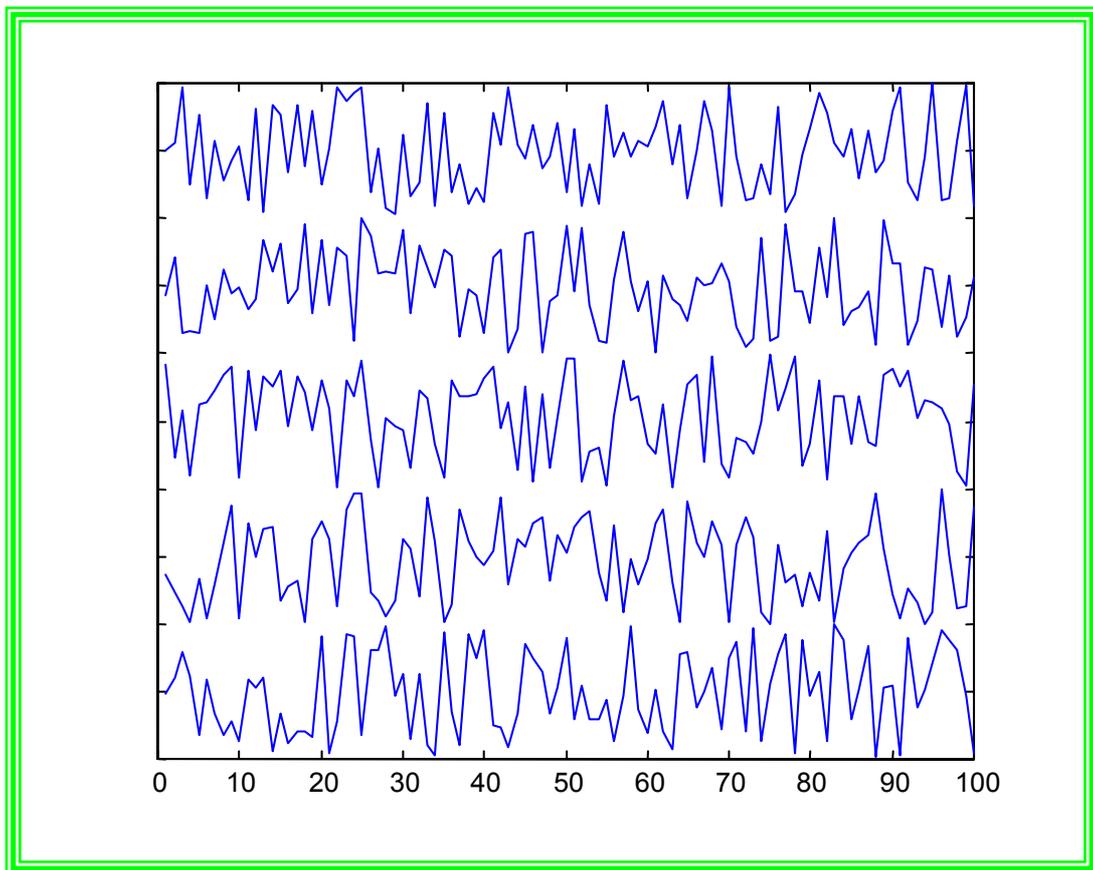


## Capitulo II

# PROCESOS ESTOCASTICOS Y ESTIMACION DE PARAMETROS



- II.1. VARIABLES ALEATORIAS
- II.2. PROCESOS ESTOCASTICOS
- II.3. ESTACIONARIDAD Y ERGODICIDAD
- II.4. PROCESOS ESTOCASTICOS Y SISTEMAS LINEALES
- II.5. PROBABILIDAD, ENTROPIA Y ESTIMACION
- II.6. ESTIMACION MAP Y ML DE LA MEDIA DE UN PROCESO
- II.7. LA CALIDAD DE UN ESTIMADOR
- II.8. ALISADO DE DATOS
- II.9. ESTIMACIÓN DE LA FUNCIÓN DE CORRELACIÓN
- II.10. MODELOS RACIONALES
- II.11. CONCLUSIONES
- II.12. EJERCICIOS
- II.13. BIBLIOGRAFIA
- APENDICE. LIMITE DE CRAMER-RAO

## II.1 VARIABLES ALEATORIAS

Normalmente en ingeniería, ante la observación de un fenómeno físico mediante una medida  $x$ , se caracteriza el desconocimiento en su generación o la imprecisión en su observación a través de la probabilidad de que un determinado valor sea el observado. En el fondo, la idea de probabilidad se asocia incorrectamente con la incapacidad de determinar con exactitud futuras observaciones. Nótese que es, en un principio, incorrecto representar del mismo modo imprecisión y aleatoriedad. El primer concepto va ligado a la problemática del sistema de medida y es ajeno a lo medido, mientras que el segundo es una característica intrínseca al proceso de generación del fenómeno observado. No obstante, la diferencia entre imprecisión y aleatoriedad no se va a reflejar en nuestras herramientas de proceso de señales y así, siempre que surja duda sobre nuestra capacidad de predecir se dirá que se esta ante una variable aleatoria.

Después de lo comentado, parece que la mejor manera de poner una etiqueta a la mencionada incapacidad de predecir, es mostrar con qué frecuencia se observan unos y otros valores de la variable. Esta función de la variable  $x$  es en esencia discreta pues medir la frecuencia de aparición de un valor  $x_0$  concreto requerirá teóricamente de infinitas medidas. Dicho de otro modo, al tomar  $N$  valores de  $x$  se puede tener la desfortuna de que el valor  $x_0$  no aparezca nunca y que si que aparezcan alguna vez valores comprendidos entre  $x_0+\varepsilon$  y  $x_0-\varepsilon$ , pero no exactamente en  $x_0$ . Es evidente, que tratar de medir la frecuencia de aparición de un valor concreto no es recomendable y es más seguro, y tiene más sentido, abordar frecuencias de aparición dentro de un intervalo  $\Delta$ . De este modo, el denominado histograma de la variable aleatoria  $x$ ,  $H(x)$ , puede obtenerse dividiendo el margen dinámico de  $x$  en intervalos de amplitud  $\Delta$ , alrededor de valores específicos de  $x$ . Sobre  $N$  medidas, se anotará el número de medidas que caen dentro del intervalo. La función se normaliza dividiendo el número de valores en cada intervalo por el número total de medidas  $N$ .

Para comprender mejor la forma de evaluar el histograma, nótese que en el intervalo  $x_0-\Delta/2$ ,  $x_0+\Delta/2$  el número de medidas encontradas es  $n$ , entonces el histograma sería (II.1),

$$H(x_0) = \frac{n}{N} = \frac{1}{N} \sum_{i=0}^{N-1} \int \delta(z - Q(x_i)) dz$$

$$H(x) = \sum_{q=-Q}^Q H(x_q) \cdot \frac{1}{\Delta} \Pi\left(\frac{x - x_0}{\Delta}\right) \quad (\text{II.1})$$

donde la función  $Q(\cdot)$  cuantifica el valor de su argumento en intervalos de tamaño  $\Delta$  y la función de interpolación se elige de forma rectangular. Así pues, un histograma tendrá siempre la apariencia de diferentes rectángulos superpuestos, tal y como se indica en la Figura II.1, donde el ancho de sus bases indica la precisión empleada o amplitud de los intervalos usados para su cálculo.

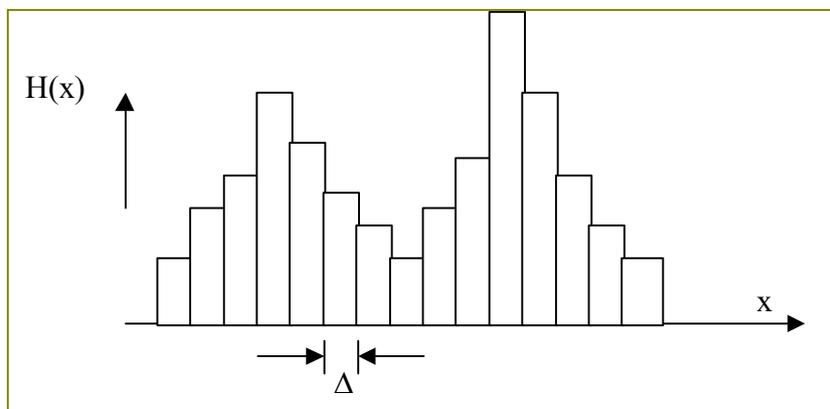


Figura II.1. Histograma de la variable aleatoria  $x$

De lo anterior se pueden intuir varias cuestiones relativas a la evaluación del histograma. Es claro que un histograma de alta resolución conlleva hacer lo más pequeño posible la anchura del intervalo. El problema es que, si el número de medidas no es grande, se pueden dejar 'vacíos' rectángulos

o convertir el histograma en rectángulos que como máximo capturan un valor. Así pues, si el número de medidas disponible es finito, solo se puede estrechar el intervalo en aquellos rectángulos más llenos. También se puede elaborar, mera cosmética, una curva continua tomando los valores medidos y usando otras funciones de interpolación, en lugar de rectángulos. Si se usan triángulos, en lugar de rectángulos, el histograma sería un conjunto de segmentos rectos unidos entre si. Existen otras posibilidades, todas ellas más o menos afortunadas en dar una imagen de curva continua al histograma básico, que se esquematiza en la figura anterior.

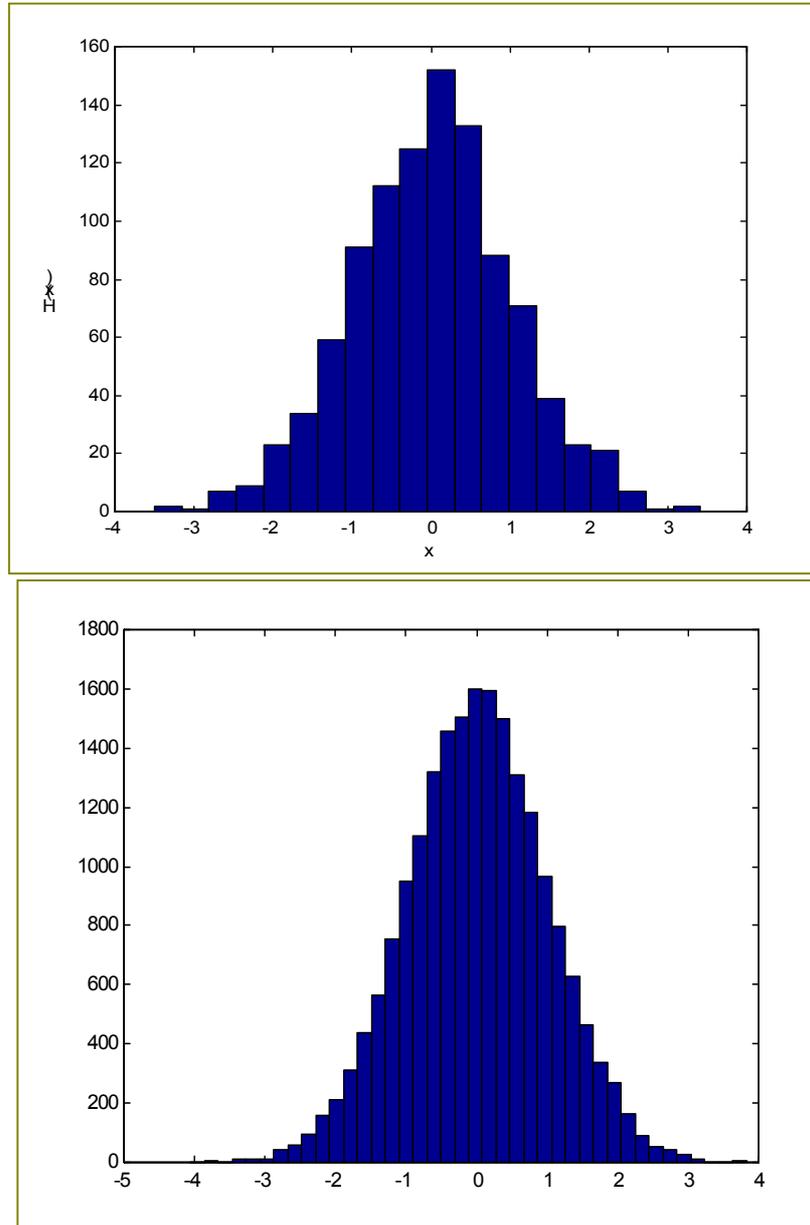


Figura II.2. (Arriba) Histograma calculado con 20 intervalos y obtenido a partir de 1000 valores de una variable aleatoria  $x$  gaussiana de media nula y varianza unidad. (Abajo) Igual pero con 40 intervalos y 20000 valores.

En el caso de que el número de medidas tienda a infinito, no hay gran problema en hacer tender el intervalo a cero y el histograma aparecería como una curva continua. Dicha curva continua, imposible de medir en la práctica, es lo que se denomina función de densidad de probabilidad de  $x$ , o abreviadamente pdf (“probability density function”) de  $x$ ,  $f(x)$ .

$$\lim_{\substack{N \rightarrow \infty \\ \Delta \rightarrow 0}} H(x) = f(x) \quad (\text{II.2a})$$

La interpretación de esta función es la siguiente: la probabilidad de que el resultado de una medida de una variable aleatoria  $x$  esté en el intervalo  $[x_0, x_1]$  es igual a la integral de esa función en dicho intervalo:

$$\Pr(x_0 \leq x \leq x_1) = \int_{x_0}^{x_1} f(x) dx \quad (\text{II.2b})$$

De esta forma, puede definirse la función de distribución  $F(x)$  como la integral de la pdf:

$$F(x) = \Pr(x \leq x_1) = \int_{-\infty}^{x_1} f(x) dx \quad (\text{II.2c})$$

Nótese que, por lo expuesto, el histograma puede considerarse un estimador de la probabilidad de que una variable aleatoria tome valores en intervalos de tamaño  $\Delta$ . Otra manera de calcular el histograma, en esta ocasión con forma de curva continua, es vía la denominada función característica. La función característica de una variable aleatoria se define como la transformada de Fourier de la función de densidad de probabilidad:

$$\Phi(\omega) = \int_{-\infty}^{\infty} \exp(-j\omega x) f(x) dx \quad (\text{II.3})$$

Como quiera que cada vez que se calcula el producto escalar de una función  $g(x)$  con la probabilidad de  $x$  lo definimos como el valor esperado  $E\{ \cdot \}$  de  $g(x)$ , la función característica puede definirse como el valor esperado de  $\exp(-j\omega x)$ . De lo anterior se puede concluir que un estimador de la función característica de  $x$ , usando valores  $x_i$  para  $i=1, \dots, N$  vendría dado por la ecuación (II.4). Al mismo tiempo, la transformada inversa de Fourier sería el estimador de la pdf.

$$\hat{\Phi}(\omega) = \frac{1}{N} \sum_{i=1}^N \exp(-j\omega x_i) \quad (\text{II.4})$$

Al margen de herramientas más completas para comprobar la calidad del estimador y que se expondrán más adelante, puede comprobarse que, al menos, el valor esperado del estimador es la función característica exacta. El estimador de la probabilidad o histograma se obtendría, muestreando la función característica con  $M < N$  puntos y tomando transformada discreta de Fourier inversa de la función resultante.

Como se ha podido ver, la caracterización de una v.a. (variable aleatoria) consiste en conocer su probabilidad con la mayor precisión posible. Pues bien, aun contando con que se disponga de  $f(x)$ , arrastrar constantemente la función en procesamiento de señal, ya sea para diseño o análisis de sistemas, sería engorroso y complejo. Hacer práctico el manejo de la pdf implica representarla paramétricamente, es decir, mediante un número pequeño de parámetros que la representen lo más correctamente posible. La aproximación utilizada, aparentemente burda o grosera y que, sin embargo, en la mayor parte de las ocasiones, es más que suficiente para solucionar diseños, preserva tan solo dos parámetros: la media o centro de gravedad y la varianza o momento de inercia por utilizar el símil de masa y parámetros mecánicos para la pdf.

$$\begin{aligned} \mu_x &= E\{x\} = \int x f(x) dx \\ \sigma_x^2 &= E\{x - \mu_x\}^2 = E\{x^2\} - \mu_x^2 = \int |x - \mu_x|^2 f(x) dx \end{aligned} \quad (\text{II.5})$$

Es fácil, alegar que dadas media y varianza hay muchas distribuciones diferentes que las comparten y es cierto. ¿Cuál es entonces la razón del éxito de esta reducción tan simplista de la información a conservar? En primer lugar, bajo el punto de vista de ingeniería, siempre se maneja potencia y relaciones de potencia (SNR= Relación señal a ruido) con lo que varianza, cuando la media es nula, al ser la potencia o energía de la variable aleatoria suele ser suficiente para los fines del procesamiento o diseño que se llevan a cabo. En

segundo lugar, y más importante, la mayor parte de las variables manejadas en comunicaciones son gaussianas y, en este caso, la información retenida es toda la necesaria para conocer  $f(x)$ . Además, cuando la variable aleatoria no es gaussiana en origen, el sistema de medida, habitualmente lineal con memoria, la convierte en gaussiana debido al teorema central del limite.

La pdf para una variable gaussiana viene dada por:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left(-\frac{|x - \mu_x|^2}{2\sigma_x^2}\right) \tag{II.6}$$

En el caso de tener un vector  $\underline{X}$  de componentes complejas, de dimensión N con una distribución gaussiana:

$$f(\underline{X}) = \frac{1}{\pi^N \cdot \det(\underline{C})} \exp\left[-(\underline{X} - \underline{\mu}_x)^H \underline{C}^{-1} (\underline{X} - \underline{\mu}_x)\right] \tag{II.7}$$

siendo

$$\begin{aligned} \underline{\mu}_x &= E\{\underline{X}\} \\ \underline{C} &= E\{(\underline{X} - \underline{\mu}_x)(\underline{X} - \underline{\mu}_x)^H\} = E\{\underline{X}\underline{X}^H\} - \underline{\mu}_x \underline{\mu}_x^H \\ \det(\underline{C}) &= \text{determinante}(\underline{C}) \end{aligned} \tag{II.8}$$

Nótese que, en este caso, efectivamente la media y la varianza (o la correlación) caracterizan completamente la variable aleatoria.

Finalmente, es de destacar que mientras los sistemas lineales cambian la distribución de potencia o densidad espectral, son los sistemas no-lineales los que cambian la densidad de probabilidad (véase figura II.3) Si una variable x de pdf  $f(x)$  pasa por una linealidad sin memoria y biunívoca  $g(x)$  entonces la probabilidad de la salida viene dada por (II.9).

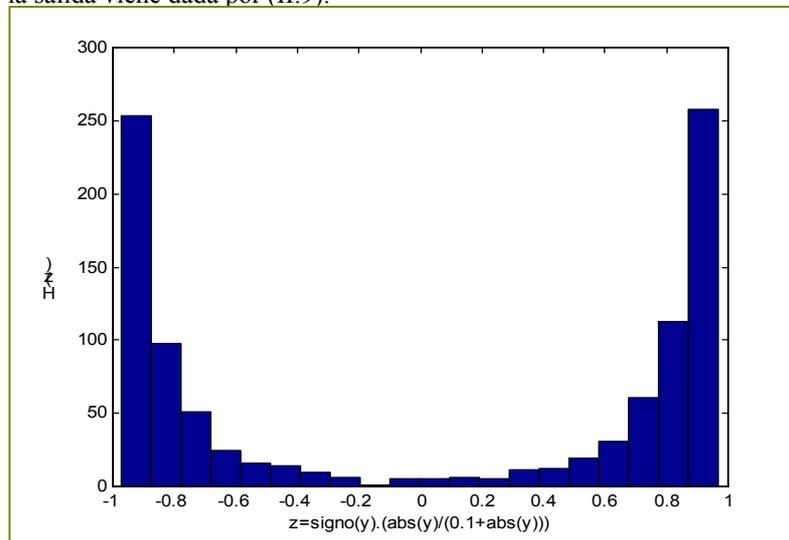


Figura II.3. Histograma de la v.a. z resultado de someter una v.a. gaussiana y a la no linealidad indicada en la parte inferior de la figura.

$$y = g(x) \qquad f_y(y) = \frac{f_x(x)}{\left| \frac{\partial g}{\partial x} \right|_{x=g^{-1}(y)}} \tag{II.9}$$

De esta propiedad se sugiere la manera de generar, a partir de una distribución uniforme otras distribuciones de probabilidad.

## II.2 PROCESOS ESTOCASTICOS

En el apartado anterior se ha expuesto la manera de caracterizar una v.a. vía la función de probabilidad o con dos de sus momentos, media y varianza. No obstante, en la mayor parte de las situaciones y aplicaciones, como en comunicaciones, la mayor parte de esas variables (tensión o corriente en dispositivos o sistemas) tienen una evolución temporal. En otras palabras, lo que en el apartado anterior aparecía como conjunto de medidas no son más que muestras, en general sucesivas, de una señal continua. En definitiva, medida de manera continua o muestreada, las medidas se corresponden con una evolución temporal. Así pues el fenómeno físico se nos manifiesta como una señal  $x(t)$  sobre la que no se es capaz de predecir exactamente a lo largo del tiempo.

No obstante, el problema no es su evolución temporal, sino que está ligado al verdadero carácter aleatorio del fenómeno físico observado. Un ejemplo evidente es la siguiente: si se diseñan varios receptores de FM y se visualiza el ruido que presentan cuando están desintonizados, se observa que, aunque contruidos bajo el mismo diseño y tipo de dispositivos de principio a fin, las señales recogidas son diferentes en su forma de onda. Es más, cuantos más equipos diferentes se visualizan más formas de onda diferentes se observarían. Por lo tanto, calificar el ruido de un receptor como una señal sería incorrecto y es más adecuado denominarle proceso ya que se manifiesta con infinitas realizaciones o formas de onda diferentes, cada una de ellas asociada a un experimento, que en el caso reseñado es una medida sobre distintos receptores. A este conjunto de señales obtenidas como resultado de distintos experimentos (cada uno de los receptores) se le denominara proceso aleatorio y se representara por  $\{x\}$ .

Al representar un proceso, tal y como indica la Figura II.4, puede verse que se trata de un conjunto de infinitas funciones, de longitud también infinita en el tiempo. La Figura II.5 muestra cinco realizaciones de, tan solo, 100 muestras de un proceso aleatorio, en el que cada una de las variables aleatorias está distribuida uniformemente. Es evidente que el primer problema que se plantea es como representar un proceso, o bien, la manera de comprimir la información de esta representación en unos pocos parámetros o funciones.

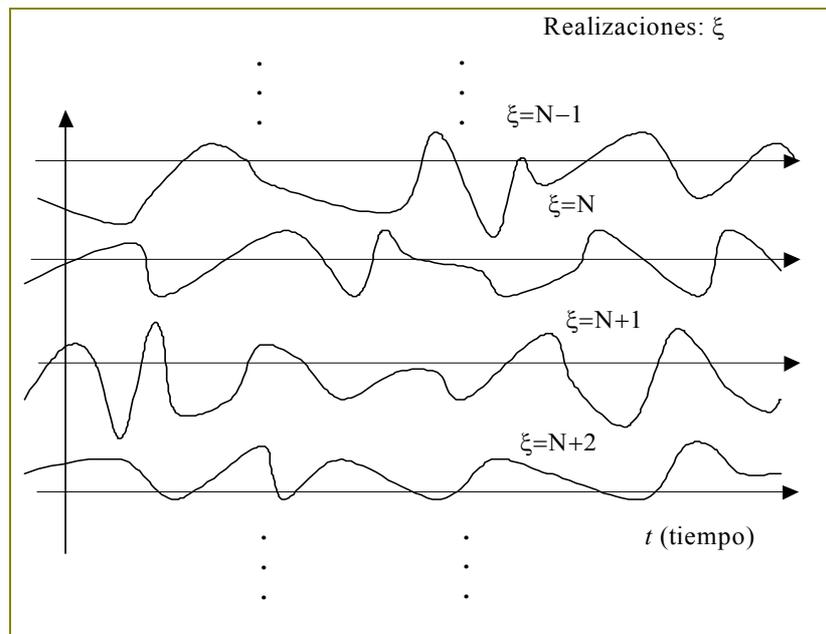


Figura II.4. Proceso aleatorio representado por sus realizaciones, resultados de distintos experimentos. Cada una de ellas lleva asociada una forma de onda en el tiempo.

La caracterización simplificada comienza, en primer lugar, por dejar fija una de las dimensiones y caracterizar en la otra. Comenzando por la caracterización cuando se ha fijado la realización  $\xi$ , se esta

ante una forma de onda de potencia media finita. Así pues, la información a reseñar a realización fija, tal y como se realiza para señales determinísticas, se resume en una sola función y dos parámetros: la autocorrelación, la potencia como autocorrelación en el origen, y la media, habitualmente nula, respectivamente.

$$\begin{aligned}
 m_x(\xi) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_T x(t, \xi) dt \\
 R_{xx}(\tau, \xi) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_T x^*(t, \xi) x(t + \tau, \xi) dt \\
 P_x(\xi) &= R_{xx}(0, \xi)
 \end{aligned}
 \tag{II.10}$$

Es habitual que los valores anteriores, siempre que T sea suficientemente grande, no dependan de la realización particular  $\xi$ .

Pasando a la representación a tiempo fijo, si se representa el proceso para un  $t_0$  fijo, se estará ante un conjunto infinito de valores (uno como resultado de cada uno de los experimentos) que reproducen el problema de representar una v.a. como ya ocurrió en el apartado anterior. Así pues, la caracterización fijando un instante de tiempo es la correspondiente a una v.a. con pdf  $f(x, t)$  a partir de la cual es posible determinar o representar por su media y su varianza:

$$\begin{aligned}
 \mu_x(t) &= \int_{-\infty}^{+\infty} x f(x; t) dx \\
 \sigma_x^2(t) &= \int_{-\infty}^{+\infty} (x - \mu_x(t))^2 f(x; t) dx
 \end{aligned}
 \tag{II.11}$$

Estos parámetros servirían para la representación de, digamos la imagen en dos dimensiones que es un proceso estocástico. Es obvio que la caracterización del proceso es todavía escasa al no disponerse de una medida de la evolución en los dos ejes, es decir, una medida cruzada en las dos dimensiones del proceso.

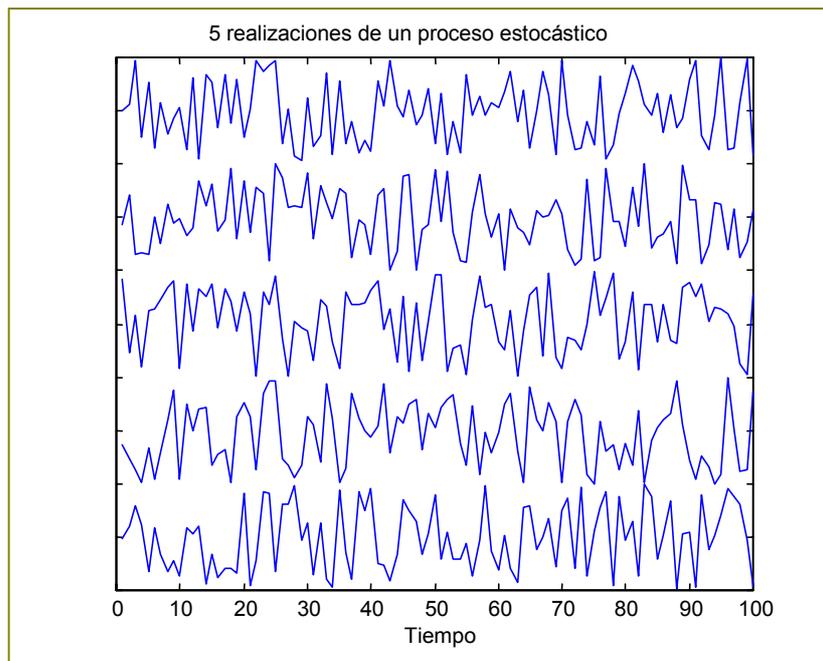


Figura II.5 Cinco realizaciones de un proceso. El corte a  $t$  fijo produce una v.a. de distribución uniforme.

Como se ha comentado la única manera de representar fenómenos aleatorios es usando sus funciones de densidad de probabilidad. De este modo, cuando fijado el tiempo se ha encontrado una v.a. se ha recurrido a representar la  $f(x,t)$  de la v.a. Es importante reseñar que en el caso de un proceso es la pdf de las variables aleatorias que lo componen la que le caracteriza. Así pues, el interés es ahora el caracterizar que ocurre con los valores que toma el proceso cuando se desplaza a lo largo de su evolución, no solo en realización sino también en tiempo. La manera mas sencilla, y a la vez grosera, de medir dicha evolución es considerar el caso de dos cortes a  $t$  fijo en instantes diferentes de tiempo. Nótese que la medida entraña establecer una relación entre dos variables aleatorias, y su caracterización requerirá el estimar o medir la función de densidad de probabilidad conjunta:

$$f(x_1, x_2; t_1, t_2) \quad (\text{II.13.a})$$

La interpretación de esta distribución conjunta es, como en la ecuación (II.2b), la siguiente: La integral de la función nos da la probabilidad de que el proceso en el instante  $t_1$  tome valores en el intervalo  $[x_a, x_b]$  y que, conjuntamente, en el instante  $t_2$  tome valores en el intervalo  $[x_c, x_d]$ :

$$\Pr(x_a \leq x(t_1) \leq x_b; x_c \leq x(t_2) \leq x_d) = \int_{x_a}^{x_b} \int_{x_c}^{x_d} f(x_1, x_2; t_1, t_2) dx_1 dx_2 \quad (\text{II.13.b})$$

De nuevo, el problema es conseguir una representación de la función anterior que sea de fácil en cálculo, estimación y manejo, y que tenga una interpretación que la haga útil para el procesado de señal. Dejando de momento de lado este problema que, a la larga es el objetivo de la caracterización de orden dos de un proceso, se va a describir un aspecto que esta íntimamente relacionado con él.

Una manera de abordar la medida cruzada sería establecer el parecido o distancia entre dos cortes a  $t$  fijo, uno realizado en  $t$  y el otro en  $t+\tau$ , que también se puede expresar como cuánto se parece el futuro al presente del proceso. Nótese que esta medida de parecido se establece como la función que al aumentar disminuye la distancia euclidea entre las realizaciones, en los dos instantes de tiempo seleccionados. Como puede verse, se asume que la parte real de la función que se denominara como parecido es la que hace disminuir la distancia y que esta medida de distancia se hace una vez normalizadas las energías de la realizaciones en cada corte. El primer comentario se debe al sentido que tiene una señal analítica y debe recordarse que tan solo su parte real tiene sentido físico. El segundo comentario es natural ya que no tendría sentido comparar dos vectores, en distancia euclidea, si cualquiera de ellos es libre de cambiar su norma.

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{m=0}^{N-1} |x(t, \xi_n) - x(t + \tau, \xi_m)|^2 = \sigma_x^2(t) + \sigma_x^2(t + \tau) - 2 \operatorname{Re} \left( \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{m=0}^{N-1} x^*(t, \xi_n) x(t + \tau, \xi_m) \right)$$

Procediendo de este modo y por pasos, tal y como se indica en la Figura II.6, el valor de la realización  $\xi_n$  en  $t$  se compararía con los valores que toman todas las realizaciones en  $t+\tau$ :

$$\text{Parecido}(\xi_n, t, \tau) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{m=0}^{N-1} x^*(t, \xi_n) x(t + \tau, \xi_m) \quad (\text{II.13.c})$$

En esta última relación hay que tener presente el hecho de que pares de valores muy frecuentes contribuyen mucho a la medida de parecido (aparecen muchas veces en el sumatorio), pero, resulta intuitivamente incorrecto que un par de valores que no salen casi nunca contribuyan con la misma importancia que los que son muy frecuentes. Una manera de solucionar el problema mencionado, que respeta la necesidad de dar mas fiabilidad a los valores mas probables, es expresar la misma ecuación (II.13.c) introduciendo la pdf de la variable aleatoria  $x(t+\tau)$  como ponderación de la importancia (o frecuencia de aparición) de los valores del integrando:

$$\text{Parecido}(\xi_n, t, \tau) = \int x^*(t, \xi_n) x_2 f(x, t, x_2; t + \tau) dx_2 \quad \text{con} \quad x_2 = x(t + \tau) \quad (\text{II.13.d})$$

donde, debe recordarse que  $x(t, \xi_n)$  no es una variable a integrar sino que es el valor fijo que toma la realización  $\xi_n$  en el instante  $t$ . Por último, para llegar a la noción completa de parecido, la medida anterior ha de extenderse a todos los puntos del corte  $t$ , es decir, integrarse para todos los valores posibles de  $x(t)$ . La función así obtenida se denomina función de autocorrelación del proceso  $\{x\}$ :

$$r_x(t, \tau) = \lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} x^*(t, \xi_n) x(t + \tau, \xi_m) \cdot f(x(\xi_n), t, x(\xi_m), t + \tau) \quad (\text{II.14a})$$

o bien, siguiendo el razonamiento anterior, se puede utilizar una pdf que caracterice conjuntamente a las variables aleatorias  $x(t)$  y  $x(t+\tau)$ , que substituida en (II.14a), da lugar a la expresión más habitual:

$$r_x(t, \tau) = \iint x_1^* x_2 f(x_1, x_2; t, t + \tau) dx_1 dx_2 = E\{x^*(t)x(t + \tau)\} \quad (\text{II.14b})$$

Nótese que en (II.14b) se retiene únicamente la media de una función de dos dimensiones. Se podría continuar reteniendo la media de probabilidades conjuntas de ordenes mayores, a tres tiempos, cuatro, etc.; no obstante, para la inmensa mayoría de aplicaciones la caracterización a la que se ha llegado con (II.14b) es suficiente para trabajar con procesos aleatorios.

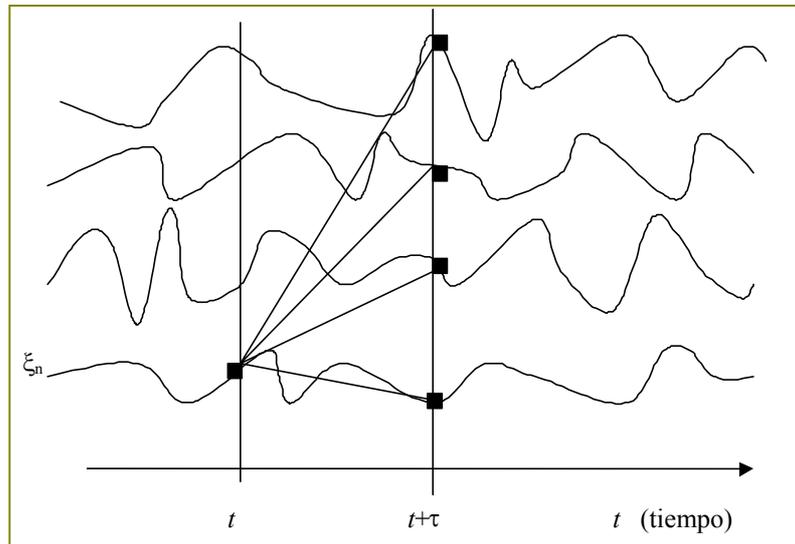


Figura II.6. Evaluación del parecido o distancia entre un valor en una realización y los valores futuros del proceso.

Antes de proseguir, se ha de mencionar que cuando las pdf mencionadas son gaussianas, se esta ante un proceso aleatorio gaussiano. Si el parecido de orden dos es nulo para cualquier desplazamiento  $\tau$ , el proceso será denominado blanco y su función de autocorrelación sería:

$$r_x(t, \tau) = N_o(t) \delta(t - \tau) \quad (\text{II.15})$$

La potencia del proceso vendrá dada por  $r_x(t,0)$  y podrá variar, como se indica, con el tiempo.

En cualquier caso es de destacar que, de una función de cuatro dimensiones  $f(x_1, x_2, t, t+\tau)$ , todo lo que se va retener es una función de dos dimensiones  $r_x(t, \tau)$ . Dicha función es básicamente el centro de gravedad, según  $x_1$  y  $x_2$ , de la función original. Aunque parezca sorprendente, la mayor parte de la ingeniería, y en especial comunicaciones, esta reducción de información es la que ha soportado todo su avance y desarrollo. Lo realmente importante es que aun existen mas simplicaciones, respaldadas por el uso afortunadamente, que convertirán el trabajo con procesos en una tarea no mucho mas compleja que si de señales determinísticas se tratase.

### II.3. ESTACIONARIEDAD Y ERGODICIDAD

Al observar las funciones que se han elegido para caracterizar un proceso, expuestas en el apartado anterior, se comprueba que todas dependen del tiempo. No obstante, la mayor parte de los fenómenos que dan lugar a procesos aleatorios mantienen sus propiedades estadísticas constantes con el tiempo, en cuyo caso se dice que son estacionarios. De este modo, si la media del proceso no depende del tiempo se dirá que el proceso es estacionario de orden uno, y si la correlación no depende del instante  $t$ , que es estacionario de orden 2. En definitiva, si se puede decir que el proceso es estacionario de orden dos, se verificara (II.16):

$$\begin{aligned} r_x(t, \tau) &= E\{x^*(t)x(t+\tau)\} = r_x(\tau) \quad \forall t \\ \sigma_x^2(t) &= \sigma_x^2 = r_x(0) \quad \forall t \end{aligned} \quad (II.16)$$

Es claro que la estacionaridad del momento de orden dos no implica la estacionaridad de la función de probabilidad conjunta.

Dado que si se verifica:

$$f(x_1, x_2; t, t+\tau) = f(x_1, x_2; \tau) \quad \forall t \quad (II.17)$$

se esta ante un criterio de estacionaridad más estricto, puesto que esta ultima implica por supuesto que su centro de gravedad también se mantiene constante con el tiempo. Cuando se verifique (II.17) se dirá que el proceso es estacionario de orden dos en sentido estricto, mientras que si tan solo se verifica (II.16) dirá que es estacionario, se suele sobrentender que de orden dos, en sentido laxo o amplio.

La estacionariedad en sentido amplio, al margen de ser de más fácil verificación, se cumple para la mayor parte de los procesos con los que se trabaja en procesado de señal. Es más, cuando la longitud de las señales con las que se trabajara sea de, digamos, NT segundos, tan solo se necesitara estacionaridad de la autocorrelación durante ese intervalo; es más, en términos prácticos, lo que será necesario es que la no estacionaridad o variabilidad no sea tan rápida como para invalidar el procesado llevado a cabo en el intervalo mencionado.

Para el caso de procesos no estacionarios, e imitando a las señales de potencia finita, se puede definir una densidad espectral de potencia variante con el tiempo y definida como la transformada de Fourier de  $r_x(t, \tau)$  en la segunda variable:

$$S_x(t, \omega) = \int r_x(t, \tau) \exp(-j\omega\tau) d\tau \quad (II.18)$$

La densidad espectral del proceso es la potencia que todas las realizaciones, en media, sitúan en un ancho de banda infinitesimal alrededor de cada frecuencia. Claramente se verifica que su área es la potencia del proceso o autocorrelación en  $\tau$  igual a cero. Cuando el proceso es estacionario lógicamente la distribución de potencia se mantendrá constante con el tiempo y será igual a (II.19).

$$S_x(\omega) = \int r_x(\tau) \exp(-j\omega\tau) d\tau \quad (II.19)$$

Además del parecido consigo mismo empleado para introducir la función de autocorrelación, se puede hablar de parecido entre dos procesos  $\{x\}$  e  $\{y\}$ , y de correlación cruzada de ambos a la función que lo caracteriza. Esta correlación cruzada se usara según su definición en (II.20) donde se incluye el caso de que uno o ambos procesos sean complejos.

$$r_{xy}(t, \tau) = E\{x^*(t)y(t+\tau)\} \quad (II.20)$$

Cuando este parecido entre ambos sea estacionario, la correlación cruzada dependerá solo de  $\tau$  y, por la misma razón, su transformada de Fourier será también una función que dependerá únicamente de la frecuencia.

$$S_{xy}(\omega) = \int r_{xy}(\tau) \exp(-j\omega\tau) d\tau \quad (\text{II.21})$$

Como puede concluir el lector, la noción de estacionaridad y su presencia en situaciones prácticas será una ayuda inestimable en el diseño y análisis de sistemas de procesado de señal. Así como, la estacionaridad es una cualidad bastante habitual en la práctica, cuando menos en intervalos finitos de tiempo, la propiedad que sigue, denominada ergodicidad, no lo es. Es más, es profundamente artificial y obedece más a una limitación que a una propiedad presente o fácil de comprobar en la práctica. El problema al que responde es la incapacidad de disponer de más de una realización del proceso en la mayor parte de las aplicaciones. La propiedad de ergodicidad explora este problema y brinda la segmentación como herramienta para emular diversas realizaciones del proceso bajo análisis.

Se dice que un proceso estacionario es ergódico cuando las funciones que entrañan valores esperados a lo largo de realizaciones pueden obtenerse también a partir de una sola realización. En otras palabras, ergodicidad de segundo orden conlleva que la autocorrelación del proceso obtenida en la ecuación II.14b coincide con la que aparece en la ecuación II.10:

$$r_x(\tau) = \iint x_1^* x_2 f(x_1, x_2; \tau) dx_1 dx_2 = R_{xx}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_T x^*(t) x(t + \tau) dt \quad (\text{II.22})$$

Aparentemente la propiedad de ergodicidad, planteada así, no parece de muy fácil comprobación. Una de las razones fundamentales por la que el principio de ergodicidad se aplica es la que sigue. Si se asume que el proceso es estacionario y se dispone de una realización que dura 2NT seg., la ergodicidad del proceso conlleva que es posible considerar que se disponen de dos realizaciones de duración NT segundos, que se obtienen de segmentar la realización única original (ver figura II.7). El proceso de división conlleva que, a mayor número de realizaciones, menor habrá de ser la duración de cada una debido a la segmentación. Otra manera de aumentar el número de realizaciones es segmentar la realización de 2NT seg. solapando los segmentos. El problema es que de esta forma se introducirá correlación artificialmente entre las supuestas realizaciones e invalidará el pretendido incremento en el número de estas. En resumen, en su sentido más general, y adecuado, la ergodicidad capacita para generar realizaciones a partir de una única, es decir, mover información del eje temporal al eje de realizaciones. Dicho de otra forma, se podrá intercambiar un promedio sobre todas las realizaciones (esperanza matemática) por un promedio usando una única realización (promedio temporal). Este concepto de ergodicidad será muy utilizado en futuros temas.

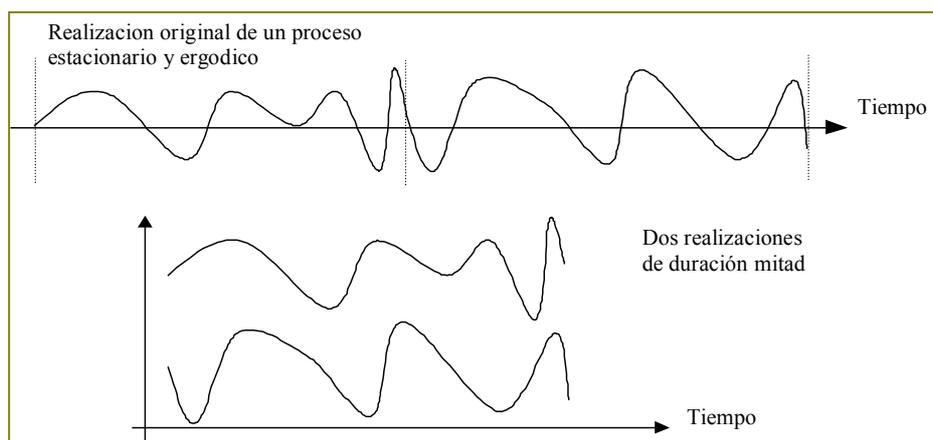


Figura II.7. El concepto de ergodicidad en un proceso estacionario. La información temporal puede pasarse a información en el eje de las realizaciones.

#### II.4. PROCESOS ESTOCASTICOS Y SISTEMAS LINEALES

Cuando las realizaciones de un proceso se aplican a un sistema lineal, a su salida se obtiene otro proceso estocástico. Las relaciones existentes entre las funciones que caracterizan el proceso de entrada  $\{x\}$  con el de salida  $\{y\}$  se obtienen vía la ecuación de convolucion en función de la respuesta impulsional del sistema lineal  $h(t)$ .

En primer lugar, la media del proceso de salida puede derivarse según se indica a continuación donde  $H(\omega)$  es la respuesta frecuencial del sistema lineal.

$$\begin{aligned} y(t) &= \int h(\zeta)x(t-\zeta)d\zeta \\ E\{y\} &= \int h(\zeta)E\{x(t-\zeta)\}d\zeta = \{si\ x\ es\ estacionario\} = \\ &= E\{x\} \int h(\zeta)d\zeta = E\{x\}H(0) \end{aligned} \quad (II.23)$$

Es interesante indagar el parecido que presenta la salida con la entrada, es decir, la correlación cruzada de los dos procesos y como afecta el sistema lineal al concepto de parecido o auto-parecido.

$$r_{xy}(t) = E\left\{x^*(\tau) \int h(\zeta)x(t+\tau-\zeta)d\zeta\right\} \quad (II.24)$$

Si el proceso  $\{x\}$  es estacionario, se obtendrá (II.25.a) y, en términos de espectro cruzado (II.25.b).

$$r_{xy}(t) = \int h(\zeta)r_x(t-\zeta)d\zeta = h(t) * r_x(t) \quad (II.25.a)$$

$$S_{xy}(\omega) = H(\omega)S_x(\omega) \quad (II.25.b)$$

Del mismo modo, puede obtenerse la correlación cruzada de  $\{y\}$  y  $\{x\}$  y también la autocorrelación del proceso de salida.

$$\begin{aligned} r_{yx}(t) &= \int h^*(\zeta)E\{x^*(\tau-\zeta)x(\tau+t)\}d\zeta = h^*(-t) * r_x(t) \\ S_{yx}(\omega) &= H^*(\omega)S_x(\omega) \end{aligned} \quad (II.26)$$

Finalmente, puede expresarse la autocorrelación del proceso de salida en función del de la entrada. En (II.27) aparecen estas relaciones entre autocorrelaciones y densidades espectrales. Es de destacar que la fase del sistema lineal no tiene influencia alguna en el cambio de la capacidad de predicción del proceso o en su autocorrelación. Esta es la razón por la cual aquellos sistemas que basan su percepción o aprendizaje mediante el uso de la función de autocorrelación no presenten sensibilidad a la fase cuando la señal se filtra linealmente. En definitiva:

$$\begin{aligned} r_y(t) &= h(t) * h^*(-t) * r_x(t) \\ S_y(\omega) &= |H(\omega)|^2 S_x(\omega) \end{aligned} \quad (II.27)$$

La utilidad de estas relaciones es crucial, entre otras razones, porque constituyen una alternativa real y práctica a la medida de sistemas lineales. Tradicionalmente se piensa que la medida de la  $h(t)$ , o de la función de transferencia puede realizarse únicamente mediante la aplicación de una delta a la entrada. Al margen de las dificultades prácticas de implementar una delta y que son materia de un curso de sistemas lineales, baste decir que el elevado rango dinámico que una delta implica hace inviable su uso. La mayor parte de los sistemas lineales lo son en un determinado rango denominado ‘pequeña señal’ y dejan de serlo, y de ser observable la  $h(t)$ , si la señal de entrada supera en dinámica la zona de pequeña señal. Es habitual encontrar que las tensiones de ruptura de dispositivos o capacidades asociadas son valores que hacen prohibitivo cualquier señal que intente parecerse a una delta. En definitiva, se hace

imprescindible el disponer de una señal de entrada que con dinámica finita permita medir la  $h(t)$ . En este sentido ha de recordarse lo fácil que es disponer de ruido blanco (densidad espectral de potencia plana), simplemente como la tensión de salida de una unión semiconductora. Al utilizar ruido blanco, las dos relaciones mostradas en (II.28) permitirían medir la  $h(t)$  o la  $H(\omega)$  del sistema:

$$H(\omega) = \frac{S_{xy}(\omega)}{S_x(\omega)} = \frac{S_{xy}(\omega)}{N_o} \quad (\text{II.28})$$

$$H(\omega) = \frac{S_y(\omega)}{S_{yx}(\omega)}$$

Claramente, en un principio, cualquiera de las dos es válida. En el tema de filtrado de Wiener se probará que la primera es la estimación de mínimo error cuadrático medio, mientras que la segunda no esta respaldada por ningún criterio. Por esta razón es fácil comprobar su sensibilidad al ruido y/o la posible distorsión no lineal del sistema a medir. Se deja al lector la prueba de este punto.

Es de destacar que, con una señal de dinámica controlada como ruido blanco, es posible medir con absoluta precisión la respuesta de un sistema lineal sin recurrir a la inviable función delta. Tanto es así, que cualquier analizador de redes utiliza ruido blanco y calculo de densidad espectral cruzada, en lugar de la inasequible e inadecuada función delta para evaluar la respuesta de amplificadores, filtros o cualquier otro sistema lineal.

Para justificar el interés de las relaciones anteriores, se planteara el problema de una manera más general. Suponiendo que se desea medir la respuesta  $H(\omega)$  de un sistema lineal y que se utiliza (II.29) para su estimación, se analizara su calidad en un sistema practico.

$$\hat{H}(\omega) = \frac{S_{xy}(\omega)}{S_x(\omega)} \quad (\text{II.29})$$

En el caso de que el sistema a medir presente, como se indica en la Figura II.8, ruido o distorsión no lineal caracterizada por la señal  $w(t)$ , al calcular la densidad cruzada, asumiendo que el ruido y/o la distorsión son independientes de la entrada, esta será igual al producto de  $H(\omega)$  por la densidad de la entrada, lo que prueba lo correcto del estimador. Factores como la calidad de la estimación de la densidad espectral cruzada, son la única limitación de la calidad final de la expresión (II.29). La cuestión adicional que se plantea es saber en qué rango frecuencial está presente el ruido y cuán fuerte es la distorsión no lineal. Por ejemplo, en un amplificador, en pequeña señal  $w(t)$  será el ruido de los dispositivos; mientras que en gran señal  $w(t)$  será principalmente distorsión no lineal. Pero el interés es saber cual es su distribución frecuencial.

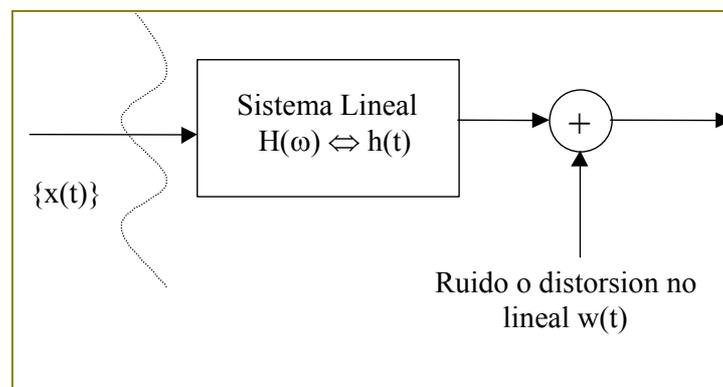


Figura II.8. Medida de un sistema lineal con ruido vía la excitación con ruido blanco  $\{x\}$ .

Para ver el impacto de  $w(t)$  en la medida de  $H(\omega)$  basta observar que un sistema lineal no genera, a la salida, frecuencias no presentes a la entrada. En otras palabras, todas las frecuencias o contenido espectral observado a la salida han de estar presentes también a la entrada (coherencia). Esta idea sugiere la definición de la denominada coherencia espectral. La definición de su módulo es la siguiente:

$$|\gamma(\omega)|^2 = \frac{S_{xy}(\omega)S_{yx}(\omega)}{S_x(\omega)S_y(\omega)} \quad (\text{II.30})$$

La cualidad crucial que esta función posee es que, para cualquier  $H(\omega)$  lineal, la coherencia espectral es idéntica a la unidad para todo el margen de frecuencia. Dicho de otro modo, si al medir la coherencia espectral, junto con la estimación (II.29), se observa que esta cae por debajo de la unidad se puede estar seguro de que, en aquellas frecuencias en las que ocurra, el sistema o tiene ruido o distorsiona.

Observando el caso de la Figura II.8, cuando  $w(t)$  esta incorrelado con la entrada, o su densidad espectral cruzada  $S_{wx}(\omega)$  es cero, las densidades cruzadas y la de la salida serian:

$$\begin{aligned} S_{xy}(\omega) &= S_x(\omega)H(\omega) \\ S_{yx}(\omega) &= S_x(\omega)H^*(\omega) \\ S_y(\omega) &= S_x(\omega)|H(\omega)|^2 + S_w(\omega) \\ \text{asumiendo que } S_{xw}(\omega) &= S_{wx}(\omega) = 0 \quad \forall \omega \end{aligned} \quad (\text{II.31})$$

Al sustituir las relaciones anteriores en la expresión de la coherencia espectral se obtiene (II.32):

$$|\gamma(\omega)|^2 = \frac{1}{1 + \frac{S_w(\omega)}{|H(\omega)|^2 S_x(\omega)}} \quad (\text{II.32})$$

Esta ultima relación evidencia que la responsabilidad de que la coherencia caiga por debajo de la unidad es únicamente la perdida de linealidad del sistema bajo medida. Es claro que, cuando el ruido o la distorsión no lineal son nulos la coherencia es la unidad en cualquier frecuencia. También es de destacar que cuando la respuesta del sistema bajo medida es cero la coherencia cae también a cero. Este último comportamiento es lógico si se medita sobre el papel del sistema, con respecto a reproducir frecuencias de la señal de entrada, cuando este no responde en absoluto o muy mal a la señal de entrada. Nótese también que la coherencia permite medir de manera precisa la relación señal a ruido a la salida del sistema, es decir, señal útil como la parte que responde a la señal y el ruido el resto. Esta expresión es de gran utilidad en la caracterización de canales de transmisión en comunicaciones:

$$|\gamma(\omega)|^2 = \frac{1}{1 + \frac{S_w(\omega)}{|H(\omega)|^2 S_x(\omega)}} = \frac{1}{1 + \frac{1}{SNR(\omega)}} = \frac{SNR(\omega)}{SNR(\omega) + 1} \quad (\text{II.33})$$

Se deja al lector la comprobación de cuáles son los efectos derivados de las situaciones donde el ruido es coherentes con la señal a la salida del sistema lineal, y hasta que punto esto ha de considerarse nocivo para la medida de la respuesta del sistema lineal.

En la figura II.9 puede observarse la coherencia espectral entre la entrada y la salida de un sistema, supuestamente, lineal. El sistema presenta distorsión no lineal en dos zonas de frecuencia por debajo de su frecuencia de corte nominal que era de 60 Khz (aquellas en las que el modulo de la coherencia cae manifiestamente por debajo de 1). De la figura puede concluirse en qué rangos de frecuencia se trata y de cual es el nivel o relación señal a ruido a efectos de la salida.

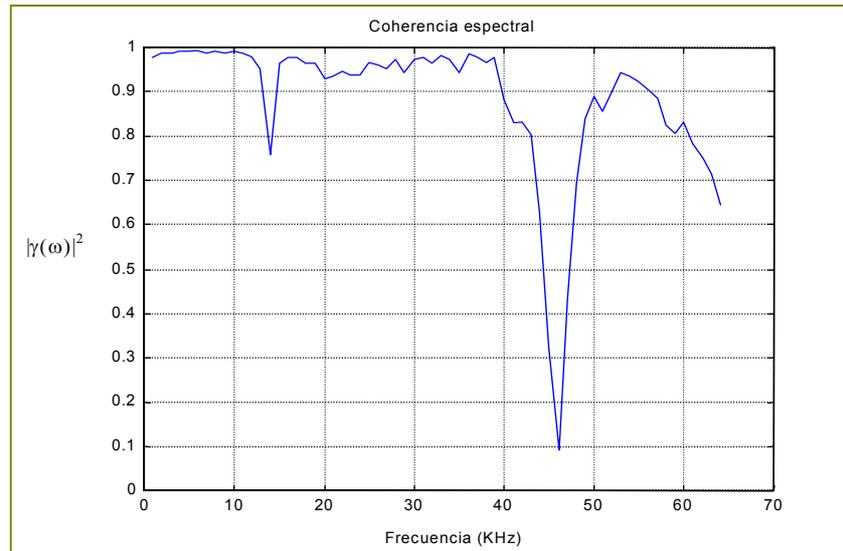


Figura II.9 Coherencia espectral medida para un sistema lineal que presenta distorsión no lineal.

## II.5. PROBABILIDAD ENTROPIA Y ESTIMACION.

El objetivo de esta sección es establecer los criterios que permiten, a partir de una única realización de un proceso, que en la mayor parte de los casos serán muestras de una señal, estimar parámetros ya sean de la propia señal, ya sean del fenómeno físico que la produce.

El concepto de estimación está íntimamente relacionado con el concepto de información. La forma de explicar intuitivamente esta relación es que plantear la estimación de un parámetro  $\theta$  a partir de una colección de muestras agrupadas en un vector  $\underline{X}$  es equivalente a plantear cuánta información del parámetro tienen las muestras disponibles. Así pues, se hace necesario revisar muy brevemente la relación del concepto de información con otro, que ya es mucho más familiar al lector, que es el de probabilidad.

Para analizar la relación entre probabilidad e información basta pensar en que la información, medida como se desee, es siempre inversa a la probabilidad. Dicho de otro modo, si alguien nos relata una cosa que pasa todos los días (muy probable), proporciona mucha menos información que si se relata una cosa que ocurre cada cien años (muy poco probable). De esta manera, una forma coherente de definir la información que nos da  $x$  sería la inversa de su probabilidad de ocurrencia. No obstante, definida de este modo, la información iría desde 1 (la menor información) hasta infinito (la máxima). Por razones prácticas interesa cambiar la escala y poner la información cero con la probabilidad uno y la información infinita con la probabilidad cero. Esto se consigue fácilmente empleando una medida logarítmica. La base del logaritmo podría ser cualquiera, sin embargo parece lógico definirla en base 2, dado que buena parte de nuestro aprendizaje se realiza en base a respuestas de sí o no cierto a nuestros planteamientos. En definitiva, se define la información  $I(x)$  asociada a una variable aleatoria según se indica en (II.34), y la unidad de medida es el "bit" (el bit es una medida de información y no una señal o un pulso).

$$I(x) = -\log_2[f(x)] \text{ bits} \quad (\text{II.34})$$

Del mismo modo, la información media o entropía de la variable aleatoria sería el promedio estadístico de la información.

$$H(x) = -\int f(x) \log_2[f(x)] dx \quad (\text{II.35})$$

Retomando el problema de estimar un parámetro dada la observación de un vector de datos, se vea que, de nuevo, la imprecisión o aleatoriedad se mezclan, e introducen confusión sobre el problema. Imaginando que el parámetro a estimar es determinístico, la imposibilidad práctica de conocerlo exactamente hace que se hable de probabilidad de que el valor correcto coincida con un valor dado  $\theta$ . Planteado de este modo, parece claro que considerar el parámetro a estimar como una v.a. se adapta mucho a la realidad. Obviamente, la calidad de la estimación ha de depender de los datos observados, por

lo que en lugar de pdf absoluta hablaremos de la pdf del parámetro  $\theta$  condicionado a los datos que se han podido observar. Si esta pdf condicional es conocida, podemos formular el estimador como aquel valor que maximice la probabilidad condicional de que el parámetro se encuentre en un intervalo  $d\theta$ .

$$\begin{aligned}\theta_{MAP} &= \arg \max_{\theta_o} \left[ \Pr \left( \theta_o - \frac{d\theta}{2} \leq \theta \leq \theta_o + \frac{d\theta}{2} / \underline{X}_n \right) \right] = \\ &= \arg \max_{\theta_o} [f(\theta_o / \underline{X}_n) d\theta] = \arg \max_{\theta_o} [f(\theta_o / \underline{X}_n)]\end{aligned}\quad (\text{II.36})$$

Al mismo tiempo, obsérvese que la hipótesis más probable será la de mínima información o, sencillamente, la mas habitual. Este criterio de probabilidad máxima, o información mínima para obtener lo mas habitual, dada la observación se denomina MAP o máximo a posteriori (es decir, se obtiene bajo la hipótesis de los datos observados).

Este criterio óptimo para la estimación de  $\theta$  tiene un soporte formal pero conlleva la determinación de una pdf condicional que es complicada en términos prácticos, ver (II.36). Un criterio, más intuitivo aparentemente, sería aquel que, dados los datos, estimase el parámetro que menos se desvía del correcto en valor cuadrático medio. Su formulación aparece en (II.37) y se le denomina estimación MSE (Mean Square Error) o, no siempre de un modo correcto, como de Mínima Varianza (MV):

$$\theta_{MSE} = \arg \min_{\theta_o} E \left\{ (\theta - \theta_o)^2 / \underline{X}_n \right\} \quad (\text{II.37})$$

La primera cuestión es responder en bajo que circunstancias el estimador de MSE coincidirá con el óptimo MAP. Para ver la relación entre ambos se encontrara una formulación alternativa de (II.37). La nueva formulación del error cuadrático medio puede verse en (II.38).

$$\begin{aligned}E \left\{ (\theta - \theta_o)^2 / \underline{X}_n \right\} &= \int (\theta - \theta_o)^2 f(\theta / \underline{X}_n) d\theta \\ \frac{\partial E \left\{ (\theta - \theta_o)^2 / \underline{X}_n \right\}}{\partial \theta_o} &= 0 = \theta_o - \int \theta f(\theta / \underline{X}_n) d\theta\end{aligned}\quad (\text{II.38})$$

Como puede verse el valor que anula la derivada es precisamente la media condicional del parámetro. Por lo tanto el estimador MSE también se denomina de media condicional.

$$\theta_{MSE} = \int \theta f(\theta / \underline{X}_n) d\theta \quad (\text{II.39})$$

Esta última formulación revela que el estimador MSE coincidirá con el MAP tan solo cuando el máximo de la distribución condicional coincida con su media. Nótese que esta situación se da para muchas pdf's, entre las cuales la Gaussiana es la más habitual. No obstante es fácil imaginar infinitas distribuciones donde el máximo diferirá de la media y por tanto el MSE sería subóptimo. Recuerde pues que la similitud entendida como distancia euclídea es solo la mejor medida de parecido cuando la distribución es Gaussiana.

Volviendo al estimador MAP, como se ha mencionado, la formulación de la probabilidad condicional es difícil. Para evidenciarlo, se pondrá un ejemplo en el que se intenta estimar la media de un proceso Gaussiano a partir de un conjunto de observaciones. Tomando  $Q$  muestras de una realización de un proceso con matriz de covarianza  $\underline{R}$  y de media  $\mu$ , es fácil expresar el vector de observaciones como:

$$\underline{X}_n = \mu \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} + \underline{w}_n = \mu \underline{1} + \underline{w}_n \quad (\text{II.40})$$

$$E\{\underline{w}_n\} = 0 \quad ; \quad E\{\underline{w}_n \underline{w}_n^H\} = \underline{\underline{R}} \quad \forall n \text{ (estacionario)}$$

A partir de la formula anterior, absolutamente general, es fácil expresar la probabilidad del vector de datos dada la media como (II.41):

$$f(\underline{X}_n / \mu) = \frac{1}{\pi^Q \cdot \det(\underline{\underline{R}})} \exp\left\{- (\underline{X}_n - \mu \underline{1})^H \underline{\underline{R}}^{-1} (\underline{X}_n - \mu \underline{1})\right\} \quad (\text{II.41})$$

La formulación de  $f(\mu / \underline{X}_n)$  a partir de (II.41) es fácil vía el teorema de Bayes, donde se puede prescindir del termino del denominador, al no influir en el proceso de maximización:

$$f(\mu / \underline{X}_n) = \frac{f(\underline{X}_n / \mu) f(\mu)}{f(\underline{X}_n)} \propto f(\underline{X}_n / \mu) f(\mu) \quad (\text{II.42})$$

Es decir, el criterio MAP conlleva la maximización de la pdf de los datos condicionada al parámetro a determinar multiplicada por la pdf del parámetro o hipótesis. Dicho de otro modo la información que proporcionan los datos del parámetro es la información de los datos que da el parámetro más una información a priori sobre el parámetro que viene dada por el conocimiento de su pdf.

Ya que es relativamente sencillo formular la pdf condicionada al parámetro (función a la que habitualmente se le denomina la verosimilitud o likelihood) y más difícil el tener un conocimiento fiable de  $f(\mu)$ , se usa la primera de manera aislada para construir un criterio de estimación, de nuevo subóptimo, al que se denomina de máxima verosimilitud (ML):

$$\theta_{ML} = \arg \max_{\theta} f(\underline{X}_n / \theta) \quad (\text{II.43})$$

Para evidenciar de manera coloquial y un tanto burda la diferencia entre las ecuaciones (II.36) y (II.43), puede reflexionarse sobre cuál de estas dos cuestiones es más fácil de responder: sabiendo que un equipo de fútbol es bueno, ¿cómo serán los siguientes N resultados  $f(\underline{X}_n / \theta)$ ?, o bien, dados N resultados, ¿cuán bueno es el equipo  $f(\theta / \underline{X}_n)$ ? Tomando el principio de que el éxito justifica y que no se debate lo injusto o justo de nuestro juicio, sino principios de probabilidad, parece claro que la segunda es más difícil o entraña una información más elaborada.

Es posible, finalmente, determinar cuando la estimación ML coincide con la estimación MAP. A partir de (II.42) es obvio que cuando la distribución de las hipótesis es uniforme (máxima entropía, máximo desconocimiento o mínima información de su distribución) entonces la estimación MAP es idéntica a la ML. Dicho de otro modo, es la información a priori que se dispone del parámetro lo que permite pasar del criterio ML al criterio MAP.

En el próximo apartado se mostrara claramente la diferencia entre ambos estimadores, concentrándose en la estimación de la media de un proceso a partir de Q muestras de una realización.

## II.6 ESTIMACION MAP Y ML DE LA MEDIA DE UN PROCESO

Por sencillez, se procede en primer lugar a derivar el estimador ML de la media a partir del vector de Q muestras. Dado que al tomar una función monótona no decreciente de otra función no se alteran sus extremos (máximos o mínimos), es habitual trabajar con la denominada log-likelihood o el logaritmo de la probabilidad condicional correspondiente. Al tomar logaritmo neperiano sobre (II.41) el único termino que depende de la media y que se ha de minimizar es su exponente cambiado de signo (II.44)

$$\Lambda(\mu) = [\underline{X}_n - \mu \underline{1}]^H \underline{R}^{-1} [\underline{X}_n - \mu \underline{1}] \quad (\text{II.44})$$

Calculando el gradiente de la log-likelihood con respecto al conjugado, ver Capitulo I, e igualar al vector cero, se obtiene el estimador deseado.

$$\mu_{ML} = \frac{\underline{1}^H \underline{R}^{-1} \underline{X}_n}{\underline{1}^H \underline{R}^{-1} \underline{1}} \quad (\text{II.45})$$

Antes de proseguir con el estimador MAP, es interesante destacar tres aspectos de este estimador. En primer lugar, el valor estimado depende de las observaciones recogidas en el vector  $\underline{X}_n$ . Como tal, si recogiéramos en ese vector valores procedentes de distintas realizaciones del proceso obtendríamos distintos valores estimados. En definitiva, la estimación obtenida va a ser una variable aleatoria siempre.

En segundo lugar, el estimador puede formularse como el filtrado del vector de datos con un filtro de respuesta impulsional  $\underline{h}$ . En otras palabras, el estimador de media en un proceso Gaussiano es lineal y, por lo tanto, puede implementarse con un sistema lineal.

$$\underline{h} = \alpha \underline{R}^{-1} \underline{1} \quad \text{donde la constante } \alpha \text{ es la inversa de } \underline{1}^H \underline{R}^{-1} \underline{1}$$

$$\mu_{ML}(n) = \underline{h}^H \begin{bmatrix} x(n) \\ x(n-1) \\ \vdots \\ x(n-Q+1) \end{bmatrix} \quad (\text{II.46})$$

Es decir, la estimación de media se actualiza en cada instante en función de los datos que se introducen a la entrada.

De aquí puede entenderse el porqué la Gaussianidad de gran parte de los fenómenos naturales ha fomentado el uso de los sistemas lineales y no al revés. Si la mayor parte de los fenómenos de interés no hubiesen sido gaussianos, tendría que haberse desarrollado, más de lo que esta hoy en día, la teoría de los sistemas no-lineales.

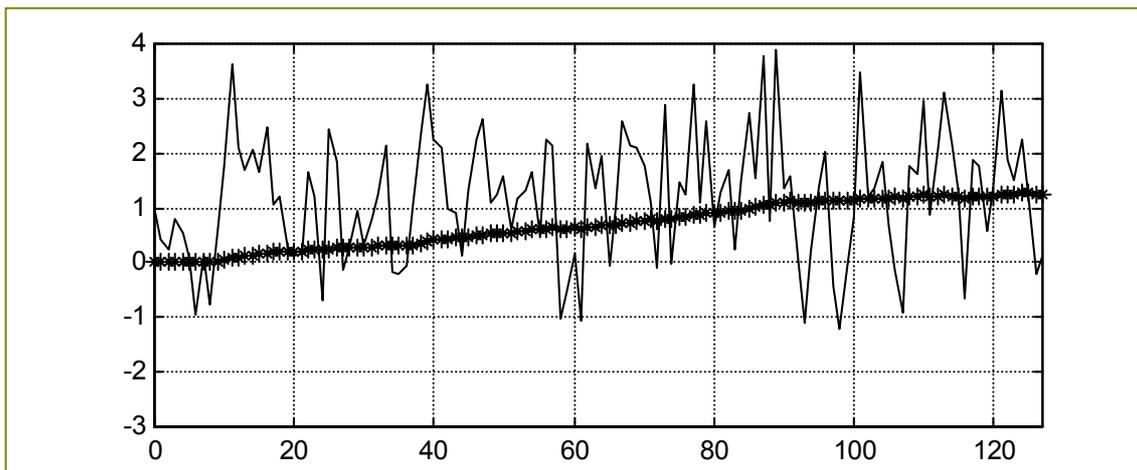


Figura II.10. Estimación de la media (valor exacto unidad) de un proceso gaussiano no blanco. La longitud del estimador es de 50 coeficientes. Obsérvese el transitorio correspondiente a un FIR de la longitud mencionada.

El tercer aspecto sobre el que se ha de llamar la atención es que, intuitivamente, la media de una colección de  $Q$  muestras debería estimarse sumando y dividiendo por  $Q$  (media aritmética). Es interesante que el criterio revela que eso no es lo mejor que se puede hacer para estimar la media. Entonces, ¿cuándo

la media aritmética es el mejor estimador de la media? O dicho de otro modo, ¿cuándo el vector  $\underline{h}$  es igual al vector de unos multiplicado por  $1/Q$ ? La respuesta es inmediata, solo cuando la matriz de correlación es diagonal entonces el estimador coincide con la media aritmética:

$$\begin{aligned}\underline{R} &= \sigma^2 \underline{I} \\ \mathbf{1}^H \underline{R}^{-1} \mathbf{1} &= \frac{Q}{\sigma^2} \\ \underline{h} &= \frac{\mathbf{1}}{Q} \\ \mu_{ML} &= \mathbf{1}^H \underline{X}_n / Q\end{aligned}\tag{II.47}$$

En resumen, sólo cuando la parte de media nula del proceso es blanca o incorrelada la media aritmética es el mejor estimador. El lector puede imaginar en cuantas situaciones esta formulación intuitiva de la media es inadecuada.

Para pasar a la estimación MAP, se ha de disponer de una información a priori como por ejemplo que, en media, la media estimada esté alrededor de un valor  $\mu_m$  y que el 90% de las veces esté comprendida en un margen  $\pm \sigma_m$ . A partir de este conocimiento, y para simplificar la formulación logarítmica del criterio MAP, se puede asumir que el parámetro se muestra con una distribución gaussiana de media y varianza como se indica en (II.48).

$$f(\mu) = \frac{1}{\sqrt{2\pi}\sigma_m} \exp\left(-\frac{1}{2} \frac{(\mu - \mu_m)^2}{\sigma_m^2}\right)\tag{II.48}$$

Al derivar para obtener el máximo del producto de la likelihood por la probabilidad a priori, criterio MAP, se obtiene la expresión (II.49):

$$\frac{\partial}{\partial \mu} \left[ \underline{X}_n - \mu \mathbf{1} \right]^H \underline{R}^{-1} \left[ \underline{X}_n - \mu \mathbf{1} \right] + \frac{(\mu - \mu_m)^2}{\sigma_m^2} = \mathbf{1}^H \underline{R}^{-1} \left[ \underline{X}_n - \mu \mathbf{1} \right] + \frac{(\mu - \mu_m)}{\sigma_m^2} = 0\tag{II.49}$$

y el estimador MAP:

$$\mu_{MAP} = \frac{\mathbf{1}^H \underline{R}^{-1} \underline{X}_n + \left(\frac{\mu_m}{\sigma_m^2}\right)}{\mathbf{1}^H \underline{R}^{-1} \mathbf{1} + \left(\frac{1}{\sigma_m^2}\right)}\tag{II.50}$$

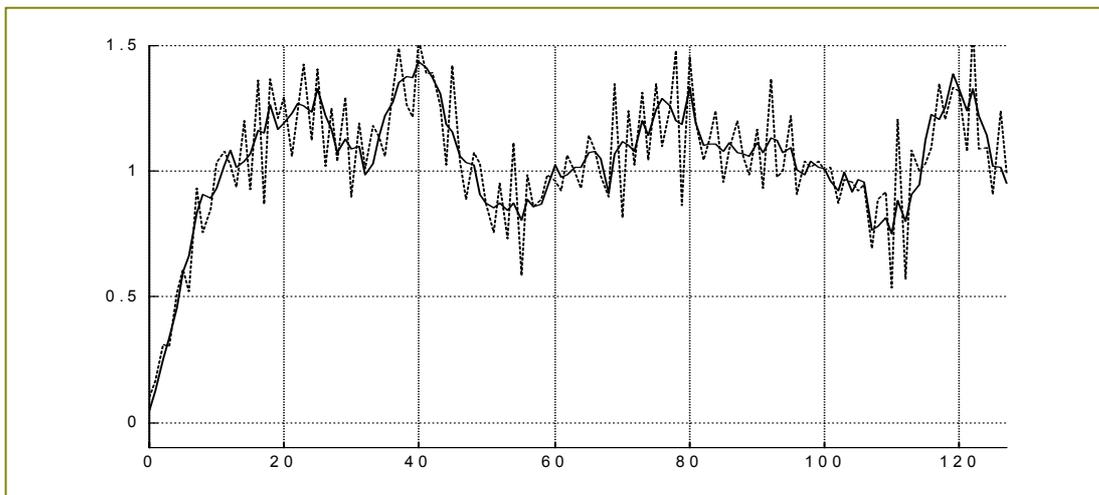


Figura II.11. Estimador de media de un proceso gaussiano no blanco. La media correcta es la unidad. El optimo en línea continua, la media aritmética en discontinua. Ambos de longitud 10 muestras. Obsérvese la mayor varianza del estimador no optimo o de media aritmética.

Obsérvese que, efectivamente, cuando el desconocimiento a priori es máximo, lo que equivale a que la varianza del parámetro tienda a infinito, el mejor estimador coincide con el ML. En el otro extremo, la certeza de que la media es siempre  $\mu_m$  (varianza cero) hace obvio el que la mejor estimación no depende de los datos y es directamente el valor mencionado. En general, MAP es un compromiso óptimo entre conocimiento previo e información de las observaciones. De nuevo, nótese la mezcla de imprecisión o desconocimiento y como se plasma ésta en parámetros que caracterizan esa imprecisión. Es de señalar que no se ha considerado el estimador de mínima varianza dado que las distribuciones implicadas eran gaussianas y, por lo tanto coincidiría con el estimador MAP.

## II.7. LA CALIDAD DE UN ESTIMADOR

A lo largo de las secciones anteriores se ha hablado de este estimador mejor que aquel otro. Surge pues la pregunta de cómo se evalúa la calidad del estimador. Considerando el caso de que el parámetro a estimar es determinista y efectivamente vale  $\theta$ , es evidente, que al estimarlo desde observaciones, la estimación es en si una v.a. A esta variable aleatoria parece intuitivo pedirle que, en media, coincida con el valor correcto. De otro modo, cuando la media del estimador no coincide con el valor correcto se dirá que el estimador tiene un sesgo  $b$  definido como sigue:

$$b^2 \equiv [E\{\hat{\theta}\} - \theta]^2 \quad (\text{II.51})$$

Claramente, el sesgo es una medida del error sistemático que comete el estimador cuando se efectúa la estimación sobre distintas realizaciones. Además del sesgo, es necesario que, además de que en media converja al valor correcto la v.a., el estimador no presente una desviación grande respecto a dicha media, es decir, su varianza sea pequeña.

$$\sigma_{\hat{\theta}}^2 \equiv E\left\{\left(\hat{\theta} - E\{\hat{\theta}\}\right)^2\right\} \quad (\text{II.52})$$

En el fondo, el sesgo y la varianza del estimador son medidas de cuan próxima es la pdf del valor estimado a una delta centrada en el valor a estimar. Es más, ni siquiera el sesgo es un problema grave ya que si el estimador mide siempre 2 mts. con un sesgo negativo de 1 mts siempre se sabrá que el valor verdadero es 3 mts. De lo dicho, no debe dársele mayor gravedad de la que tiene al hecho de que un estimador sea sesgado, lo importante es que el sesgo no dependa de los datos y sea conocido.

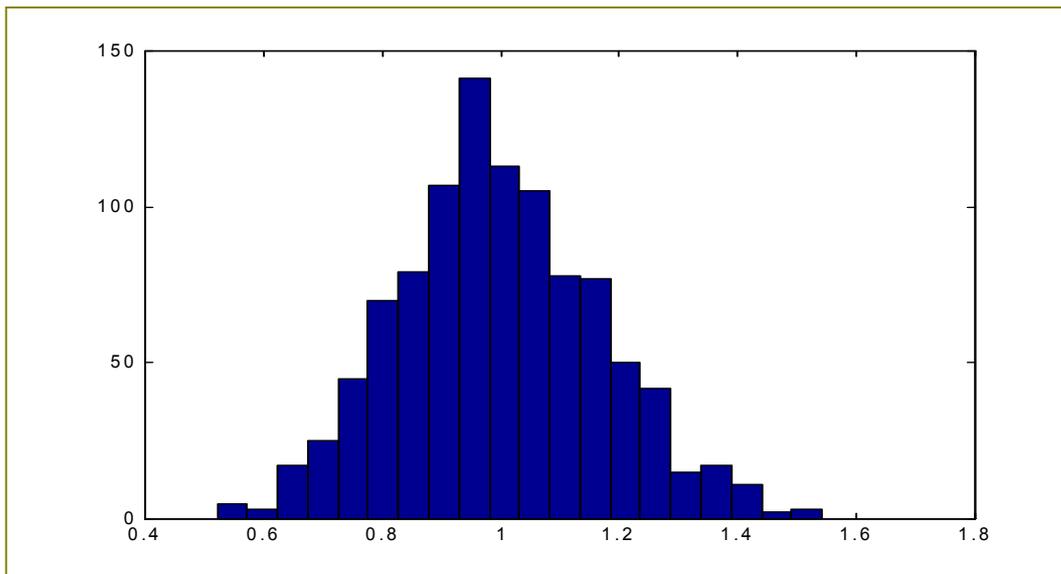


Figura II.12. En la figura puede verse el histograma de la media estimada en cada instante por el estimador óptimo de longitud 10 muestras para el proceso de la figura anterior. Nótese que la media es uno y la varianza es del orden de 0.2.

Una razón que avala el uso de los dos parámetros mencionados es que el error cuadrático medio del estimador con respecto al valor correcto (no confundir con la formulación del estimador de mínima

varianza) es igual a la suma del sesgo al cuadrado más la varianza. Esta expresión es fácil de probar recordando el carácter determinista del parámetro  $\theta$ :

$$E\{(\hat{\theta} - \theta)^2\} = E\{[(\hat{\theta} - E\{\hat{\theta}\}) - (\theta - E\{\hat{\theta}\})]^2\} = b^2 + \sigma_{\hat{\theta}}^2 \quad (\text{II.53})$$

A continuación, se aplicaran estas dos medidas de calidad al estimador de la media de un proceso derivado en el apartado anterior.

Al tomar el valor esperado del estimador, lo único que es aleatorio es el vector de datos, y su media es la media correcta multiplicada por el vector de unos. Con lo que, se puede concluir que el estimador de máxima verosimilitud de la media es insesgado.

$$E\{\mu_{ML}\} = \frac{\mathbf{1}^H \underline{\underline{R}}^{-1} E(\underline{X}_n)}{\mathbf{1}^H \underline{\underline{R}}^{-1} \mathbf{1}} = \mu \quad (\text{II.54})$$

Con respecto a la varianza, en (II.55) puede verse su expresión.

$$\begin{aligned} \sigma_{ML}^2 &= E\left\{\left(\frac{\mathbf{1}^H \underline{\underline{R}}^{-1} \underline{X}_n}{\mathbf{1}^H \underline{\underline{R}}^{-1} \mathbf{1}} - \mu \frac{\mathbf{1}^H \underline{\underline{R}}^{-1} \mathbf{1}}{\mathbf{1}^H \underline{\underline{R}}^{-1} \mathbf{1}}\right)^2\right\} = E\left\{\left(\frac{\mathbf{1}^H \underline{\underline{R}}^{-1} E\{\underline{X}_n - \mu \mathbf{1}\}}{\mathbf{1}^H \underline{\underline{R}}^{-1} \mathbf{1}}\right)^2\right\} \\ &= \frac{\mathbf{1}^H \underline{\underline{R}}^{-1} E\{\underline{N}_n \underline{N}_n^H\} \underline{\underline{R}}^{-1} \mathbf{1}}{(\mathbf{1}^H \underline{\underline{R}}^{-1} \mathbf{1})^2} = \frac{1}{\mathbf{1}^H \underline{\underline{R}}^{-1} \mathbf{1}} \end{aligned} \quad (\text{II.55})$$

De esta última expresión se han de destacar dos cosas. La primera es que en el caso de que la matriz de correlación sea diagonal (en general también puede afirmarse pero menos evidente) la varianza tiende a cero cuando la longitud  $Q$  tiende a infinito. Cuando un estimador tiene esta propiedad se dice que es consistente.

La segunda se refiere al valor de dicha varianza. Ya se ha visto que el método de estimación de un parámetro no es único y depende del criterio de partida (ML, MAP, MSE o cualquier otro). En cada caso, el estimador obtenido tendrá una varianza distinta, que será siempre mayor que cero si el vector de observaciones es finito. Cabe preguntarse, si hay una cota mínima para la varianza obtenida en la estimación insesgada de un parámetro, la respuesta es que sí que existe siempre que el estimador sea insesgado. Esa cota mínima viene dada por el límite de Cramer-Rao:

$$\sigma_{estimator}^2 \geq \frac{1}{E\left\{\left|\frac{\partial}{\partial \theta} \ln(f(\underline{X}_n/\theta))\right|^2\right\}} \quad (\text{II.56})$$

La deducción de esta cota aparece en el apéndice A. El conocimiento de esta cota es de gran utilidad para cualquier problema de estimación ya que, cuanto más próxima este la varianza de un estimador a ella, se podrá decir de forma absoluta que mejor es el estimador; eso si, recordando que la validez del uso de la cota anterior esta condicionada al carácter insesgado del estimador.

El estimador de media derivado en la sección anterior es eficiente pues alcanza la cota inferior de Cramer-Rao como puede verse a continuación. El segundo termino de la ecuación (II.56) es:

$$\begin{aligned} \frac{\partial}{\partial \mu} \ln(f(\underline{X}_n/\mu)) &= \mathbf{1}^H \underline{\underline{R}}^{-1} (\underline{X}_n - \mu \mathbf{1}) \\ E\left\{\mathbf{1}^H \underline{\underline{R}}^{-1} (\underline{X}_n - \mu \mathbf{1}) (\underline{X}_n - \mu \mathbf{1})^H \underline{\underline{R}}^{-1} \mathbf{1}\right\} &= \mathbf{1}^H \underline{\underline{R}}^{-1} \mathbf{1} \end{aligned} \quad (\text{II.57})$$

y por lo tanto, la varianza encontrada en (II.55) coincide con la cota de Cramer-Rao.

De nuevo hemos de insistir en la distribución gaussiana de las observaciones, ya que es ella la que facilita el diseño del estimador como un sistema lineal. No obstante, como podrá verse en el siguiente apartado, el tomar sesgo y varianza como parámetros de calidad permite derivar estimadores ad-hoc, mediante sistemas lineales, incluso sin entrar en la distribución de los datos.

## II.8. ALISADO DE DATOS.

El problema que se pretende resolver es la estimación de una forma de onda determinista en ruido, sea cual sea la distribución de este, con el criterio de controlar sesgo y varianza. Un diseño elaborado del sistema requeriría de un conocimiento a priori de la forma de onda buscada. En muchas aplicaciones, como en el caso de detección de pulsos o simplemente eliminar ruido de una trayectoria en sistemas de seguimiento, se desconoce la forma de onda buscada, y aun así se requiere eliminar de la señal recibida el ruido aditivo que esta presenta. Cuando el desconocimiento de la forma de onda es absoluto, un recurso es simplemente alisar los datos vía un integrador. El proceso a llevar a cabo esta representado en la Figura II.13 donde se plantea el alisado vía un integrador de T seg. El objetivo es diseñar el periodo de integración T, posiblemente variante con el tiempo, de modo que sesgo y varianza sean lo mas pequeños posible. Al mismo tiempo, esta aplicación ilustra como sesgo y varianza suelen presentar un compromiso, compromiso que se dilucida minimizando el error cuadrático medio, como podrá verse.

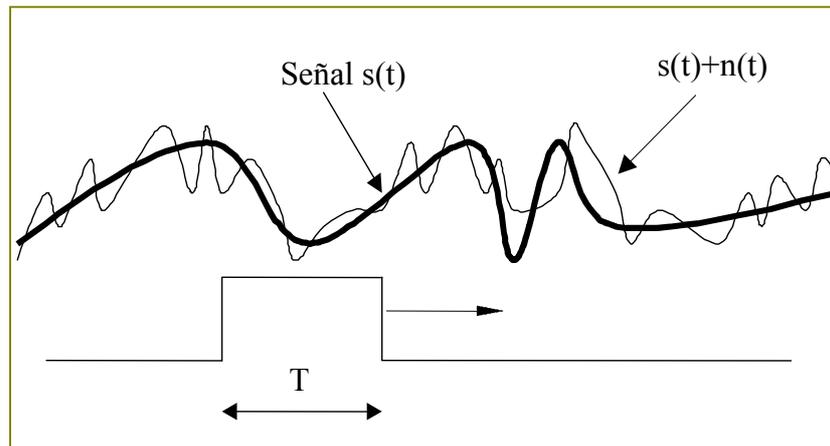


Figura II.13. Alisado de datos como respuesta a un integrador de duración T seg.

La salida del integrador en un instante t viene dada por (II.58), donde se ha elegido como instante de salida el centro del intervalo de integración para evitar distorsión o carga de fase. Esta expresión recoge, en las dos formas en que se puede expresar la convolucion, las contribuciones de la señal deseada s(t) y el ruido n(t).

$$\hat{s}(t) = \frac{1}{T} \int_{t-\frac{T}{2}}^{t+\frac{T}{2}} (s(\tau) + n(\tau)) d\tau = \frac{1}{T} \int_{-T/2}^{T/2} (s(\phi + t) + n(\phi + t)) d\phi \quad (\text{II.58})$$

Claramente, tomando como referencia la segunda integral, el sesgo estará originado por la integración de la señal deseada y la varianza por la integral del ruido. De este modo, se puede formular el sesgo de la estimación como la diferencia de s(t) y la respuesta del integrador a s(t):

$$b(t) = s(t) - \frac{1}{T} \int_{-T/2}^{T/2} s(\phi + t) d\phi \quad (\text{II.59})$$

Para poder operar, más allá de la expresión anterior, se supondrá que, en el intervalo de integración  $T$ , la función no varía bruscamente y que puede aproximarse por su desarrollo en serie de Taylor alrededor del punto  $t$  hasta el término de orden 3. Nótese la calidad de la aproximación pues lo habitual en otras ocasiones es limitarse al orden 1; dicho de otro modo, se realiza una aproximación poco burda y con muy poco impacto en la validez del resultado final:

$$s(t + \phi) \cong s(t) + \phi s'(t) + \frac{\phi^2}{2} s''(t) + \frac{\phi^3}{6} s'''(t) \quad (\text{II.60})$$

Al introducir la aproximación en la integral, realizada según  $\phi$ , el primer término da lugar a  $s(t)$  que se cancela con el que entra en la definición del sesgo. El segundo y el cuarto son funciones impares integradas en un intervalo par y por tanto se anulan. En definitiva, el sesgo se produce únicamente con la derivada segunda de la señal deseada.

$$b(t) = \frac{T^2 s''(t)}{24} \quad (\text{II.61})$$

Es decir, el sesgo aumenta con el tiempo de integración y es máximo en los puntos donde la forma de onda a estimar presenta sus máximos y sus mínimos, pues los suaviza.

Con respecto a la varianza, será directamente la potencia del ruido a la salida, tal y como se expresa en el primer término de (II.62). Si se asume que el ruido es blanco a la entrada y de potencia  $\sigma^2$ , se puede expresar la varianza directamente como la potencia mencionada dividida por el periodo de integración. Como era de prever también, al aumentar el tiempo de integración disminuye la varianza y el estimador es consistente.

$$\text{var} = E \left\{ \left| \frac{1}{T} \int_{-T/2}^{T/2} n(t + \phi) d\phi \right|^2 \right\} = \frac{\sigma^2}{T} \quad (\text{II.62})$$

Llegado este punto se está en un compromiso entre sesgo y varianza en lo que se refiere a la determinación del  $T$  óptimo. El dilema se resuelve recordando que el error cuadrático medio es la suma de sesgo al cuadrado más la varianza:

$$MSE = b^2(t) + \text{var} = \frac{T^4 s''(t)^2}{24^2} + \frac{\sigma^2}{T} \quad (\text{II.63})$$

Derivando e igualando a cero se obtiene el  $T$  óptimo:

$$T_{opt} = \sqrt[5]{\frac{144\sigma^2}{s''(t)^2}} \quad (\text{II.64})$$

Esta expresión nos permite construir un sistema adaptativo en su periodo de integración, dependiendo del nivel de ruido con que se han realizado las observaciones, y de si se está en un extremo o en un punto de inflexión de la señal deseada o la trayectoria buscada. La precisión de la expresión permite obtener resultados muy buenos, aunque se tenga bastante imprecisión a la hora de suministrarle el nivel de ruido y la derivada segunda. También es de destacar, que no siempre un buen estimador, como el que nos concierne es de sesgo cero. En la mayor parte de las situaciones, el diseño del estimador con sistemas lineales, es un compromiso entre distorsión o sesgo y ruido o varianza.

## II.9. ESTIMACION DE LA FUNCION DE CORRELACION

Como ha podido verse en el Capítulo I, la función o bien la matriz de correlación, es necesaria cuando en un sistema lineal se expresa la potencia de la salida en función de la respuesta del sistema.

Además, la distribución espectral, no solo en lo que atañe al concepto de ancho de banda sino en el diseño de sistemas lineales, es una función crucial, de conocimiento obligado, en cada etapa de proceso. Aunque solo fuera por las razones enumeradas quedaría justificado el presente apartado; no obstante, se puede decir que en, prácticamente, el cien por cien de los contenidos de este curso, la función de correlación será una función necesaria. Definida como un valor esperado o como el momento de orden uno de una función de probabilidad conjunta, su calculo directo es poco menos que imposible. Como de costumbre la situación en la práctica es la de disponer de N muestras y con ellas se ha de obtener la mejor estimación posible de la función buscada.

Considerando que las muestras disponibles se disponen en un vector de N componentes  $\underline{X}_n(N)$ , se va a abordar el problema de la estimación de la matriz de correlación e indirectamente se obtendrá la estimación de sus componentes, es decir, de la función de correlación. Como premisa, se supondrá que el proceso es estacionario, al menos, en la duración del segmento de datos disponible y con el que se va a realizar la estimación.

La matriz de correlación esta definida según (II.65), donde se evidencia el carácter Toeplitz y hermitico de la matriz, es decir, todos los elementos en cualquier diagonal son iguales.

$$\underline{R} = E\{\underline{X}_n(N)\underline{X}_n^H(N)\} = \begin{bmatrix} r(0) & r(1) & r(2) & \cdots & r(N-2) & r(N-1) \\ r(-1) & r(0) & r(1) & \cdots & r(N-3) & r(N-2) \\ r(-2) & r(1) & r(0) & \cdots & r(N-4) & r(N-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ r(-N+2) & r(-N+3) & r(-N+4) & \cdots & r(0) & r(1) \\ r(-N+1) & r(-N+2) & r(-N+3) & \cdots & r(-1) & r(0) \end{bmatrix} \quad (\text{II.65})$$

La primera estimación y más cruda sería tomar el único producto disponible del promedio estadístico anterior, y además, ese único producto sin ningún tipo de promedio.

$$\hat{\underline{R}} = \underline{X}_n(N)\underline{X}_n^H(N) \quad (\text{II.66})$$

Este estimador no da lugar a una matriz Toeplitz, sin embargo es insesgado ya que el valor esperado de cualquier elemento coincide con el valor correcto,

$$E\{r_{i,j}\} = E\{x(n-i)x^*(n-j)\} = r_x(j-i) \quad (\text{II.67})$$

Sin embargo, la varianza del estimador sería:

$$\text{var}\{r_{i,j}\} = E\{r_{i,j}^2\} - r_x^2(j-i) \quad (\text{II.68})$$

Se calcula a continuación el primer termino de la expresión anterior, suponiendo el proceso gaussiano. En consecuencia, usando que el valor esperado de un producto de cuatro v.a. gaussianas de media nula es:

$$E\{x_1x_2x_3x_4\} = E\{x_1x_2\}E\{x_3x_4\} + E\{x_1x_3\}E\{x_2x_4\} + E\{x_1x_4\}E\{x_2x_3\}$$

la varianza del estimador anterior resulta ser:

$$\text{var}\{r_{i,j}\} = 2r_x^2(j-i) + r_x^2(0) - r_x^2(j-i) = r_x^2(j-i) + r_x^2(0) \quad (\text{II.69})$$

Esta varianza es un valor muy elevado, basta fijarse en la estimación de  $r_x(0)$ , la varianza sería  $1,41r(0)$  es decir, se estima 1 Km con una varianza de +/- 1.41 Km !! Claramente el estimador aunque es insesgado es totalmente inadmisibile en la práctica.

La primera solución para estimar  $r_x(m)$  de otro modo, con menor varianza, no es otra que promediar hasta donde sea posible. Dado que la matriz estimada no es Toeplitz y la real lo es, la primera idea es tomar como estimador de correlación la suma a lo largo de cada diagonal de la matriz (II.65). En efecto, al proceder de este modo, el estimador de correlación, el elemento que pondríamos en la diagonal  $m$  de la matriz estimada:

$$\begin{aligned} \underline{R} &= \text{Toeplitz}(\hat{R}) \\ r(m) &= \sum_{\forall(i-j)=m} r_{i,j} = \sum_{n=0}^{N-|m|-1} x^*(n)x(n+m) \quad m = 0, N-1 \end{aligned} \quad (\text{II.70})$$

En este punto surge el conflicto, el estimador (II.70) tiene una varianza inferior pero tiene un sesgo que depende del valor estimado. Se le denomina estimador sesgado con los valores de media y varianza mostrados en (II.71). No obstante, y dado que la potencia de un proceso es un parámetro crucial en ingeniería, para que el sesgo de la autocorrelación en cero (es decir la potencia) sea nulo es necesario normalizar el estimador (II.70).

$$\begin{aligned} E\{\tilde{r}(m)\} &= (N-|m|)r(m) \\ \text{var}\{\tilde{r}(m)\} &= \sum_{q=-N}^N (N-|q|) \left[ |r(q)|^2 + ra(m-q)rb(m+q) \right] \\ \text{siendo } ra(s) &= E[x(n).x(n+s)] \quad \text{y } rb(s) = ra^*(s) \end{aligned} \quad (\text{II.71})$$

De hecho la normalización produce el siguiente estimador, al dividir por el número de datos disponibles:

$$\begin{aligned} \hat{r}(m) &= \frac{1}{N} \sum_{n=0}^{N-|m|-1} x^*(n)x(n+m) \\ E\{\hat{r}(m)\} &= \left(1 - \frac{|m|}{N}\right) r(m) \\ \text{var}\{\hat{r}(m)\} &= \frac{1}{N} \sum_{q=-N}^N \left(1 - \frac{|q|}{N}\right) \left[ |r(q)|^2 + ra(m-q)rb(m+q) \right] \end{aligned} \quad (\text{II.72})$$

Una alternativa menos utilizada es la denominada del estimador insesgado que aparece en (II.73) junto con su varianza:

$$\begin{aligned} \tilde{r}(m) &= \frac{1}{N-|m|} \sum_{n=0}^{N-|m|-1} x^*(n)x(n+m) \\ E\{\tilde{r}(m)\} &= r(m) \\ \text{var}\{\tilde{r}(m)\} &= \frac{1}{(N-|m|)^2} \sum_{q=-N}^N (N-|q|) \left[ |r(q)|^2 + ra(m-q)rb(m+q) \right] \end{aligned} \quad (\text{II.73})$$

En este estimador, se normaliza por el número de términos que intervienen en el sumatorio.

La primera diferencia entre ambos estimadores es que intercambian sesgo y varianza. En el primer estimador el sesgo aumenta con  $m$ , y su varianza crece con  $m$ , pero proporcional a la propia función (ver segundo termino de (II.72)). Este termino suele decrecer con  $m$ , con lo que su evolución es la mayor parte de las veces constante con  $m$ . En el caso del estimador insesgado, su varianza crece independientemente de la función a medida que  $m$  aumenta. Así pues, en cualquier caso y al margen de que estimador se emplee al final, el error cuadrático medio aumenta cuanto mayor es el índice, es decir,

los valores de correlación para desplazamientos cortos son siempre más fiables que para desplazamientos largos. Con todo es en el estimador insesgado donde mayor crecimiento toma el error cuadrático medio.

El hecho de que la varianza crezca con el índice y en la medida que éste se acerca a N empeore, motiva que en la práctica no sea recomendable usar más de N/3 (30% de la longitud de los datos disponibles) valores de la correlación estimada pues su fiabilidad disminuye dramáticamente.

Si el aumento de la varianza fuese un argumento válido para descartar el estimador insesgado, se podría alegar que, al ser los sesgos distintos, el error cuadrático medio podría ser un argumento para rechazar el insesgado en favor del otro. Sin entrar en este detalle existe otro hecho fundamental para rechazar el estimador insesgado. Esta respuesta a que estimador emplear nace al querer indicar como sería el estimador de la densidad espectral de potencia, dada la matriz de autocorrelación. Es más, la consideración de cómo obtener la densidad espectral aclarará las dudas sobre cual de los dos estimadores disponibles es el más adecuado.

Dada una matriz de autocorrelación de dimensión Q (matriz QxQ), la densidad espectral de potencia se puede formular como se indica en (II.74), donde se usa el carácter Toeplitz de la matriz de autocorrelación y que la norma del vector  $\underline{S}$  es igual a Q. Esta expresión coincide con la definición de densidad espectral ya que, cuando el orden Q tiende a infinito, se obtiene efectivamente la densidad espectral del proceso como la transformada de Fourier de su función de autocorrelación.

$$S(\omega) = \lim_{Q \rightarrow \infty} \frac{1}{\underline{S}^H \underline{S}} \underline{S}^H \underline{R} \underline{S} = \lim_{Q \rightarrow \infty} \frac{1}{Q} \underline{S}^H \underline{R} \underline{S} = r(0) + \lim_{Q \rightarrow \infty} \sum_{q=-Q}^Q r(q) \exp(jq\omega) \quad (\text{II.74})$$

$$\underline{S}^H = [1 \quad \exp(-j\omega) \quad \dots \quad \exp(-jQ\omega)]$$

Guardando en mente la expresión anterior, si se piensa de nuevo en el problema del estimador de la matriz de correlación formado por un único vector de dimensión N y olvidando lo comentado después, se puede razonar del siguiente modo: si estimar la matriz con un solo vector de N muestras produce problemas de varianza, lo lógico sería coger la realización del proceso, asumirlo ergódico y dividirlo en  $\lfloor N/Q \rfloor$  (donde  $\lfloor \cdot \rfloor$  indica entero más próximo) segmentos posiblemente solapados, y promediar los estimadores de cada realización para obtener el total. La formulación del nuevo estimador de correlación sería:

$$\hat{\underline{R}} = \frac{1}{M} \sum_{m=0}^{M-1} X_{mQ}(Q) X_{mQ}^H(Q) \quad (\text{II.75})$$

donde se hace evidente el promedio de todos los segmentos posibles al tomar M segmentos de longitud Q. La solidez de este estimador es indiscutible bajo el punto de vista de la densidad espectral ya que lo único que hace es limitar el valor de Q a un valor finito. Cuando esta matriz se introduce en la expresión de la densidad espectral se comprueba que conduce al estimador sesgado. Dicho de otro modo, la versión en términos de función de correlación asociada a esta matriz, usada de modo natural para la estimación de la densidad espectral, equivale a la transformada de Fourier del estimador sesgado.

$$\hat{S}(\omega) = \frac{1}{\hat{\underline{S}}^H \hat{\underline{S}}} \hat{\underline{S}}^H \hat{\underline{R}} \hat{\underline{S}} = \frac{1}{Q} \hat{\underline{S}}^H \hat{\underline{R}} \hat{\underline{S}} = \hat{r}(0) + \sum_{q=-Q}^Q \hat{r}(q) \exp(jq\omega) \quad (\text{II.76})$$

$$\text{siendo } \hat{r}(q) = \frac{1}{N} \sum_{n=0}^{N-|q|-1} x^*(n) x(n+q) \quad \text{para } q = -Q, \dots, Q$$

Como (II.75) es una matriz definida positiva la densidad espectral obtenida con su uso en el último término de (II.74) es siempre positiva, como corresponde a una densidad espectral. En otras palabras, al margen de manejar un conjunto de 2Q+1 valores finitos la transformada de Fourier de la secuencia formada por el estimador sesgado de -Q a +Q es siempre positiva, como ha de ocurrir en cualquier candidato a estimar una densidad espectral de potencia.

En resumen, si se requiere un estimador de la matriz de correlación se usará siempre (II.75). Si tan solo interesan los valores de la correlación se usará el estimador sesgado, en la certeza que el empleo de la matriz de correlación se corresponde con el estimador sesgado de la función de autocorrelación, a efectos de la estimación de la densidad espectral de potencia. Es preciso insistir que la matriz, como tal, es un estimador insesgado de la matriz de autocorrelación; es al usarla para calcular la densidad espectral, cuando aparece el estimador sesgado de autocorrelación como su transformada de Fourier inversa.

El estimador insesgado de autocorrelación está en desuso ya que su transformada de Fourier de un segmento del estimador no es necesariamente positiva. La causa de la anterior se hace evidente cuando se examina el valor esperado de ambos estimadores de densidad espectral para  $Q$  puntos,

$$\begin{aligned} \text{insesgado } E\{\hat{r}(q)\} &= \begin{cases} r(q); & |q| \leq Q \\ 0; & \text{resto} \end{cases} \\ \text{sesgado } E\{\hat{r}(q)\} &= \begin{cases} \left(1 - \frac{|q|}{Q}\right)r(q); & |q| \leq Q \\ 0; & \text{resto} \end{cases} \end{aligned} \quad (\text{II.77})$$

El primero se trunca con un rectángulo de duración  $2Q+1$  cuya transformada es a veces negativa; mientras que, en el segundo, se trunca con un triángulo de duración también  $2Q+1$  cuya transformada es siempre positiva. En los estimadores espectrales, la media del estimador es igual a la transformada de Fourier de la elección tomada en (II.77), es decir, la media del estimador espectral es la convolución del espectro correcto por la función que acompaña a  $r(q)$  en la formulación anterior. Como el primero es la multiplicación de la función de autocorrelación correcta por un pulso rectangular, su estimador será la convolución del espectro correcto con una señal no siempre positiva mientras que el segundo sí. Por esta razón y no otra, en el segundo caso se puede asegurar que el valor esperado del estimador espectral será siempre positivo mientras que en el primero no.

Lo expuesto en esta sección es más un contenido propio de análisis espectral, como habrá podido intuirse en su presentación. La razón de incluirla, dentro de un tema sobre estimación con un nivel de detalle elevado, es con el fin evidenciar los problemas en el uso de la correlación estimada y poner de relieve que, cuando el estimador no es una función lineal de las observaciones como es el caso de la autocorrelación, los enfoques óptimos ML, MAP, etc. conllevan una gran dificultad en su formulación. Aunque esta formulación es posible, es difícil encontrar en la literatura un tratamiento diferente del expuesto, basado únicamente o directamente en sesgo y varianza de propuestas ad-hoc de estimadores, tal y como se hizo en el caso de alisado de datos. Nótese, en cualquier caso, que la función de verosimilitud responde a la expresión que sigue:

$$\begin{aligned} f(\underline{X}_m(Q) / \underline{R}) &= \frac{1}{\pi^Q \det(\underline{R})} \exp\left(-\frac{1}{2} \underline{X}_m^H(Q) \underline{R}^{-1} \underline{X}_m(Q)\right) \quad \text{para } M = 1 \\ f(\underline{X}_m(N) / \underline{R}) &= \prod_{m=1}^M f(\underline{X}_{mQ}(Q) / \underline{R}) \quad \text{para } M \text{ segmentos} \end{aligned}$$

puediéndose formular el estimador ML de la matriz de autocorrelación correspondiente que resulta ser:

$$\hat{\underline{R}}_{ML} = \frac{1}{M} \sum_{n=0}^{M-1} \underline{X}_n \cdot \underline{X}_n^H$$

De nuevo, este ha de ser el que siempre se ha de emplear. Si se desea la función de autocorrelación, lo normal es sumar a lo largo de las diagonales la anterior matriz. Recuerde que esto último es equivalente al estimador sesgado pero es más rápido, cómodo y recomendable como método computacional para usar la autocorrelación en aplicaciones de procesamiento de señal. Note también que en el caso de que el número de segmentos sea inferior al orden de la matriz entonces el rango de la matriz estimada no será completo y su determinante sería cero y carecería de inversa. Puede demostrarse que en este caso, el estimador ML viene dado por los  $M(<Q)$  autovectores y autovalores no nulos de la matriz anterior según sigue:

$$\underline{\underline{\Theta}} = \frac{1}{M} \cdot \sum_{n=0}^{M-1} \underline{X}_n \cdot \underline{X}_n^H \quad \underline{\underline{\Theta}} \cdot \underline{e}_q = \lambda_q \cdot \underline{e}_q \quad q=1, Q \quad \text{con} \quad \lambda_q = 0, \forall q \geq M+1$$

$$\hat{\underline{\underline{R}}}_{ML}^{-1} = \sum_{p=1}^Q \lambda_p^{-1} \cdot \underline{e}_p \cdot \underline{e}_p^H \quad \lambda_p = \lambda_M, \forall p \geq M+1$$

Note que la matriz de correlación como tal, para evitar su carácter singular, se construye asignando a todos los autovalores cero, es decir a los Q-M, el autovalor mínimo obtenido. En cualquier caso, la situación anterior en proceso de muestras temporales es anómala y Q suele ser bastante mas pequeño que M. Este no es el caso en otros sistemas donde en lugar de tiempo se procesan muestras en espacio, código o frecuencia.

## II.10. MODELOS RACIONALES.

Hasta el momento la exposición se ha limitado a la mera caracterización de señales temporales vistas como realizaciones de procesos. A partir de las observaciones de una realización se ha expuesto los criterios ML y MAP para la estimación de parámetros. Incluso en estos casos (ver por ejemplo (II.41)), se ha hecho patente la necesidad de conocer la matriz de autocorrelación para la correcta estimación de parámetros. El hecho de que la matriz defina las distribuciones condicionales proporciona, de por si, un interés especial al apartado anterior, dedicado precisamente a la estimación de dicha matriz de autocorrelación.

Aunque no de modo general, muchos procesos pueden modelarse como la salida de un sistema lineal a cuya entrada se le aplica ruido blanco, máxime en el caso Gaussiano. Es decir, tal y como se indica en la Figura II.14, en muchos casos prácticos, se puede modelar el proceso como la respuesta  $x(n)$  de un sistema lineal, de respuesta impulsional  $h(n)$ , cuando se le aplica ruido blanco  $w(n)$  de potencia  $\sigma^2$  a su entrada. Una de las cuestiones fundamentales es: ¿Hasta qué punto se puede dar por cierto el modelo y cual seria su utilidad? Para responder a la pregunta anterior, es de gran ayuda conocer la autocorrelación del proceso  $x(n)$  o su densidad espectral de potencia, si efectivamente el modelo fuese correcto para el proceso observado.

Para estudiar la estructura que el modelo de la Figura II.11 da a la autocorrelación, se dividirá la presentación en tres tipos de sistemas lineales a saber: Modelo de solo ceros de orden P o MA(P) (Moving Average), modelo de solo polos de orden Q o AR(Q) (Auto Regressive) y modelo de ceros y polos o ARMA(P,Q).

La exposición seguirá este orden, que atiende básicamente a la complejidad estructural del modelo lineal del proceso. Aunque, como mas adelante se podrá observar, es el caso autoregresivo el que presenta mas ventajas de cara a estimar sus parámetros y a las cualidades del estimador espectral correspondiente.

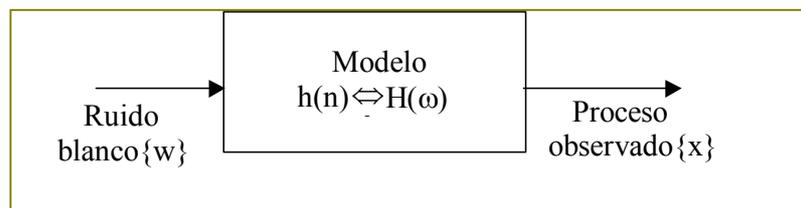


Figura II.14. Modelo lineal para el proceso  $\{x\}$  como respuesta de un sistema lineal al que se le aplica ruido blanco a su entrada.

Un modelo MA viene caracterizado la siguiente ecuación:

$$x(n) = \sum_{p=0}^P b(p)w(n-p) = b(n) * w(n) \quad (\text{II.78})$$

$$X(\omega) = B(\omega)W(\omega) \quad \text{siendo } B(\omega) = \sum_{p=0}^P b(p)\exp(-jp\omega)$$

Donde  $X(\omega)$  y  $W(\omega)$  son las transformadas de Fourier de  $x(n)$  y  $w(n)$  respectivamente. Este tipo de modelos aparece en propagación multicamino, en los que la señal observada aparece como la denominada interferencia inter-símbolo en comunicaciones (caso no gaussiano); o en acústica, en espacios sin reverberación y en las que las llegadas rápidas ('early arrivals') obedecen a este modelo (véase la figura II.12).

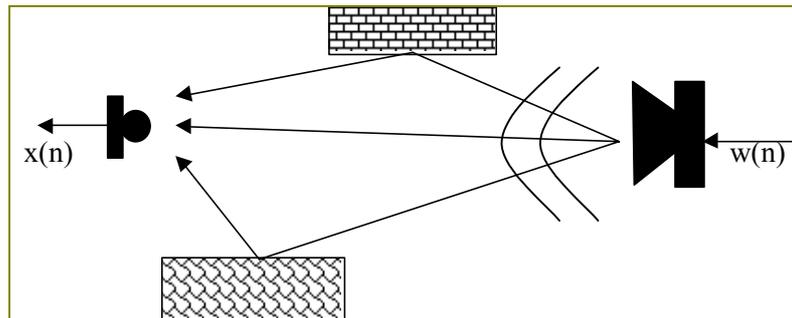


Figura II.15. Propagación multicamino esquematizando un modelo MA de orden 3.

Para calcular la función de autocorrelación de este tipo de procesos, basta multiplicar ambos lados de la ecuación de filtrado por  $x(n+m)$ , o  $x(n-m)$  ya que la correlación es hermitica, y tomar valor esperado. Al hacerlo, teniendo en cuenta que  $w(n)$  es ruido blanco, se obtiene:

$$r_x(m) = \sum_{p=0}^P b(p)E\{w(n-p)x^*(n-m)\} = \sum_{p=0}^P \sum_{s=0}^P b(p)b^*(s)E\{w(n-p)w^*(n-s-m)\} =$$

$$= \begin{cases} \text{El valor esperado} \\ \text{es cero si } p \neq m+s \end{cases} = \sigma_w^2 \sum_{s=0}^P b^*(s)b(m+s) \quad (\text{II.79})$$

Es decir, la autocorrelación observada es proporcional a la autocorrelación de los coeficientes del filtro. Otra manera de observar la relación anterior es usar el hecho de que la densidad espectral de la salida de un sistema lineal es la de la entrada (constante en nuestro caso) por el modulo al cuadrado de la función de transferencia del sistema lineal.

$$S_x(\omega) = \sigma_w^2 |H(\omega)|^2 = \sigma_w^2 |B(\omega)|^2 = \sigma_w^2 B(z)B^*(1/z^*) \Big|_{z=\exp(j\omega T)} \quad (\text{II.80})$$

De su transformada inversa puede obtenerse (II.79) directamente.

Centrándose en el problema acústico de la figura II.15, se puede estimar la autocorrelación mediante el estimador sesgado, y a partir de ahí los coeficientes  $b(\cdot)$  del modelo, que no son más que los respectivos coeficientes de reflexión de los paneles donde han tenido lugar las reflexiones. En problemas de comunicaciones con propagación multicamino tiene gran interés el denominado problema inverso, es decir, la identificación del sistema  $B(z)$  y su compensación mediante otro sistema lineal. Aparentemente el problema en los dos casos parece sencillo. No obstante, al intentar resolverlo, se hace evidente que solo se puede conocer el modulo de  $B(z)$  pero no su fase, si únicamente usa la autocorrelación de la salida. Tan solo información adicional, como que se trata de un sistema de fase mínima o mínimo retardo (todas las

raíces de  $B(z)$  dentro del círculo unidad, la respuesta impulsional en algún valor, etc.) pueden resolver la indeterminación de la fase que el problema inverso representa. De todos modos, un sistema que compensase la sala (en el problema acústico) o el canal de propagación (en el caso de comunicaciones) conllevaría implementar el filtro inverso de  $B(z)$  con lo que la solución de fase mínima se haría obligada.

Se pasará a continuación al caso de los denominados modelos de solo polos o autoregresivos. Es obvio que se considera un modelo estable, es decir, dichos polos se han de encontrar estrictamente dentro del círculo unidad. En el caso autoregresivo, la  $H(z)$  del modelo es la que se indica en (II.81), donde se ha normalizado la  $h(0)$  a la unidad sin que el modelo pierda su carácter general:

$$H(z) = \frac{1}{1 + \sum_{q=1}^Q a(q)z^{-q}} = 1/A(z) \quad (\text{II.81})$$

La señal del proceso y su espectro de potencia son:

$$x(n) = w(n) - \sum_{q=1}^Q a(q)x(n-q)$$

$$S(\omega) = \frac{\sigma^2}{|A(\omega)|^2} = \frac{\sigma^2}{A(z)A^*(1/z^*)} \Big|_{z=\exp(j\omega T)} \quad (\text{II.82})$$

Respecto a la función de autocorrelación, ésta presenta una recursión que es fácil de obtener de la multiplicación en ambos lados de la señal  $x(n)$  por  $x(n-m)$  y tomar valor esperado. Nótese que hay dos situaciones diferentes: cuando  $m$  es cero existe correlación entre  $w$  y  $x$  mientras que para  $m$  distinto de cero no.

$$r(m) = \sigma_w^2 \delta(m) - \sum_{q=1}^Q a(q)r(m-q) \quad (\text{II.83})$$

Esta relación es interesante pues muestra que infinitos segmentos de  $Q+1$  muestras de correlación son anuladas al ponderarlas con la secuencia de coeficientes. Esta relación es fácil de observar, escribiendo la formula anterior en forma matricial:

$$\begin{bmatrix} r(0) & r(-1) & \cdots & r(-Q) \\ r(1) & r(0) & \cdots & r(-Q+1) \\ \vdots & \vdots & \ddots & \vdots \\ r(m) & r(m-1) & \cdots & r(m-Q) \\ r(m+1) & r(m) & \cdots & r(m-Q+1) \\ \vdots & \vdots & & \vdots \end{bmatrix} \begin{bmatrix} 1 \\ a(1) \\ \vdots \\ a(Q) \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \end{bmatrix} \quad (\text{II.84})$$

De hecho, es como si el filtro FIR, formado por los coeficientes de  $A(z)$ , al procesar la función de correlación, desde el origen, produjese una delta discreta. De otra manera, cualquier valor de la función de correlación de índice mayor o igual a  $Q$  puede expresarse como una combinación de los valores anteriores. Esto último implica que el modelo AR(Q) tiene una autocorrelación de duración infinita pero que, conocidos los primeros  $Q$  valores y los coeficientes del modelo es directamente extrapolable.

Si ahora se pretende resolver el problema inverso, puede comprobarse que este tiene una solución fácil y única. Estimados los primeros  $Q$  valores y tomando las ecuaciones  $m=1$  a la  $m=Q$ , en orden vertical descendente, en (II.84), se dispondrá de  $Q$  ecuaciones con  $Q$  incógnitas que permiten encontrar, sin ninguna indeterminación, los coeficientes del modelo. Una vez obtenidos los coeficientes,

usando la ecuación  $m=0$  se obtendrá la potencia del ruido. No es de extrañar que se pueda resolver el problema inverso, el propio modelo garantiza que  $A(z)$  es de fase mínima; de otro modo no sería viable.

Sin embargo, en la práctica los errores de estimación de la matriz de correlación pueden llegar a comprometer la propiedad de fase mínima de  $A(z)$ . Sin llegar a demostrarlo, baste asegurar que el uso del estimador sesgado de la correlación garantizará el carácter definido positivo de la matriz de correlación y la propiedad de fase mínima.

Los modelos AR aparecen también en acústica de salas cuando, una vez pasadas los ecos cercanos o early arrivals, comienza a recogerse la reverberación. De hecho todo sistema reverberante de orden  $Q$  produce, al introducirle ruido blanco, exactamente un modelo AR del mismo orden. El tema de filtrado de Wiener profundizará más en este aspecto, por lo que los detalles adicionales se relegan a ese momento.

Por último, quedan aquellos procesos que pueden modelarse con un modelo de ceros y polos, por ejemplo medida de sonido reverberante con ruido aditivo. La función de transferencia pasa a ser (II.85):

$$H(z) = \frac{\sum_{p=0}^P b(p)z^{-p}}{1 + \sum_{q=1}^Q a(q)z^{-q}} = B(z) / A(z) \quad (\text{II.85})$$

la salida como (II.86):

$$x(n) = \sum_{p=0}^P b(p)w(n-p) - \sum_{q=1}^Q a(q)x(n-q) \quad (\text{II.86})$$

y es fácil verificar que la respuesta impulsional  $h(n)$  cumple la siguiente relación:

$$\begin{bmatrix} b(0) \\ b(1) \\ \vdots \\ b(P) \end{bmatrix} = \begin{bmatrix} h(0) & 0 & \cdots & 0 \\ h(1) & h(0) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ h(P) & h(P-1) & \cdots & h(P-Q) \end{bmatrix} \begin{bmatrix} 1 \\ a(1) \\ \vdots \\ a(Q) \end{bmatrix} \quad (\text{II.87})$$

Con respecto a la correlación, para derivar una manera eficiente de calcularla es necesario expresarla en términos de su transformada  $z$ . La transformada  $z$  de la correlación, al ser la del proceso de entrada una delta será igual a:

$$R(z) = \sigma_w^2 H(z)H^*(1/z^*) \quad \text{o bien} \quad S(\omega) = \sigma_w^2 H(\omega)H^*(\omega) \quad (\text{II.88})$$

Sustituyendo  $H(z)$  en función de los polinomios numerador y denominador,

$$R(z) = \sigma_w^2 \frac{B(z)}{A(z)} H^*(1/z^*) \quad \text{o bien} \quad R(z)A(z) = \sigma_w^2 H^*(1/z^*)B(z) \quad (\text{II.89})$$

lo cual implica que la convolución de la autocorrelación del proceso ARMA con los coeficientes del denominador es igual a la convolución de su respuesta impulsional anti-causal ( $h(-n)$ ) con los coeficientes del numerador. Esta propiedad puede escribirse matricialmente como se indica en (II.90a) y es extremadamente útil para dilucidar las posibilidades de resolver el problema inverso.

$$\begin{bmatrix} r(0) & r(-1) & \dots & r(-Q) \\ r(1) & r(0) & \dots & r(-Q+1) \\ \vdots & \vdots & \ddots & \vdots \\ r(P) & r(P-1) & \dots & r(P-Q) \\ \vdots & \vdots & \ddots & \vdots \\ r(P+1) & r(P) & \dots & r(P+1-Q) \\ \vdots & \vdots & \ddots & \vdots \\ r(P+Q+1) & r(P+Q) & \dots & r(P+1) \\ \vdots & \vdots & & \vdots \end{bmatrix} \begin{bmatrix} 1 \\ a(1) \\ \vdots \\ a(Q) \end{bmatrix} = \begin{bmatrix} h(0) & h(1) & \dots & h(P) \\ 0 & h(0) & \dots & h(P-1) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & h(0) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \end{bmatrix} \begin{bmatrix} b(0) \\ \vdots \\ b(P) \end{bmatrix} \quad (\text{II.90a})$$

De la observación de las ecuaciones anteriores, es evidente que la determinación de los parámetros del modelo puede realizarse usando (a) únicamente la respuesta impulsional, (b) conjuntamente la respuesta impulsional y la función de correlación, o (c) únicamente la función de correlación. Nótese que tomando las ecuaciones P+1 hasta P+Q de (II.90), se obtendrían Q ecuaciones con Q incógnitas que permitirían calcular los coeficientes del denominador. Estas ecuaciones se denominan de Yule-Walker extendidas, por su similitud con las de los modelos AR. Una vez calculados los coeficientes del denominador, si se conocen P puntos de la respuesta impulsional, usando (II.87) puede determinarse el numerador. Si se desconoce la respuesta impulsional, filtrando la señal original con el filtro A(z) se estaría ante un proceso MA, remitiéndose a lo expuesto para estos previamente.

Por último, si se dispone de la respuesta impulsional solamente, es posible reordenar (II.87) de la siguiente forma (si Q≥P) para obtener los parámetros de numerador y denominador:

$$\begin{bmatrix} \vdots & 0 & 0 & 0 & \dots & 0 \\ \vdots & h(0) & 0 & 0 & \dots & 0 \\ \vdots & h(1) & h(0) & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & & 0 \\ \vdots & h(P-1) & h(P-2) & h(P-3) & \dots & 0 \\ \hline \vdots & h(P) & h(P-1) & h(P-2) & \dots & h(P-Q+1) \\ \vdots & h(P+1) & h(P) & h(P-1) & \dots & h(P-Q+2) \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \vdots & h(Q+P-1) & h(Q+P-2) & h(Q+P-3) & \dots & h(P) \end{bmatrix} \begin{bmatrix} b(0) \\ b(1) \\ \vdots \\ b(P) \\ a(1) \\ \vdots \\ a(Q) \end{bmatrix} = \begin{bmatrix} h(0) \\ h(1) \\ \vdots \\ h(P) \\ h(P+1) \\ \vdots \\ h(Q+P) \end{bmatrix} \quad (\text{II.90b})$$

Nótese que esta última ecuación entraña el uso de muestras de la respuesta impulsional de índice muy elevado, las cuales no siempre están disponibles o bien están muy contaminadas por ruido, por lo cual es más habitual recurrir a (II.90a). Volviendo al problema de acústica en salas, la medida de la respuesta impulsional adolece del problema de que pasado un cierto intervalo de tiempo, la reverberación y el ruido predominan sobre los valores de h(n) que son, habitualmente de energía decreciente. Por lo anterior, es solo viable el asumir que se dispone de una buena estimación de la h(n) para los primeros instantes. Por otro lado, la medida con ruido permite conocer bien los primeros valores de autocorrelación. En definitiva, la solución del modelo con ambos tipos de conocimiento es muy útil (Figura II.16 ).

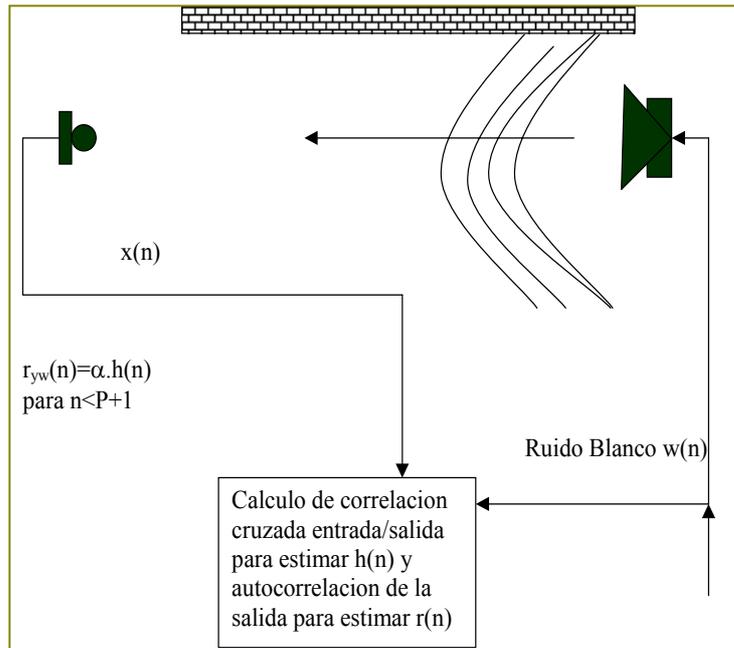


Figura II.16. Medida de  $h(n)$  y de  $r(n)$  en sistemas de propagación.

Al margen de los problemas de determinación de la fase en la estimación MA, las ecuaciones de Y-W extendidas usan términos de la función de autocorrelación muy altos y por lo tanto de una gran varianza con lo que su resolución a partir de estimadores de la autocorrelación no son, habitualmente, muy fiables. Existen técnicas de SVD o descomposición en valores principales (ver capítulo I) para mitigar el efecto del ruido de estimación, con todo, el procedimiento entraña problemas de la calidad resultante..

Todos los problemas mencionados desaparecen si se conocen los  $P$  valores de la respuesta impulsional. En este caso la solución del problema es exacta, tanto para numerador como para denominador, y todo sin requerir valores altos de la función de correlación ni de respuesta impulsional. Este procedimiento se basa en tomar las  $Q$  primeras ecuaciones en (II.90.a) y escribirlas del modo que sigue:

$$\underline{\underline{Ra}} = \underline{\underline{H}}^T \underline{\underline{b}} \tag{II.91}$$

Teniendo en cuenta la relación (II.87), al sustituirla en la anterior, la solución de los coeficientes del denominador pasa a ser el autovector generalizado de autovalor más próximo a la unidad del problema planteado en:

$$\underline{\underline{Ra}} = \underline{\underline{H}}^T \underline{\underline{Ha}} \text{ equivalente a } \underline{\underline{R\hat{a}}} = \lambda \underline{\underline{H}}^T \underline{\underline{H\hat{a}}} \tag{II.92}$$

Los coeficientes del denominador llevados a (II.87) permitirían obtener el numerador, solucionando completamente el problema inverso. Nótese que el polinomio que tiene por coeficientes el autovector obtenido desde (II.92) no tiene por que ser de fase mínima; no obstante, a efectos de obtener un estimador espectral no sería un problema y, por otra parte, encontrar su versión de fase mínima es posible determinando los ceros del polinomio que están fuera del círculo unidad.

De la presentación anterior es de destacar como la mezcla de información de orden uno (la respuesta impulsional) y de orden 2 (la función de correlación) simultáneamente es la única manera eficaz de determinar con precisión ambos conjuntos de coeficientes.

## II.12. CONCLUSIONES

El tema ha comenzado con una revisión del concepto de histograma y de la caracterización de variables aleatorias, como introducción al apartado de procesos estocásticos. En este caso la caracterización se torna muy compleja y se recurre a los conceptos de estacionaridad y ergodicidad. Es de destacar que este último, aparentemente teórico, conlleva toda la formalidad a la que se puede recurrir al tratar de reducir varianza en estimación. Segmentación y ergodicidad son evidencian como conceptos paralelos. Se cierra esta breve revisión con los efectos de sistemas lineales en procesos, y su importancia en la evaluación de la respuesta de sistemas lineales, en lugar de la teórica medida como respuesta a un impulso.

El problema de estimación de parámetros se aborda desde los principios de información y probabilidad. La diferencia entre estimación MAP y ML se resume en términos de conocimiento a priori del parámetro a estimar. A continuación se revisa como puede valorarse la calidad del estimador en términos de sesgo y varianza. De hecho, los criterios de calidad pueden utilizarse directamente en el diseño del estimador cuando la arquitectura de este ha sido elegida previamente. El caso de alisado de datos es analizado bajo esta perspectivas como exponente de la calidad de estimadores ad-hoc.

Dentro de la estimación de la función de correlación, se han descrito las alternativas y las, hasta cierto punto, sorprendentes diferencias entre la estimación de la matriz y de la función de correlación. Es de destacar que, el estimador de la matriz es básicamente insesgado. Es cuando el estimador insesgado de la matriz de autocorrelación se emplea para la estimación del espectro cuando aparece el estimador sesgado para la función de autocorrelación. Se ha de insistir que es mucho más interesante retener la estimación de la matriz que el de la función, de este modo, el estimador insesgado de esta es tan solo un resultado lateral del primer y más importante problema.

Por último, se ha presentado el impacto que el modelo lineal de un proceso tiene sobre la caracterización de éste. La solución del problema inverso, es decir, dadas medidas del proceso encontrar los parámetros del modelo se ha analizado para los tres casos por separado MA, AR y ARMA. Salvo en el caso AR, se ha hecho evidente la necesidad de información adicional, para resolver sin indeterminación el problema inverso. En concreto, se ha puesto en evidencia la potencia de disponer del conocimiento parcial de la respuesta impulsional del modelo, junto al habitual de correlación.

Todo lo descrito en este tema es de una importancia crucial para desarrollar los conceptos de estimación espectral que aparecen el siguiente tema.

## II.13. EJERCICIOS.

- 1.- Proponga un estimador de la pdf de una v.a. vía una estimación previa de la función característica. Analice el sesgo y la varianza del estimador propuesto.
- 2.- Indique qué sistema no lineal permite pasar de una distribución exponencial par (Laplace) a una distribución uniforme.
- 3.- Calcule la expresión de un estimador ML de la media, de orden 2 (dos muestras  $x(n)$  y  $x(n-1)$ ) de un proceso, sabiendo que su matriz de autocovarianza es:

$$\begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix}$$

indicando el impacto del parámetro alfa en el estimador y que valores de este distancian más el estimador óptimo de la media aritmética.

- 4.- Proponga un estimador de la correlación cruzada y reconstruya el proceso de discusión realizado en el texto para el caso de autocorrelación. (alternativas, sesgo, varianza y estimador de la densidad espectral cruzada).

- 5.- Un proceso estocástico puede formarse a partir de un modelo de señal determinista en el que uno o más de sus parámetros se convierte en aleatorio. Imaginando la señal  $x(\theta, t)$  al convertir el parámetro  $\theta$  en

una v.a. se da lugar al proceso  $\{x\}$ . Convertir el parámetro en una v.a. quiere decir que las realizaciones se obtiene a partir de diferentes valores del parámetro  $\theta$ . Cada una de estas, o contemplado el proceso a realización fija, será una señal determinista. Muestre que en este caso la autocorrelación definida en el texto pasa a ser la siguiente:

$$r_x(t, \tau) = \int x(\theta, t)x(\theta, t + \tau)f(\theta)d\theta$$

6.- Considere la señal  $x(t) = a \cos(\omega_0 t + \theta)$ , calcule la autocorrelación del proceso y encuentre la/s condición/es para que éste sea estacionario en los siguientes casos.

- $a$  es una v.a. de media  $A_0$  y varianza  $\sigma_a^2$ , al mismo tiempo  $\theta$  es una v.a uniforme entre 0 y  $\pi$ .
- La amplitud es determinista y la fase es uniforme.
- La fase es uniforme y la frecuencia es una variable aleatoria de distribución  $S_x(\omega)$  par alrededor del origen. Relacione el espectro de potencia obtenido con el ancho de banda para una modulación de FM de banda ancha.

7.- Dados los valores de autocorrelación siguientes:

$$r_x(0) = 1 \quad r_x(1) = 0.5 \quad r_x(2) = 0.1$$

y sabiendo que se trata de un modelo AR(2) calcular la expresión de la función de autocorrelación  $r_x(3)$ .

8.- Demuestre que la observación de un proceso AR en ruido aditivo siempre produce un modelo ARMA. Especifique el tipo de modelo obtenido en el caso de que el ruido aditivo sea blanco.

9.- Considerando que la suma de Q sinusoides complejas es la solución de una ecuación en diferencias de orden Q, si las Q señales forman  $s(n)$ , entonces se verifica que

$$\sum_{q=0}^Q a(q)s(n-q) = 0$$

si ahora se forma la señal  $x(n)=s(n)+w(n)$  donde el segundo es ruido gaussiano y blanco. Considere que la señal  $s(n)$  son dos sinusoides a frecuencias normalizadas de 0.1 y 0.25 fase cero, con una snr de 3 y 10dB respectivamente

- ¿Cuál es la densidad espectral del proceso  $x(n)$ ?
- Demuestre que el proceso  $x(n)$  tiene una estructura ARMA, que se obtiene al sustituir  $s(n)=x(n)-w(n)$  en la ecuación en diferencias. ¡La densidad espectral del proceso ARMA calculada con el modelo es constante a todas las frecuencias!
- Discuta la diferencia de las respuestas a y b teniendo presente que el proceso no es estacionario y que lo que ha calculado con el apartado b es en realidad la transformada de Fourier de la autocovarianza del proceso.
- Convierta ahora el proceso en estacionario y compruebe que en este caso la ecuación en diferencias ya no es valida y por lo tanto el proceso resultante no responde a un modelo ARMA.

10.- Demuestre que la mayor parte de las modulaciones lineales son procesos no estacionarios. Indique en consecuencia como se llega a la densidad espectral de estas. Pruebe que la situación es la misma para una señal digital

$$x(t) = \sum_{k=-\infty}^{\infty} a(k)p\left(t - \frac{k}{r}\right) \quad \text{con } a(k) \text{ igual a } +/-A \text{ volt.}$$

11.- Determine el impacto del producto duración ancho de banda en la estimación de la autocorrelación de un proceso que es ruido blanco cuando esta se realiza sobre un segmento de T seg. Razone porqué se denomina al producto TB grados de libertad en relación con la interpretación de ergodicidad incluida en el texto.

12.- Dado un vector  $\underline{X}_n$  de Q muestras con la siguiente estructura:

$$\underline{X}_n = a \cdot \underline{S} + \underline{w}_n$$

donde a es la envolvente compleja,  $\underline{S}$  es igual al vector  $[1 e^{jw} \dots e^{j(Q-1)w}]^T$  y  $\underline{w}_n$  es un vector de ruido blanco de potencia en cada una de sus componentes igual a  $\sigma^2$ . La verosimilitud del vector dato con respecto a los parámetros de envolvente, frecuencia y potencia de ruido viene dada por:

$$\Lambda(a, \underline{S}, \sigma^2) = \text{Ln}\{\text{pr}(\underline{X}_n / a, \underline{S}, \sigma^2)\} = K_o - Q \cdot \text{Ln}(\sigma^2) - \frac{1}{\sigma^2} \cdot [(\underline{X}_n - a\underline{S})^H (\underline{X}_n - a\underline{S})]$$

siendo Ln(.) el logaritmo neperiano y pr(.) densidad de probabilidad.

a) ¿Cuál es el estimador de máxima verosimilitud de a observado el vector  $\underline{X}_n$   $a_{ML}(n)$ ?

Sustituya la expresión del estimador en  $\Lambda(a, \underline{S}, \sigma^2)$  para obtener la verosimilitud que le permita estimar otro parámetro.

b) Compruebe que la nueva verosimilitud puede escribirse como sigue:

$$\Lambda(\underline{S}, \sigma^2) = K_o - Q \cdot \text{Ln}(\sigma^2) - \frac{1}{\sigma^2} \cdot \text{Traza}[\underline{P} \cdot \underline{X}_n \underline{X}_n^H]$$

siendo  $\underline{P}$  igual a  $\left( \underline{I} - \frac{\underline{S}\underline{S}^H}{Q} \right)$ , sabiendo que  $\underline{P} \cdot \underline{P}^H = \underline{I}$  (la matriz identidad), y usando que para

dos vectores cualesquiera se verifica que  $\underline{v}^H \cdot \underline{z} = \text{Traza}[\underline{z} \cdot \underline{v}^H]$ .

c) ¿Cuál sería  $\Lambda(\underline{S}, \sigma^2)$  en el caso de disponer de N vectores de datos  $\underline{X}_n$  independientes entre si.? Al escribir la expresión de la verosimilitud use que:

$$\underline{R} = \frac{1}{N} \cdot \sum_{n=0}^{N-1} \underline{X}_n \cdot \underline{X}_n^H$$

Recuerde que el operador traza(.) es lineal y puede intercambiarse el orden con sumatorios o valores esperados.

d) Demuestre que

$$E(\underline{R}) = |a|^2 \cdot \underline{S} \cdot \underline{S}^H + \sigma^2 \cdot \underline{I}$$

e) Demuestre que:

$$\text{Traza}[\underline{P} \cdot \underline{R}] = \text{Traza}[\underline{R}] - \frac{\underline{S}^H \cdot \underline{R} \cdot \underline{S}}{Q}$$

Este factor sale repetidamente en lo que sigue. Denómelo  $\phi$  en los siguientes apartados.

f) Calcule el estimador ML de la potencia de ruido  $\sigma_{ML}^2$  usando la expresión de  $\Lambda(\underline{S}, \sigma^2)$ .

- g) Compruebe que el estimador de la potencia del apartado anterior esta sesgado según la expresión siguiente:

$$E[\sigma_{ML}^2] = \frac{Q-1}{Q} \cdot \sigma^2$$

- h) Compruebe que utilizando el estimador ML del nivel de ruido la verosimilitud resultante es:

$$\Lambda(\underline{S}) = K_1 - Q \cdot N \cdot [\ln(\sigma_{ML}^2) - 1]$$

y demuestre que la estimación óptima de frecuencia coincide con el máximo obtenido del estimador WOSA (Ver Capitulo III), con ventana rectangular de los datos.

13.- Considere una realización de un proceso gaussiano de media nula y duración 2NT segundos, siendo T el periodo de muestreo de la señal original.

Para estimar la potencia se propone el siguiente estimador:

$$\hat{P}_x = \frac{1}{2N-1} \cdot \sum_{n=-N+1}^{N-1} x^2(n)$$

- a) Demuestre que el estimador es insesgado.

Recordando que en un proceso gaussiano se verifica que:

$$E[x_1 \cdot x_2 \cdot x_3 \cdot x_4] = E[x_1 \cdot x_2] E[x_3 \cdot x_4] + E[x_1 \cdot x_3] E[x_2 \cdot x_4] + E[x_1 \cdot x_4] E[x_3 \cdot x_2]$$

- b) Demuestre que la varianza del estimador viene dada por:

$$\text{var}(\hat{P}_x) = \frac{2}{2N-1} \cdot \sum_{m=-N+1}^{m=N-1} \left(1 - \frac{|m|}{2N-1}\right) r_x^2(m)$$

siendo  $r_x(m)$  la función de autocorrelación del proceso.

Considerando que el numero de muestras es grande ( $N \rightarrow \infty$ ) y que el espectro del proceso es de forma rectangular y ancho de banda B,

- c) Demuestre que la varianza del proceso tiende a

$$\text{var}(\hat{P}_x) \rightarrow \frac{P_x^2}{B \cdot (2N-1)}$$

- d) Justifique porque se denomina a  $B(2N-1)$  “numero de grados de libertad” coincidiendo con el numero de medidas independientes que se pueden realizar en un proceso de estimación.

Considere ahora que dispone de las muestras de  $x(n)$  en el intervalo 0,N-1. Una alternativa al estimador anterior es la que se expresa a continuación:

$$\tilde{P}_x(n) = \beta \cdot \tilde{P}_x(n-1) + \alpha \cdot x^2(n)$$

- e) Calcule la relación entre  $\alpha$  y  $\beta$  para que el estimador sea insesgado.

- f) Compruebe que para que este ultimo estimador sea idéntico al de los apartados anteriores, es necesario que  $\beta$  dependa de n y encuentre su expresión, manteniendo el estimador insesgado.

14.- Una de las distribuciones mas extrañas es la distribución de Cauchy,

$$\Pr(a) = \frac{1}{\pi} \frac{1}{1+a^2}$$

debido a que es una distribución de media nula, varianza infinita y la suma de dos variables aleatorias que son Cauchy también es Cauchy, es decir, no verifica el teorema central del limite. Demuestre estas propiedades tan anómalas de esta distribución.Cuál es su función característica?

15.- Demuestre la expresión del valor esperado del periodograma probando que salvo en el caso de que se trate de ruido blanco, este siempre será un estimador sesgado.

16.- Indique que es lo incorrecto en el razonamiento que sigue:

“ a) Sea  $\{g\}$  ruido blanco aplicado a un sistema lineal  $H(w)$  que produce el proceso  $\{x\}$ . La DFT de un record de  $\{x\}$  es igual a  $H(w)$  por la DFT del record correspondiente  $g(n)$ . b) El periodograma de  $\{x\}$  sería igual al de  $\{g\}$  multiplicado por  $|H(w)|^2$ . c) Entonces el valor esperado del periodograma de  $\{x\}$  es  $|H(w)|^2$  por el valor esperado del periodograma de  $\{g\}$ . d) Como el valor esperado del periodograma de  $\{g\}$  es insesgado por ser este ruido blanco, entonces también lo es el de  $\{x\}$ ”.

17.- Se dispone de dos procesos gaussianos  $x(n)$  e  $y(n)$  entre los que existe una relación. El presente ejercicio resuelve el problema de dado  $y(n)$  obtener una estimación de  $x(n)$ . Para ello ha de resolver los apartados que siguen.

En primer lugar note que la probabilidad conjunta de  $x$  e  $y$  se puede escribir a partir de la definición de un vector  $\underline{z}$  según se indica a continuación.

$$\underline{z} = \begin{bmatrix} x \\ y \end{bmatrix} \quad E(\underline{z}) = \begin{bmatrix} m_x \\ m_y \end{bmatrix} \quad R_{\underline{z}} = \begin{bmatrix} r_{xx} & r_{xy} \\ r_{yx} & r_{yy} \end{bmatrix}$$

a) Escriba, salvo constantes, la función de distribución  $\Pr(x,y)$  (idéntica a la de  $\Pr(\underline{z})$ ).

A continuación se expresa una forma de escribir la inversa de la matriz de covarianza de  $\underline{z}$ ,

$$R_{\underline{z}}^{-1} = \begin{bmatrix} 1 & 0 \\ -r_y^{-1} \cdot r_{yx} & 1 \end{bmatrix} \begin{bmatrix} (r_x - r_{xy} \cdot r_y^{-1} \cdot r_{yx})^{-1} & 0 \\ 0 & r_y^{-1} \end{bmatrix} \begin{bmatrix} 1 & -r_y^{-1} \cdot r_{xy} \\ 0 & 1 \end{bmatrix} = \underline{\underline{G}} \cdot \underline{\underline{D}} \cdot \underline{\underline{G}}^H$$

es decir, se trata de una forma de diagonalizar la inversa de la matriz de covarianza.

b) Calcule las dos componentes  $x_1$  y  $x_2$  del siguiente producto  $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \underline{\underline{G}}^H \cdot (\underline{z} - E(\underline{z}))$

c) Muestre ahora, con los resultados de los dos apartados anteriores, que el exponente de la probabilidad conjunta puede escribirse como dos sumandos

$$(x_1)^2 \cdot (r_x - r_{xy} \cdot r_y^{-1} \cdot r_{yx})^{-1} + (x_2)^2 \cdot r_y^{-1}$$

siendo  $x_2$  igual a  $y - m_y$ .

En definitiva ahora esta en disposición de calcular, a partir de la distribución conjunta  $\Pr(x,y)$  y de  $\Pr(y)$ , la condicional de  $x$  dado  $y$  ( $\Pr(x/y) = \Pr(x,y)/\Pr(y)$ ). Con esta función  $\Pr(x/y)$ ,

d) calcule cual es el estimador de media condicional de  $x$  dado  $y$ .

e) cual es la varianza de la estimación anterior.

f) Pruebe, en general, que el estimador de media condicional es el de menor error cuadrático medio

g) Cuales serian las expresiones del estimador del apartado (d) y (e) si en lugar de escalares hubiesen sido dos vectores.

18.- Un proceso  $x(n)$  esta generado al aplicar ruido blanco al filtro siguiente:

$$H(z) = 1 - b(1) \cdot z^{-1}$$

a) Indique como puede estimar  $H(z)$  a partir de  $N$  muestras del proceso  $x(n)$ , sabiendo que el modelo es de fase mínima.

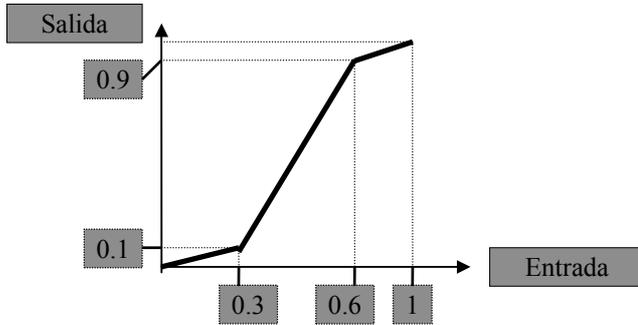
Si el sistema pasa a ser

$$H(z) = \frac{1}{1 + a(1)z^{-1} + a(2)z^{-2}}$$

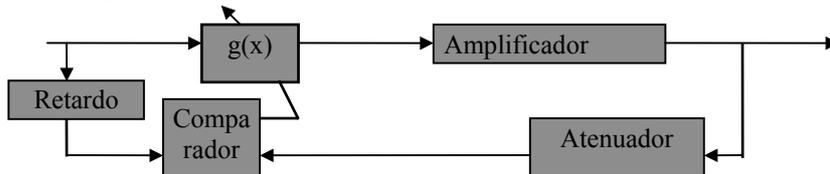
- b) Indique como obtendría los coeficientes también a partir de N muestras del proceso x(n).
- c) Responda a la misma pregunta caso de que el sistema fuese

$$H(z) = \frac{1 - b(1).z^{-1}}{1 + a(1)z^{-1} + a(2).z^{-2}}$$

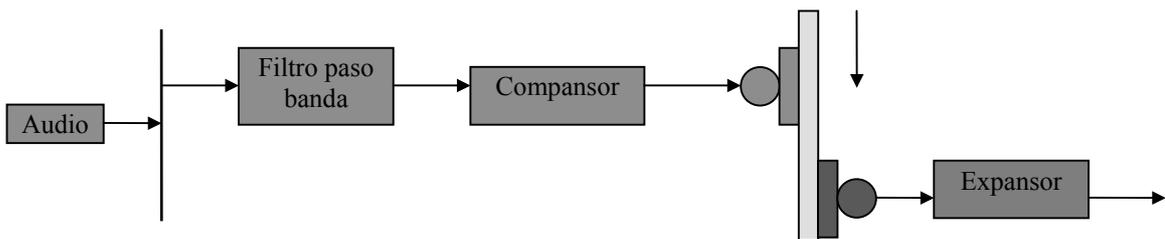
19.- Considere un amplificador de potencia, cuya respuesta entrada-salida es la siguiente:



- a) Cual es la función de distribución a la salida del amplificador cuando la entrada es uniforme? Para compensar la distorsión, tanto de corte como de saturación, se recurre al siguiente esquema:



- b) Cual ha de ser g(x) para que el amplificador entregue a la salida una señal libre de distorsión?
- c) Si la entrada es uniforme, ¿Cual es la distribución de z en el esquema anterior?
- d) Cual es la misión del atenuador y del retardo?
- e) En el caso de sistemas de grabación, se pretende, en cada banda de frecuencia, compensar la distorsión del proceso de magnetización. Al igual que en convertidores A/D, el objetivo se consigue parcialmente, persiguiendo que la distribución de los valores de corriente al cabezal sea uniforme. Tomando la distribución de la señal de audio de tipo exponencial simétrica (la misma a la salida del filtro paso-banda), indique las expresiones exactas del compansor y el expansor. (el sistema Dolby A utiliza tres bandas de frecuencia mínimo).



20.- Sea  $x(n), x(n-1), x(n-2)$  un segmento de tres valores de las muestras de una señal  $x(t)$  muestreada a 10 KHz. La señal  $x(t)$  es paso banda compleja y esta formada por una senoide de frecuencia 2KHz de amplitud y fase desconocidas; además contiene ruido aditivo gaussiano de media nula y correlación  $r_n(0)=3, r_n(1)=-1, r_n(2)=1$  (¿El ruido continuo tendría los mismos valores en  $0, T$  y  $2T$  ? Es decir, la correlación de las muestras difiere o no de la correlación muestreada?). Denominando  $b$  a la envolvente compleja de la senoide,

- a) ¿Cual es la expresión de  $\Pr(b/x(n), x(n-1), x(n-2))$  asumiendo que la distribución de  $b$  es gaussiana de media  $b_m$  y varianza  $\sigma_m$ ?
- b) Cual es el estimador MAP de  $b$ ?
- c) Cual es el estimador ML de  $b$ ? Discuta las diferencias con el anterior.

- d) Calcule numéricamente el sesgo y la varianza del estimador ML. Compruebe que la varianza es mínima al coincidir con el límite de Cramer-Rao
- e) Compruebe que el estimador ML es consistente (considere que el ruido es blanco) e indique las ventajas/desventajas de coger un segmento mayor de tres muestras.
- Suponga que conoce la presencia de otra senoide de frecuencia 4 KHz amplitud 1 Vol. y fase cero. Como cambia el estimador ML para este caso?. Comente el interés de estimación espectral en este caso.

21.- Dado un segmento de Q muestras,  $\underline{X}_n$  de un proceso Gaussiano, de media  $\underline{\mu}$  y matriz de covarianza  $\underline{R}$ , conteste a las siguientes preguntas:

- a) ¿Cual es el estimador de Máxima Verosimilitud (ML) de la media del proceso  $\hat{\underline{\mu}}$  y cual es su varianza?
- b) Repita el apartado anterior para el caso de usar la media aritmética de las Q muestras.

Para el siguiente apartado use la desigualdad siguiente:  $\lambda_{\min}(\underline{\phi}) \cdot (\underline{a}^H \cdot \underline{a}) \leq \underline{a}^H \cdot \underline{\phi} \cdot \underline{a} \leq \lambda_{\max}(\underline{\phi}) \cdot (\underline{a}^H \cdot \underline{a})$

donde  $\underline{a}$  es un vector cualquiera y  $\lambda$  denota el autovalor mínimo y máximo de la matriz correspondiente. Recuerde que los autovalores de la inversa son la inversa de los autovalores.

- c) Compruebe que tanto el estimador ML como el de la media son consistentes en el sentido que la cota superior de la varianza tiende a cero cuando Q tiende a infinito.
- d) Asumiendo que  $\underline{a}^H \cdot \underline{\phi} \cdot \underline{a} = (\underline{a}^H \cdot \underline{\phi}^{0.5}) (\underline{\phi}^{0.5} \cdot \underline{a}) = \underline{u}^H \cdot \underline{u}$  y usando la desigualdad  $\underline{u}^H \cdot \underline{u} \cdot \underline{v}^H \cdot \underline{v} \geq |\underline{u}^H \cdot \underline{v}|^2$  demuestre que la varianza del estimador de media aritmética es siempre mayor o igual que el ML.

Asuma que la longitud del estimador es 3 (Q=3) y que tan solo dispone de los valores de autocorrelación r(0) y r(1). En esta situación, necesita calcular r(2). Si el proceso es AR(1) ( $x(n)=a \cdot x(n-1)+w(n)$ ),

- e) indique como calcularía r(2) sin recurrir a los datos de nuevo y solamente a partir de los valores de correlación que ya conoce. Pruebe que, al no recurrir a los datos, la varianza de la correlación estimada decrece con el índice, en lugar de aumentar que es lo que ocurriría al usar los datos.

22.- La señal y(n) es la suma de un proceso AR(1) y ruido blanco de media nula y potencia  $\sigma_w^2$ . Ambos procesos son independientes entre si.

$$y(n) = x(n) + w(n)$$

se pretende estimar el parámetro a(1) del proceso AR a partir de la observación del proceso {y}.

a.- Cual es la expresión del parámetro a(1) en funcion de la autocorrelación del proceso {x}?

Si se propone estimar el parámetro del proceso AR a partir de {y} según la expresión siguiente:

$$\hat{a}(1) = -\frac{r_{yy}(1)}{r_{yy}(0)}$$

b.- Si se define la relacion señal a ruido como  $SNR = \frac{r_{xx}(0)}{\sigma_w^2}$ , demostrar la relacion entre la estimacion del parámetro propuesta en funcion del parámetro correcto y la SNR definida.

c.- Cual es la expresión de la densidad espectral de potencia del proceso {y} y cual es el modelo de este proceso?

d.- Usando las ecuaciones de Y-W extendidas, estimar a(1) y el nivel de potencia  $\sigma_w^2$

23.- Dadas 2N-1 muestras de un proceso {x} resumidas en el vector  $\underline{X} = [x(-N+1) \quad x(-N+2) \quad \dots \quad x(N-2) \quad x(N-1)]^T$ , se sabe que la estructura del proceso es una componente determinista mas ruido gaussiano blanco (media nula) y varianza  $\sigma^2$ . Su estructura temporal es  $x(n)=(A+B \cdot n)+w(n)$ .

a.- Pruebe que el vector de datos puede expresarse como  $\underline{X} = \underline{\Phi} \underline{a} + \underline{w}$ , donde  $\underline{w}$  contiene las muestras de ruido, el vector  $\underline{a}$  viene dado por  $\underline{a} = \begin{bmatrix} A \\ B \end{bmatrix}$  y la matriz  $\underline{\Phi}$  por  $\underline{\Phi} = \begin{bmatrix} \mathbf{1} & \underline{N} \end{bmatrix}$  (matriz de  $2N-1$  por  $2$ ).

b.- Compruebe que se verifica que

$$\underline{\Phi}^H \underline{\Phi} = \frac{1}{(2N-1)S} \begin{bmatrix} S & 0 \\ 0 & (2N-1) \end{bmatrix} = \begin{bmatrix} 1/(2N-1) & 0 \\ 0 & 1/S \end{bmatrix}$$

siendo  $S = \sum_{n=-N+1}^{N-1} n^2 = \frac{N \cdot (N-1) \cdot (2N-1)}{3}$

c.- Derive el estimado ML del vector  $\underline{a}$  y compruebe que es insesgado.

d.- Compruebe que la forma del estimador del parámetro B es  $B^{ML} = \underline{N}^H \underline{X} / S$

e.- Calcule la varianza del estimador del vector  $\underline{a}$  y compruebe que el estimador de A y de B son consistentes (su varianza tiende a cero cuando N tiende a infinito).

## II.14 REFERENCIAS

[1] A. Papoulis. "Probability, Random Variables and Stochastic Processes". Mc. Graw Hill, New York. 1991.

[2] LL. Sharf. "Statistical Signal Processing, Detection, Estimation and Time Series Analysis". Addison-Wesley Reading, MA 1991.

[3] C.W. Helstrom. "Probability and stochastic processes for engineers". Macmillan Publishing Company. New York. 1984

[4] J.A. Cadzow. "Foundations of Digital Signal Processing and Data Analysis". Macmillan Publishing Company. New York. 1987.

[5] N. Abramson. "Information Theory and Coding". McGraw- Hill, New York, 1966.

[6] A. Leon Garcia. "Probability and Random Processes for Electrical Engineers". Addison Wesley, Reading, Mass., 1994.

[7] H.L. Van Trees. "Detection, Estimation and Modulation Theory". Part I. Jhon Wiley, New York, 1969.

[8] U. Grenader, G Szego. "Toeplitz Forms and Their Applications. Berkeley Press, 1958.

## APENDICE. LIMITE DE CRAMER-RAO

El límite de Cramer-Rao es una cota inferior para la varianza de cualquier estimador no sesgado. Si la varianza de un estimador alcanza esa cota se dice que el estimador es eficiente. Supongamos que disponemos de  $N$  observaciones de una realización de un proceso. El conjunto de las  $N$  variables aleatorias asociadas proceso  $\underline{X}_n = [x(n), \dots, x(n+N-1)]$  está caracterizado por una función de densidad de probabilidad  $f(\underline{X}_n / \theta)$  que depende de un parámetro  $\theta$  (en general puede ser un vector de varios parámetros) que queremos estimar a partir de las observaciones.

**Teorema de Cramer-Rao:** Si  $f(\underline{X}_n / \theta)$  es una función derivable en  $\theta$  la varianza de cualquier estimador insesgado de  $\theta$  está acotada por la expresión

$$\sigma_{\text{estimador}}^2 \geq \frac{1}{E \left\{ \left| \frac{\partial}{\partial \theta} \ln(f(\underline{X}_n / \theta)) \right|^2 \right\}}$$

evaluada en el verdadero valor de  $\theta$ . El estimador existe siempre y cuando podamos expresar la derivada de la siguiente forma:

$$\frac{\partial}{\partial \theta} \ln(f(\underline{X}_n / \theta)) = \alpha(\theta) (\hat{\theta}(\underline{X}_n) - \theta)$$

donde  $\alpha(\theta)$  es un término que no depende de  $\underline{X}_n$  y  $\hat{\theta}(\underline{X}_n)$  es el estimador que tiene varianza mínima.

**Demostración:** Usando el teorema de Schwartz podemos escribir,

$$E \left\{ (\hat{\theta}(\underline{X}_n) - \theta)^2 \right\} E \left\{ \left| \frac{\partial}{\partial \theta} \ln(f(\underline{X}_n / \theta)) \right|^2 \right\} \geq \left| E \left\{ (\hat{\theta}(\underline{X}_n) - \theta) \frac{\partial}{\partial \theta} \ln(f(\underline{X}_n / \theta)) \right\} \right|^2$$

Si la parte de la derecha de la desigualdad valiera 1 ya tendríamos el teorema demostrado.

$$0 = \frac{\partial}{\partial \theta} E \{ \hat{\theta}(\underline{X}_n) - \theta \} = \frac{\partial}{\partial \theta} \int \dots \int (\hat{\theta}(\underline{X}_n) - \theta) f(\underline{X}_n / \theta) dx(n) \dots dx(n+N-1)$$

Esta expresión vale cero si el estimador es insesgado. Si los límites de integración (es decir, el dominio de definición de  $f(\underline{X}_n / \theta)$ ) no dependen del parámetro  $\theta$  podemos intercambiar derivada e integrales:

$$\begin{aligned} 0 &= \frac{\partial}{\partial \theta} \int \dots \int (\hat{\theta}(\underline{X}_n) - \theta) f(\underline{X}_n / \theta) dx(n) \dots dx(n+N-1) = \\ &= \int \dots \int \frac{\partial}{\partial \theta} \left[ (\hat{\theta}(\underline{X}_n) - \theta) f(\underline{X}_n / \theta) \right] dx(n) \dots dx(n+N-1) = \\ &= \int \dots \int \frac{\partial f(\underline{X}_n / \theta)}{\partial \theta} (\hat{\theta}(\underline{X}_n) - \theta) dx(n) \dots dx(n+N-1) - \int \dots \int f(\underline{X}_n / \theta) dx(n) \dots dx(n+N-1) \end{aligned}$$

y la última de las integrales vale 1. Desarrollando la última expresión:

$$\begin{aligned} 1 &= \int \dots \int \frac{\partial f(\underline{X}_n / \theta)}{\partial \theta} (\hat{\theta}(\underline{X}_n) - \theta) dx(n) \dots dx(n+N-1) = \\ &= \int \dots \int \frac{1}{f(\underline{X}_n / \theta)} \frac{\partial f(\underline{X}_n / \theta)}{\partial \theta} f(\underline{X}_n / \theta) (\hat{\theta}(\underline{X}_n) - \theta) dx(n) \dots dx(n+N-1) = \\ &= E \left\{ \frac{\partial}{\partial \theta} \ln(f(\underline{X}_n / \theta)) (\hat{\theta}(\underline{X}_n) - \theta) \right\} \end{aligned}$$

Este es el término que queríamos que valiera 1, con lo cual queda demostrada la cota. La cota se obtiene siempre y cuando se cumplan las condiciones del teorema de Schwartz para obtener el óptimo:

$$\frac{\partial}{\partial \theta} \ln(f(\underline{X}_n / \theta)) = \alpha(\hat{\theta}(\underline{X}_n) - \theta)$$

Además, puede demostrarse que la varianza mínima es  $1/\alpha$ .