

Fundamentos del procesamiento estadístico de señales

Mi pequeño resumen – Lucas Camino

Índice

	<i>pág.</i>
Prefacio	1
Capítulo 1	1
1. Introducción	1
1.1 Estimación en el procesamiento de señales	1
1.1.1 Funcionamiento del radar	1
1.1.2 Funcionamiento del sonar	2
1.1.3 Funcionamiento del habla	2
1.2 El problema de la estimación matemática	3
1.2 El problema de la estimación matemática	3
1.2 El problema de la estimación matemática	3
1.2 El problema de la estimación matemática	3
1.2 El problema de la estimación matemática	3
1.2 El problema de la estimación matemática	3
Capítulo 2	8
2. Estimación insesgada de mínima varianza	8
2.1 Introducción	8
2.2 Resumen	8
2.3 Estimadores insesgados	8
2.4 Criterio de mínima varianza	9
2.5 Estimador insesgado de mínima varianza	9
2.6 Encontrando el estimador insesgado de mínima varianza (MVU)	9
2.7 Extensión para vector de parámetros	10

Prefacio

El libro apunta a enseñar procesamiento estadístico de señales a gente relacionada con el diseño e implementación de algoritmos de proc de señales, quienes buscan tener algoritmos óptimos para ser implementados en computadoras digitales.

Se asume que los conjuntos de datos con los que se trabajará serán señales continuas o secuencias de puntos.

Capítulo 1

1. Introducción

1.1 Estimación en el procesamiento de señales

La teoría de la estimación aplica en muchos sistemas: radares, sonares, habla, análisis de imágenes, biomedicina, comunicaciones, control, sismología.

El problema siempre es el mismo: **conocer (estimar) el valor de un grupo de parámetros.**

1.1.1 Funcionamiento del radar

Con el radar buscamos conocer la posición de objetos. La distancia entre un objeto y la antena del radar se llama rango (R). El tiempo [seg] t_0 en que se recibe la señal electromagnética que se envía, rebota y vuelve a la antena se puede entender como

$$t_0 = \frac{2R}{c}$$

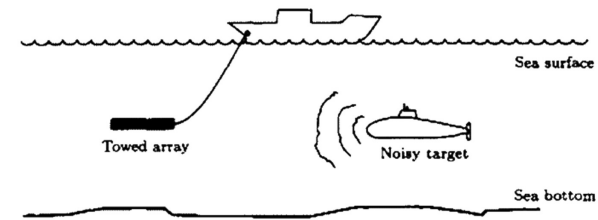
dado que la señal tiene que viajar una distancia R y volver (por eso 2 veces), y lo hace a una velocidad c , que es la velocidad de la propagación electromagnética. Es así que midiendo t_0 podemos calcular R.

La señal emitida es bonita y conocida. La recibida llega deteriorada por pérdida de energía por la propagación, ruido ambiental y *delays de tiempo ocasionados por los componentes electrónicos del receptor*.

La señal analógica recibida será convertida a digital por el radar para ser ingresada a una computadora.

1.1.2 Funcionamiento del sonar

Queremos conocer la posición de un objeto bajo el agua. El objetivo es ruidoso y este ruido se propaga por el agua y es recibido por un arreglo de sensores que cuelga del barco. De ahí la señal se lleva en digital a una computadora.



Debido a la distinta posición y de los sensores dentro del arreglo, la misma señal de ruido llega a cada uno desfasada. Nosotros conocemos la velocidad del sonido en el agua (c), la distancia entre los sensores (d) y podemos medir el tiempo t_0 que es la diferencia de tiempo que existe entre que cada sensor del arreglo recibe la señal de interés. Con estos datos podemos calcular β según la expresión

$$\beta = \arccos\left(\frac{c \cdot t_0}{d}\right)$$

Se puede ver que β será un valor estimado.

1.1.3 Funcionamiento del habla

El método más simple de reconocimiento del habla se hace por fonemas, es decir, sonidos del habla por separados.

La forma de onda de los fonemas vocálicos es periódica. El período de estas señales se denomina *pitch*. Para detectar vocales “a” y “e” podemos guardar tres muestras de cada uno de los sonidos hechos por una persona (entrenar al sistema) y luego analizar una nueva muestra y comparar con las señales previamente almacenadas. En base a la comparación de las señales es que definimos qué vocal dijo la persona.

Una persona puede variar la forma de pronunciar la misma vocal, por lo que el *pitch* de las señales de una misma vocal puede cambiar entre una y otra muestra.

Es por eso que podemos recurrir a analizar la envolvente espectral de las señales para compararlas ya que la transformada de Fourier de una señal periódica es la transformada de un único período replicada. **El cambio del período sólo afecta al espacio entre muestras del espectro pero no sus valores.**

El modelo usado para obtener la envolvente espectral de una señal se llama LPC (linear predictive coding). **Los parámetros del método determinan la envolvente de la señal.**

En todos estos sistemas tenemos el problema de conocer valores de parámetros basándonos en la forma de señales continuas, y como las computadoras son digitales, tenemos el mismo problema basado en señales digitales.

Queremos conocer el valor de un parámetro θ a partir de una muestra de N elementos que no es ni más ni menos que una señal discreta. En términos matemáticos, representamos esta señal como una función $x[n]$, donde cada valor de n representa una muestra. Construimos una expresión matemática que sea función de la muestra para poder conocer el valor del parámetro, es decir, construimos un **estimador** $\hat{\theta} = g(x[0], x[1], \dots, x[N-1])$.

1.2 El problema de la estimación matemática

El primer paso para encontrar un buen estimador es modelar matemáticamente los datos. Dado que los datos son por naturaleza aleatorios, los describimos con su función de densidad de probabilidad (FDP). La notación teórica es:

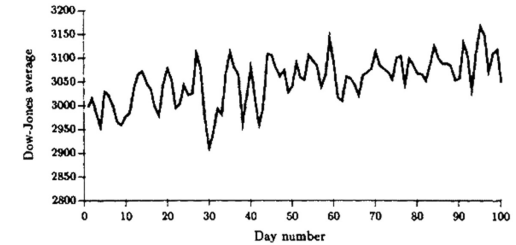
$$FDP = p(x[0], x[1], \dots, x[N-1]; \theta)$$

La FDP depende de un parámetro θ , cuyo valor se desconoce. Si el valor de θ cambia, obtendremos distinta FDP.

Si θ fuese la media poblacional, observando la FDP podríamos adivinar por dónde andaría el valor de θ . El “sentido común” es crítico al momento de hallar un buen estimador.

En un problema real no contamos con una FDP pero debemos elegir un estimador que sea no sólo consistente con los conocimientos previos y limitaciones de un problema sino que además sea matemáticamente manejable.

1.2.1 Ejemplo: “Hypothetical Dow-Jones average”



Tenemos estos datos los cuales, a pesar de ser fluctuantes, tienden a ir aumentando.

Para determinar si esto es cierto podemos pensar que estamos frente a una señal $x[n]$, la cual está contaminada por ruido.

Si la señal $x[n]$ es una línea recta que aumenta, su pendiente debe ser positiva, por lo que matemáticamente escribimos:

$$x[n] = A + B \cdot n, \text{ siendo } B \text{ la pendiente}$$

y si le sumamos ruido, el cual modelamos como ruido blanco gaussiano, llegamos a la expresión

$$x[n] = A + B \cdot n + w[n], n \in [0; N-1]$$

$$w[n] \sim N(0, \sigma^2) \text{ y cada muestra es independiente de la otra}$$

Como conocemos los parámetros del ruido, dado que es gaussiano, sólo nos quedan por descubrir los valores de los parámetros A y B . Si pensamos al conjunto de parámetros θ como un vector,

$$\theta = [AB]^T$$

Si el conjunto de datos \mathbf{x} es $\mathbf{x} = [x[0], x[1], \dots, x[N-1]]^T$, nuestra FDP es

$$p(\mathbf{x}; \theta) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \cdot e^{-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A - Bn)^2}$$

La modelización del conjunto de datos como una recta es consistente con el saber que el promedio ronda el valor 3000 (esto lo indica matemáticamente el valor A) y que en promedio el valor de los datos aumente lo indica un valor $B > 0$.

El rendimiento de un estimador está atado a los factores que hayamos asumido al momento de modelar la FDP. Podríamos pensar que el estimador es robusto si pequeños cambios de la FDP no afectan el rendimiento del estimador.

Estimadores basados en FDPs se asocian a la **estimación clásica**, donde el valor de los parámetros es desconocido pero determinístico. En el ejemplo de Dow-Jones sabemos de antemano que $2000 < A < 4000$. Sería imposible, según los datos brindados que $A < 2000$ ó $A > 4000$. En cambio, esperaríamos encontrar un valor para A que esté entre los 2800 y 3200.

Tal enfoque se denomina **estimación bayesiana**. El parámetro que intentamos estimar se considera como la realización de la variable aleatoria θ . **ALGO QUE NO ENTIENDO PASA** y los datos son descritos por la FDP combinada

$$p(x, \theta) = p(x|\theta) \cdot p(\theta)$$

$p(\theta)$ es la FDP anterior, que reunía los conocimientos que teníamos de θ antes de observar datos, y $p(x|\theta)$ es la FDP condicionada por el conocimiento de los datos x , conociendo θ .

Una vez que tengamos una FDP el problema que tenemos es el de determinar el mejor estimador, o función de la muestra óptima.

Se puede pensar a un estimador como una regla que le asigna valor a θ para cada realización de x .

1.3 Evaluar el rendimiento del estimador

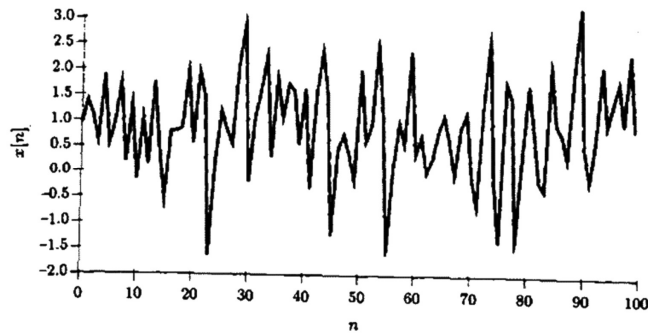


Fig 1.7

Con un vistazo rápido pareciera que $x[n]$ es un DC A en ruido (DC referencia a la corriente continua, lo que es una función constante). Si $x[n]$ es un valor constante más ruido, podemos modelarla matemáticamente como

$$x[n] = A + w[n], \text{ } w[n] \text{ es un proceso de media cero.}$$

Nos gustaría estimar el valor de A . Intuitivamente, como A es el nivel medio de $x[n]$ y la media de $w[n]$ es cero, podemos establecer el estimador

$$\hat{A} = \frac{1}{N} \cdot \sum_{n=0}^{N-1} x[n]$$

Ahora deberíamos preguntarnos:

1. ¿Qué tan cercano será el valor de \hat{A} al del parámetro A ?
2. ¿Existen mejores estimadores para la media muestral?

Para el conjunto de datos de la figura 1.7 resulta que $\hat{A}_1=0.9$, lo que es cercano al valor real que es $A=1$.

Para este mismo caso podríamos construir el estimador $\hat{A}_2=x[0]$, el cual, intuitivamente, podría no parecer un estimador funcional dado que no usa toda la información proporcionada por la muestra. No hay ningún tipo de *promedio* que ayude a disminuir los efectos del ruido. Sin embargo, para el conjunto de datos de la figura 1.7 este nuevo $\hat{A}_2=0.95$, lo que es más cercano al verdadero valor de A . ¿Podemos, entonces, decir que \hat{A}_2 es mejor que \hat{A}_1 ?

Como un estimador es una función que depende de los valores de la muestra y ésta es variable, un estimador es también variable. El valor de un estimador se entiende como **el valor del estimador en una realización**.

Para evaluar el rendimiento de un estimador debemos hacerlo estadísticamente. Una opción sería repetir el experimento que generó los datos y aplicar nuevamente cada estimador para cada conjunto de datos. Sólo así podríamos ver cuál estimador produjo menor error en la mayoría de los casos.

Supongamos que repetimos el experimento fijando $A=1$, agregando a la señal distintas realizaciones de ruido $w[n]$, de manera que obtendremos un conjunto de realizaciones para $x[n]$. Luego, evaluamos los estimadores para cada conjunto de datos y dibujamos los histogramas (gráfico que describe la cantidad de veces -frecuencia- que un estimador produjo cierto valor). Un histograma puede usarse para aproximar la FDP.

Con 100 realizaciones se obtienen estos histogramas

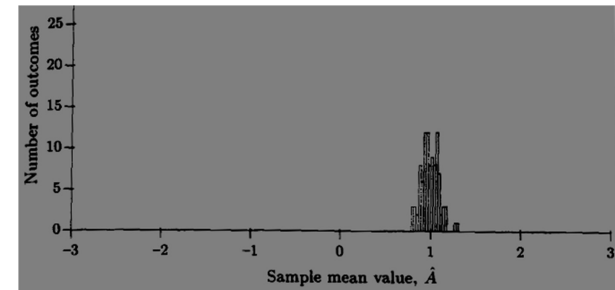


Fig 1.8 - \hat{A}_1

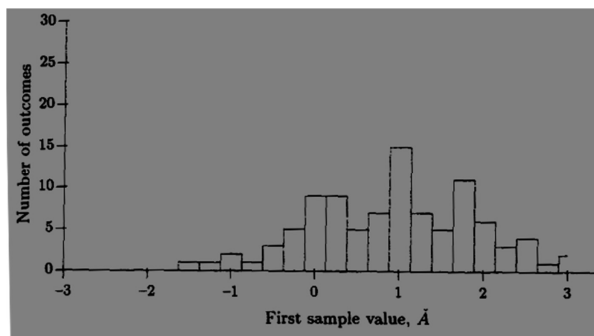


Fig 1.8 - \hat{A}_2

Ahora queda en evidencia que el estimador \hat{A}_1 es mejor porque usualmente produce un valor más cercano a $A=1$. Un escéptico dirá que la conclusión podrá no ser la misma si el experimento se realizara 1000 o 10.000 veces.

Para **probar** que el estimador \hat{A}_1 es mejor que el \hat{A}_2 podemos establecer que tiene menor varianza. Los supuestos que debemos mantener es que las realizaciones del ruido $w[n]$, además de tener media cero, tienen igual varianza σ^2 y que cada realización es independiente de la otra.

Vemos que \hat{A}_1 y \hat{A}_2 son insesgados, ya que $E(\hat{A}_1) = E(\hat{A}_2) = A$, y que $\text{var}(\hat{A}_1) = \text{var}(\hat{A}_1) = \frac{\sigma^2}{N}$ y $\text{var}(\hat{A}_2) = \sigma^2$. Como las varianzas de cada realización $w[n]$ son independientes y $\text{var}(\hat{A}_1) < \text{var}(\hat{A}_2)$. Asumiendo que $w[n]$ es gaussiano, podemos afirmar que la probabilidad de error es menor para \hat{A}_1 que para \hat{A}_2 .

Dos puntos importantes del ejemplo previo que deben grabarse en nuestras mentes:

1. Un estimador es una variable aleatoria y como tal su rendimiento puede sólo ser completamente descrito por dos maneras: estadísticamente o por su FDP (histograma).
2. El uso de simulaciones por computadora para evaluar el rendimiento de una estimación nunca es conclusivo. En el mejor de los casos el rendimiento satisface un grado de confianza determinado. En el peor de los casos podemos tener resultados erróneos debido a la falta de veces que repetimos un experimento o a errores de simulación.

Otro problema con el que nos encontraremos a menudo se debe al balance entre rendimiento y complejidad de cómputo. Mejores resultados se obtienen con mayor complejidad de cómputo. Si bien podemos encontrar muy buenos estimadores, a veces conviene utilizar peores estimadores (alternativos) pero que se “comporten mejor” computacionalmente. Este criterio queda a definición del usuario.

Capítulo 2

2. Estimación imparcial de varianza mínima

2.1 Introducción

En este capítulo buscaremos estimadores determinísticos que generen, en promedio, el verdadero valor del parámetro. Luego, encontraremos aquél que tenga menor varianza. Todo esto provocará que la mayor parte del tiempo el estimador se aproxime al valor del parámetro.

2.2 Resumen

Los estimadores insesgados de varianza mínima en general NO EXISTEN. Si existen pueden encontrarse por medio de los métodos de la cota de Cramer Rao y el concepto de “estadística suficiente”.

Si no existe un estimador imparcial de varianza mínima o si los dos métodos anteriores fallan, se puede construir un estimador menos óptimo haciendo una aproximación lineal.

2.3 Estimadores imparciales

Un estimador es imparcial si, en promedio, el estimador produce el valor del verdadero parámetro. El valor de un parámetro θ está en el intervalo $[a;b]$, si $\hat{\theta}$ es imparcial, en promedio valdrá θ .

Entonces, si $\hat{\theta}$ es imparcial, anotamos que:

$$E(\hat{\theta}) = \theta, \quad a < \hat{\theta} < b$$

Cuando el libro dice “unbiased” creo que está refiriéndose a “insesgado”

Los estimadores “unbiased” tienden a tener sus funciones de densidad de probabilidad centradas en el valor del parámetro. Esto no ocurre siempre.

Que un estimador sea insesgado no quiere decir que sea bueno. Sólo indica que en promedio valdrá lo que vale el parámetro.

Los estimadores insesgados tienen un error sistemático. Un sesgo persistente siempre resultará en un estimador pobre.

Podemos obtener un nuevo estimador que surja del promedio de estimadores insesgados. Como todos tienen esperanza del estimador igual al parámetro, el promedio de ellos también la tendrá. En este caso, la varianza del nuevo estimador será menor que la de todos los otros ya que

$$\begin{aligned} \text{var}(\bar{\theta}) &= \frac{1}{n^2} \sum_{i=1}^n \text{var}(\hat{\theta}_i) \\ &= \frac{\text{var}(\hat{\theta}_1)}{n} \end{aligned}$$

En cambio, si los estimadores son insesgados, la esperanza de cada estimador será igual al parámetro más un sesgo, definiendo al sesgo $b(\theta)$ como

$$b(\theta) = E(\hat{\theta}) - \theta$$

En estos casos no importa cuántos estimadores promediamos para formar un nuevo estimador. Este último NUNCA cubrirá el verdadero valor del parámetro.

2.4 Criterio de mínima varianza

Cuando buscamos el estimador óptimo debemos decidir por medio de criterios.

Un criterio común es el llamado Error Cuadrático Medio (ECM(θ)), que se define:

$$ECM(\theta) = \text{var}(\hat{\theta}) + \text{sesgo}^2$$

$$ECM(\theta) = \text{var}(\hat{\theta}) + b^2(\theta)$$

$$ECM(\theta) = \text{var}(\hat{\theta}) + (E(\hat{\theta}) - \theta)^2$$

Este criterio no termina siendo muy bueno y se prefiere encontrar un estimador con la mínima varianza.

2.5 Estimador insesgado de mínima varianza

El problema surge de preguntarnos si existe un estimador insesgado con la mínima varianza. Generalmente no existe. Podemos tener dos casos en donde, para ambos, hay 3 estimadores insesgados.

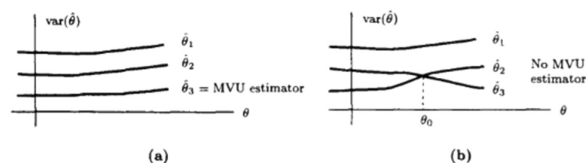


Figure 2.3 Possible dependence of estimator variance with θ

En el caso (a) podríamos decir que el estimador 3 es el mejor, pero para el caso (b) no hay un claro mejor estimador ya que el 3 es mejor para ciertos valores y el 2 para otros.

Hay casos en donde no hay un único estimador de mínima varianza posible.

2.6 Encontrando el estimador insesgado de mínima varianza (MVU)

No existe un procedimiento que permita encontrar el mejor estimador y de menor varianza. Sin embargo hay aproximaciones. Podemos:

- 1- Encontrar la cota inferior de Cramer-Rao.
- 2- Aplicar el teorema de Rao-Blackwell-Lehmann-Scheffe
- 3- Pedirle a los estimadores que no sólo sean insesgado sino también lineales. Luego, encontrar el estimador que menor varianza tenga.

Con 1 y 2 nos acercaremos al estimador de mínima varianza. Con 3 lo obtendremos sólo si el estimador es lineal en todos los datos.

Cramer-Rao dice que para cualquier estimador insesgado la varianza es mayor o igual a cierto valor. Si existiera un estimador cuya varianza es igual a la de Cramer-Rao, ése es el estimador de mínima varianza. Podría ocurrir que ningún estimador tenga la varianza perfecta que dice el método. Aún así podría existir un estimador de mínima varianza. Ahí es cuando avanzamos a 2.

El procedimiento 2 “encuentra una estadística suficiente, una que usa todos los datos de manera eficiente y luego encuentra una función de la estadística suficiente que es un estimador insesgado de x ”. **FALTA ACLARAR EL TEOREMA.**

El procedimiento 3 requiere que el estimador sea lineal, lo que es a veces difícil. Obviamente se podrá encontrar un estimador lineal de mínima varianza para ciertos conjuntos de datos.

2.7 Extensión para vector de parámetros

Tenemos un vector de parámetros $\theta = [\theta_1, \theta_2, \dots, \theta_p]^T$. Un estimador $\hat{\theta} = [\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p]^T$ será insesgados si se cumple:

$$E(\hat{\theta}_i) = \theta_i \quad a_i < \theta_i < b_i \quad \text{para todo } i.$$