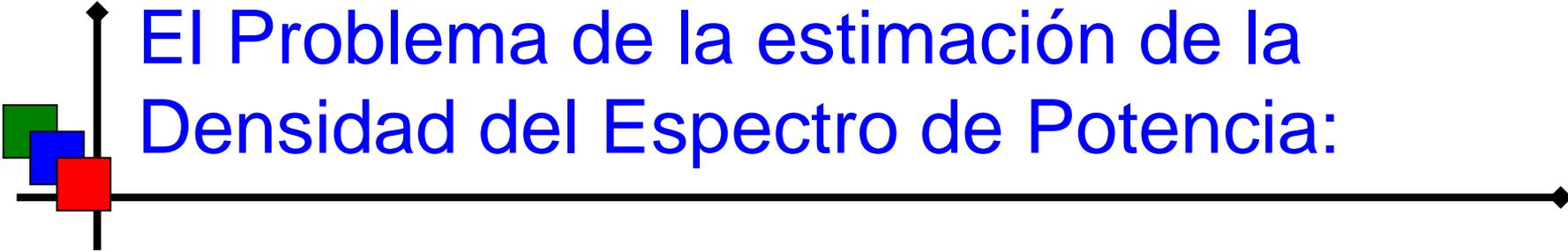
$$\hat{p}_k = \frac{1}{N} \sum_{n=0}^{N-k-1} \varepsilon_n^* \varepsilon_{n+k}$$

Para *lag* cero tendremos:

$$\hat{p}_0 = \frac{1}{N} \sum_{n=0}^{N-1} |\varepsilon_n|^2 = \hat{\sigma}_\varepsilon^2$$

Las fluctuaciones o desviación media estándar de este estimador van a disminuir a medida que N aumente en proporción a $1/\sqrt{N}$:

$$\hat{p}_0 = \hat{\sigma}_\varepsilon^2 \pm \frac{\hat{\sigma}_\varepsilon^2}{\sqrt{N}} = \hat{\sigma}_\varepsilon^2 \pm \hat{\sigma}_{\hat{p}_0}$$



El Problema de la estimación de la Densidad del Espectro de Potencia:

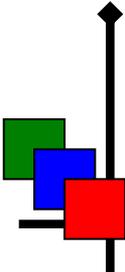
Para *lags* k distintos de cero, como el proceso es no correlacionable, la autocovarianza será igual a cero más menos la fluctuación aleatoria:

$$\hat{p}_k = 0 \pm \frac{N - k}{N} \frac{\hat{\sigma}_\varepsilon^2}{\sqrt{N}} = 0 \pm \hat{\sigma}_{\hat{p}_k}$$

Donde vemos que las fluctuaciones disminuyen en proporción al número de términos $(N - k)/N$ debido al deterioro de la calidad del estimador con el aumento del *lag* k .

La transformada de Fourier del estimador de la autocovarianza nos da un estimador de la densidad del espectro de potencia conocido como peridiograma:

$$\hat{P}(\omega) = \sum_{k=-(N-1)}^{(N-1)} \hat{p}_k e^{-i\omega k} = \hat{p}_0 + 2 \sum_{k=1}^{(N-1)} \hat{p}_k \cos(\omega k)$$



El Problema de la estimación de la Densidad del Espectro de Potencia:

Es decir que el estimador de la densidad del espectro de potencia es la suma de números aleatorios. Esta suma es también un número aleatorio que tiene una fluctuación estadística proporcional al número de términos $\sqrt{N-1}$:

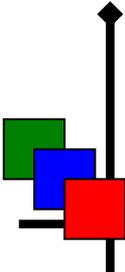
$$\hat{P}(\omega) = \hat{\sigma}_\varepsilon^2 \pm \sqrt{N-1} \frac{\hat{\sigma}_\varepsilon^2}{\sqrt{N}} \approx \hat{\sigma}_\varepsilon^2 \pm \hat{\sigma}_\varepsilon^2$$

Cuando N aumenta las fluctuaciones del peridiograma no disminuyen, sino que tienden a un valor constante igual a $\hat{\sigma}_\varepsilon^2$, es decir que aumentando la longitud de la secuencia no se produce una mejora en el estimador de la densidad del espectro de potencia. Es cierto que al aumentar la longitud aumenta la resolución en frecuencia, pero no aumenta la resolución estadística, ya que las fluctuaciones de este estimador del espectro de potencia son del mismo orden que el valor del espectro de potencia que queremos estimar. En lenguaje estadístico se dice que **el peridiograma es un estimador no consistente de la densidad del espectro de potencia**, independientemente de la longitud de la secuencia. Al aumentar la longitud de la secuencia solo logramos mejorar la resolución espectral sin ninguna mejora de la resolución estadística.

El Problema de la estimación de la Densidad del Espectro de Potencia:

Variable Aleatoria	Fluctuaciones
ε_n	$\hat{\sigma}_\varepsilon$
\hat{p}_k	$\frac{N - k}{N} \frac{\hat{\sigma}_\varepsilon^2}{\sqrt{N}}$
$\hat{P}(\omega)$	$\sqrt{N - 1} \frac{\hat{\sigma}_\varepsilon^2}{\sqrt{N}} \approx \hat{\sigma}_\varepsilon^2$

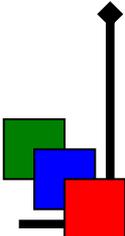
Básicamente el motivo de este resultado es que si bien las fluctuaciones de la autocovarianza disminuyen con $1/\sqrt{N}$, la sumatoria de la transformada de Fourier, agrega a las fluctuaciones del peridiograma un nuevo factor $\sqrt{N - 1}$, que impide que las fluctuaciones del peridiograma disminuyan aunque N aumente. Cuando aumentamos la longitud del registro, aumenta la resolución espectral pero la resolución estadística permanece constante.



Resolución Estadística a Expensas de Resolución Espectral:

Dado una secuencia aleatoria de longitud N , una forma de aumentar la resolución estadística es partiendo la secuencia en M subsecuencias de longitud N/M , calcular el peridiograma de cada una de ellas y promediarlos. Esta es una forma de simular promedios estadísticos. Podemos hacer esto porque estamos haciendo la hipótesis de que el proceso es ergódico, es decir, que los promedios estadísticos son iguales a los promedios temporales. Claro que al acortar la longitud de las subsecuencias a N/M ganamos resolución estadística a expensas de resolución en frecuencia, es decir que estamos tomando una decisión de compromiso entre una y otra resolución.

El problema con el que nos enfrentamos es encontrar la mejor forma de estimar la densidad del espectro de potencia de una secuencia infinitamente larga de ruido aleatorio cuando solo disponemos de un segmento de dicha secuencia. Los procesos aleatorios son procesos de energía infinita que no poseen espectros de potencia pero, por el teorema de Wiener-Khintchine, si es posible estimar la densidad de su espectro de potencia. En todos los casos algún tipo de hipótesis tendremos que hacer con respecto a los datos no registrados.



Proceso Aleatorio MA

Método de Blackman and Tukey:

Quando estimamos el espectro de potencia de un proceso aleatorio, ¿existe una mejor hipótesis que suponer que los datos son nulos fuera de la ventana de observación? Para responder esta pregunta vamos a suponer que la secuencia en cuestión x_n de longitud N_x es un proceso aleatorio estacionario que se genera excitando con ruido blanco ε_n un filtro MA a_n de orden N . El ruido blanco es un concepto abstracto que consiste en una señal aleatoria infinita cuyo espectro es constante para todas las frecuencias y su fase debe variar aleatoriamente para producir la componente innovativa. La salida de nuestro filtro MA estará dada por:

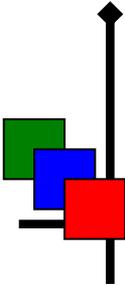
$$x_n = \varepsilon_n * a_n$$

$$X(Z) = E(Z)A(Z)$$

Es decir que el espectro de potencia de este proceso MA estará dado por:

$$|X(\omega)|^2 = (N_x - N)\sigma_\varepsilon^2 |A(\omega)|^2 \propto |A(\omega)|^2$$

El problema de estimación del espectro de potencia se reduce a determinar los coeficientes del filtro MA y calcular la densidad del espectro de potencia del mismo.



Proceso Aleatorio MA

Método de Blackman and Tukey:

Si nuestros datos x_n fueron generados por un proceso MA, entonces no podemos suponer que valen cero fuera de la ventana de observación como lo hace el peridiograma. Sin embargo como x_n es generado por un proceso MA su correlación será distinta de cero únicamente para los primeros N lags, a partir de este lag los datos serán ruido aleatorio no correlacionado. Es decir que sabemos que el dato en sí mismo es distinto de cero pero que la autocorrelación tiene que ser igual a cero para un lag mayor que N . Esta es una suposición menos estricta que la que hace el peridiograma.

La hipótesis es que los lags mayores que N son iguales a cero, pero que cuando los estimamos a partir de la ventana de datos disponibles, erróneamente obtenemos valores distintos de cero como consecuencia de no haberlos observado en una ventana infinitamente larga. Este es el método de estimación espectral de Blackman and Tukey.

La confiabilidad de los estimadores de la autocorrelación disminuye a medida que aumenta el lag de autocorrelación debido a la disminución de la zona de solapamiento con datos disponibles, esto es consecuencia de no haber observado los datos en una ventana infinitamente larga.



Proceso Aleatorio MA

Método de Blackman and Tukey:

El hecho de haber observado un dato de longitud infinita en una ventana de longitud N_x , nos conduce a la conclusión de que debemos utilizar ventaneos o funciones de peso que le den más peso a los *lags* centrales del estimador de la autocovarianza, con lo cual el estimador de la densidad del espectro de potencia nos quedará:

$$\hat{P}(\omega) = \sum_{k=-(N-1)}^{N-1} \hat{p}_k w_k e^{-i\omega k}$$

Donde la ventana w_k será una ventana de longitud $2N - 1$. Cuando esta ventana es una ventana rectangular tendremos el caso del proceso MA de orden N con el cual presentamos el método. Distintas ventanas como las de Bartlett o Welch, pueden ser utilizadas para implementar el método de Blackman and Tukey, sin embargo la elección de la ventana raramente está relacionada con alguna propiedad del dato o del proceso que se está analizando. Además la elección de diferentes órdenes N del proceso MA produce estimadores radicalmente distintos del espectro de potencia. Por lo general no se poseen criterios adecuados que permitan elegir de manera apropiada el orden del proceso.



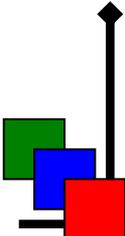
Proceso Aleatorio MA

Método de Blackman and Tukey:

La utilización de ventanas aumenta la resolución estadística (es decir reduce el *bias* o sesgo, y la varianza del estimador del espectro de potencia) a expensas de la resolución en frecuencias. Multiplicar a la auto-correlación en tiempo por la ventana es equivalente a convolucionar el peridiograma con la transformada de Fourier de la ventana. Lo cual para toda ventana razonable, implica una suavización del peridiograma reduciendo las fluctuaciones estadísticas.

También es posible efectuar el promedio de subsecuencias más cortas como discutimos anteriormente y combinar distintos tipos de ventaneo con el promedio de diferentes segmentos, generando una gran cantidad de estimadores distintos. También el dato puede ser segmentado con distintos ventaneos, podemos calcular los peridiogramas de los diferentes segmentos y luego promediar los peridiogramas. Incluso los distintos segmentos pueden solaparse o no.

La elección de las ventanas, de los segmentos y el orden del proceso MA, sin que exista un criterio basado en algún conocimiento que podamos tener del proceso en cuestión, pone en evidencia cuan empíricos pueden ser estos métodos de estimación espectral.



Proceso Aleatorio AR

Método de Yule-Walker:

En este caso suponemos que la señal en cuestión x_n , a la que le queremos estimar el espectro de potencia, fue generada excitando con ruido blanco ε_n un proceso AR de orden M :

$$X(Z) = \frac{E(Z)}{B(Z)}$$

O lo que es equivalente:

$$X(Z)B(Z) = E(Z)$$

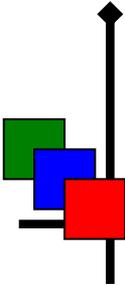
$$x_n * b_n = \varepsilon_n$$

Es decir que el espectro de potencia de este proceso AR está dado por:

$$|X(\omega)|^2 = \frac{(N_x + M)\sigma_\varepsilon^2}{|B(\omega)|^2} \propto \frac{1}{|B(\omega)|^2}$$

La idea es resolver la ecuación $x_n * b_n = \varepsilon_n$ para obtener b_n , óptimo en el sentido de los cuadrados mínimos, este operador está jugando el papel de un filtro que predice el error ε_n :

$$b_n = \Phi_{xx}^{-1} \begin{pmatrix} (N_x + M)\sigma_\varepsilon^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$



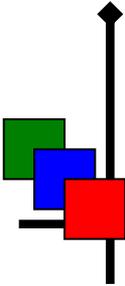
Proceso Aleatorio AR

Método de Yule-Walker:

Luego podremos estimar el espectro de potencia de nuestra señal como la inversa de la densidad del espectro de potencia del filtro AR salvo factor de escala.

Para que el método funcione debemos considerar un modelo autoregresivo en el cual el orden M del proceso sea mucho menor que la longitud N_x del dato.

Este método requiere de un estimador de la autocorrelación para poder calcular los coeficientes b_n del filtro autoregresivo. Obtendremos distintas soluciones para el modelo autoregresivo según utilicemos distintos estimadores de la autocorrelación. También deberemos elegir el orden del proceso AR, el cual como en el caso del modelo MA, generalmente no posee relación directa con lo que podamos llegar a saber sobre el proceso que genera la secuencia en cuestión.

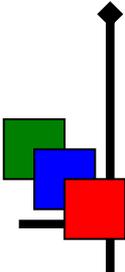


Proceso Aleatorio AR

Método de Yule-Walker:

Como es de esperarse la estructura de solo polos del proceso AR, en general produce más detalles o picos en el estimador del espectro de potencia obtenido, en comparación con el estimador obtenido modelando la secuencia como un proceso MA.

El modelo AR extiende la función de autocorrelación del dato aún más allá (hasta infinito) de lo que podría llegar a estimarse utilizando la ventana de observación. Por lo cual podríamos decir que mientras que el modelo MA hace menos suposiciones sobre el comportamiento del dato fuera de la ventana de registración que las que hace el peridiograma, el modelo AR hace aún menos suposiciones que el modelo MA.



Teoría de la Información de Shannon

Entropía informática:

La Teoría de la Información fue propuesta por Claude E. Shannon y Warren Weaver a finales de la década de 1940. Es una rama de la probabilidad y de la estadística que estudia todo lo relacionado con la información, en particular con los métodos de compresión y transmisión, sin que se produzca pérdida o deterioro de la misma.

Una de sus ramas es la criptografía que estudia la forma de codificarla.

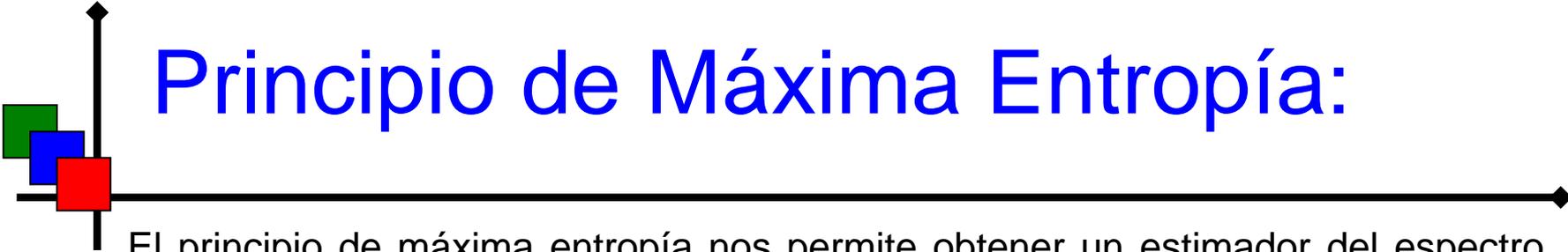
Se ocupa además, de cuantificarla, de analizar los diferentes modos de representarla, como así también del estudio de la capacidad de los sistemas de comunicación para transmitirla y procesarla.

El concepto básico de entropía en Teoría de la Información está vinculado a la incertidumbre que existe en cualquier experimento o señal aleatoria. Es una medida del desorden que contiene un sistema. El contenido de la información de un suceso es una función decreciente de la probabilidad de su aparición.

Una forma de cuantificar la aleatoriedad de un suceso, es mediante la entropía informática. Dada una variable aleatoria X con una función densidad de probabilidades $p_X(x)$ se define su entropía informática mediante la siguiente expresión:

$$H(X) = - \int_{-\infty}^{+\infty} p_X(x) \log[p_X(x)] dx$$

Si para representar la información se usaran valores en base \mathbf{b} entonces es conveniente utilizar el logaritmo en base \mathbf{b} . Si la información se va a representar mediante código binario se usa el logaritmo en base 2.



Principio de Máxima Entropía:

El principio de máxima entropía nos permite obtener un estimador del espectro de potencia correspondiente a la secuencia estacionaria infinita más aleatoria y menos predecible posible y que aún sea consistente con los datos observados. La entropía es una medida del desorden o de la no predictibilidad. De acuerdo con la teoría de información de Shannon la entropía de una serie de tiempo es proporcional a la integral del logaritmo de su densidad de espectro de potencia. Es decir que debemos maximizar la integral:

$$\int_{-\pi}^{+\pi} \log[P(\omega)] d\omega = \text{máximo}$$

Sujeta a las siguientes condiciones:

$$\int_{-\pi}^{+\pi} P(\omega) e^{i\omega k} d\omega = p_k$$

Resolviendo este problema de maximización con condiciones, aplicando cálculo de variaciones y multiplicadores de Lagrange, se llega a que el estimador de la densidad del espectro de potencia que satisface al principio de máxima entropía está dado por:

$$P(\omega) = \frac{K}{|B(\omega)|^2}$$



Principio de Máxima Entropía:

Donde K es una constante y $|B(\omega)|^2$ es el espectro de potencia de un proceso autoregresivo de orden M , cuyos coeficientes están dados por:

$$b_n = \Phi_{xx}^{-1} \begin{pmatrix} K \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Es decir que si los datos observados satisfacen el principio de máxima entropía, podrán ser representados como un proceso autoregresivo de orden M . Todavía resta el problema de cómo estimar los coeficientes del proceso AR. Una de las formas de hacerlo es utilizando el método de Yule-Walker que ya vimos y al cual se llegó por un camino diferente. La desventaja de este método es que estima los coeficientes de autocorrelación (o los de autocovarianza según como lo planteamos) a partir de la ventana de observación, los cuales como vimos se hacen menos confiables a medida que aumenta el lag de autocorrelación. La otra forma de hacerlo es utilizando el método de Burg el cual estima los coeficientes b_n del proceso autoregresivo directamente a partir de los datos de la ventana de observación y calcula simultáneamente los estimadores de la autocorrelación.



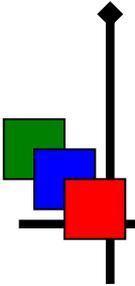
Principio de Máxima Entropía:

Observación: fíjese que si el espectro de potencia se hace cero en una banda de frecuencia, la integral que define a la entropía diverge dando una entropía infinitamente negativa. Según la teoría de la información de Shannon la interpretación de esto debe ser que tales series de tiempo no poseen aleatoriedad y por lo tanto son totalmente predecibles.



El Método de Burg:

El problema de los métodos de estimación espectral que hemos visto hasta ahora, es la poca confiabilidad de los estimadores de los coeficientes de autocorrelación que se calculan a partir de una ventana de observación de longitud finita. Este problema se agrava de sobremanera si la ventana de observación es corta. En 1967 John P. Burg presentó un paper en la 37° Convención de la Sociedad de Geofísicos de Exploración (SEG), en la ciudad de Oklahoma, que tuvo un impacto revolucionario en el análisis espectral. Burg mostró como estimar la autocorrelación evitando los efectos de la ventana de observación de longitud finita y sin hacer ninguna hipótesis sobre el dato fuera de la ventana de observación. El procedimiento de Burg es recursivo, comienza diseñando un filtro que predice el error de longitud 2, que intenta convertir el dato de entrada en ruido blanco. Luego iterativamente va calculando la autocorrelación a *lags* mayores simultáneamente con filtros predictores más largos. Mediante esta recursión es posible extender el filtro predictivo hasta alcanzar un orden M . Finalmente el espectro de potencia del dato es estimado como la inversa del espectro de potencia del filtro que predice el error.



El Método de Burg

Primera iteración:

$$\begin{pmatrix} \phi_0 & \phi_1 \\ \phi_1 & \phi_0 \end{pmatrix} \begin{pmatrix} 1 \\ f \end{pmatrix} = \begin{pmatrix} K \\ 0 \end{pmatrix}$$

$$\phi_0 = \sum_{i=0}^{N-1} x_i^2$$

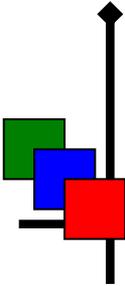
$$\phi_1 = -\phi_0 f$$

$$K = \phi_0(1 - f^2)$$

$$E_{RMS} = \frac{1}{N-1} \sum_m [(x_n * (1, f))_m]^2 + \frac{1}{N-1} \sum_m [(x_{-n} * (1, f))_m]^2$$

$$\frac{\partial E_{RMS}}{\partial f} = \text{mínimo}$$

$$f = \frac{2 \sum_{i=0}^{N-2} x_{i+1} x_i}{\sum_{i=0}^{N-1} x_i^2 + \sum_{i=0}^{N-2} x_{i+1}^2}$$



Bibliografía:

- Karl, John H. (1989), An introduction to Digital Signal Processing, Academic Press. Chapter Eleven.
- Oppenheim, A.V. and Schafer, R. (1975), Digital Signal Processing, Prentice-Hall.