



MANUAL DE USUARIO DE AnálisisSerial

(c) 2007 - Instituto de Ingeniería Eléctrica
Facultad de Ingeniería
Universidad de la República Oriental del Uruguay.
Software Libre distribuido bajo licencia GNU-GPL v3

18 de Junio de 2010
última revisión: 22 de febrero de 2014
Ruben Chaer
Montevideo - Uruguay.

1. Introducción.

El programa ANALISSERIAL es un utilitario auxiliar a la plataforma de simulación de sistemas de energía eléctrica SimSEE. Analisisserial es útil para analizar series temporales de datos y crear un modelo de Correlaciones en Espacio Gaussiano con Histograma CEGH.

2. Modelo CEGH.

Dada un conjunto de series temporales, que cubren el mismo horizonte temporal y con muestreo uniforme y sincronizado entre las series, es posible identificar un modelo que represente al conjunto de series en forma óptima manteniendo algunas características importantes de las series.

La primer pregunta es ¿para qué se quiere el modelo?.

La primer respuesta es, que el modelo sirve en SimSEE para generar series temporales sintéticas con las mismas características que el conjunto de series de datos utilizadas para crear el modelo. Además de la simple posibilidad de generar series temporales sintéticas con características similares a las históricas, el modelo intenta captar la estructura del proceso aleatorio permitiendo así dado un estado dado del proceso generar realizaciones de las salidas dado ese estado. Esta posibilidad, de poder generar salidas a partir del conocimiento del estado, es la que posibilita considerar el proceso estocástico los algoritmos de Programación Dinámica Estocástica.

En la programación dinámica estocástica, el cálculo de la función de costo futuro sobre el espacio de estados, se realiza en forma recursiva desde el FUTURO hasta el PRESENTE (tiempo inverso) por lo que es necesaria una forma coherente de generar los valores de las series aleatorias (ejemplo aporte a las represas) a partir del estado del sistema. En definitiva, si el conjunto de series carece de estado (o sea no tiene memoria) no sería necesario identificar un modelo que capture dicha memoria para representarla en el algoritmo de programación dinámica estocástica. Pero si la memoria del sistema oculto que genera las series de datos importa, será necesario disponer de un modelo que represente dicha memoria para poder llevar a cabo optimizaciones que dependan del estado. Dicho en otras palabras, cuando se representa “EL ESTADO” del sistema, se debe incluir el Estado de los procesos estocásticos (cuyo estado resulte relevante a los efectos de lo que se está estudiando) para que sea considerado en el algoritmo de programación dinámica estocástica. Si se necesita identificar el estado debemos tener un modelo del sistema oculto que genera las series. Un tipo de modelo posible es el CEGH, otro tpo de modelo puede ser una cadena de estados de Markov.

A continuación se hace un resumen del modelo CEGH, suficiente para entender el resultado generado por ANALISSERIAL. Para una descripción más detallada de los fundamentos de los modelos CEGH ver [1].

La idea detrás del modelo CEGH es lograr un modelo que sea capaz de generar series con igual histograma de amplitudes que las series originales y manteniendo las funciones de correlación cruzadas entre las series y de las series con sigo mismas y con sus pasados.

La clave de los modelos CEGH está en: 1) La función densidad de probabilidad de un proceso estocástico queda totalmente determinada por la función de Auto-correlación (para un proceso de varias variables se generaliza la definición de auto-

correlación con la matriz de covarianzas) y 2) La función de auto-correlación de una señal es la anti-transformada de Fourier del espectro de potencia de la señal.

Entonces, dado un modelo que genere igual espectro de potencia que el de las series de datos históricos, se tendrá un modelo que genera series con igual funciones de auto-correlación que las históricas y entonces si el proceso es gaussiano, como las funciones de densidad de probabilidad quedan determinadas por los coeficientes de la función de auto-correlación se tendrá un modelo que genera series con las mismas funciones de densidad de probabilidad y por tanto capaz de dar las mismas medidas de probabilidad.

Esto permite, dada una serie temporal, trabajar en el dominio de la frecuencia para la síntesis de un filtro lineal cuya respuesta en frecuencia coincida con el espectro de potencia de la señal. Ese filtro, cuando en su entrada tiene ruido blanco, genera series temporales con igual función de auto-correlación que la serie original. De esta forma para una serie tenemos una herramienta potente de generar modelos que capten la memoria de un sistema.

Otro resultado importante de los sistemas lineales es que la salida de un filtro lineal que es alimentado con ruido blanco gaussiano es gaussiana.

El modelo CEGH combina ambos resultados creando un filtro lineal que en un espacio transformado genera series temporales que representan procesos gaussianos con las funciones de correlación y auto-correlación que las series originales de datos con las que se identificó el modelo. Estas series temporales, son transformadas al espacio real (donde dejan de ser gaussianas) mediante un conjunto de funciones de deformación no-lineales que obligan a las series así transformadas a tener el mismo histograma de amplitudes que las series originales de datos.

La Fig.1 resume lo anterior, donde $rbg(t)$ es un vector de fuentes de ruido blanco gaussiano que ataca un filtro lineal generando el vector de salidas $xg(t)$. Las $xg(t)$ son las salidas de procesos gaussianos con las correlaciones impuestas por el filtro lineal. Luego, las series xg son transformadas por un conjunto de transformaciones no-lineales $TNL(x, t)$ obteniendo así el vector de series sintéticas $y(t)$.

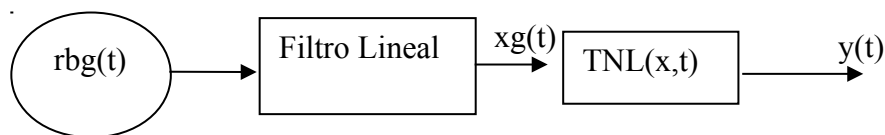


Fig. 1: Modelo CEGH. Procedimiento de síntesis de valores.

El programa ANALISSERIAL realiza una análisis de un conjunto de series de datos y como resultado produce el “Filtro Lineal” y las transformaciones TNL creando así lo necesario para definir el modelo CEGH del proceso. El modelo generado es utilizable en SimSEE, tanto en la etapa de optimización como en la simulación.

3. Ejecutando ANALISISERIAL.

El programa ANALISISERIAL se distribuye junto a SimSEE. Para ejecutar el programa basta ir con el explorador de windows a la carpeta `c:\simsee\bin` y abrir el ejecutable “`analisisserial.exe`”. La pantalla que se abre al ejecutarlo es la que se muestra en la Fig.2.

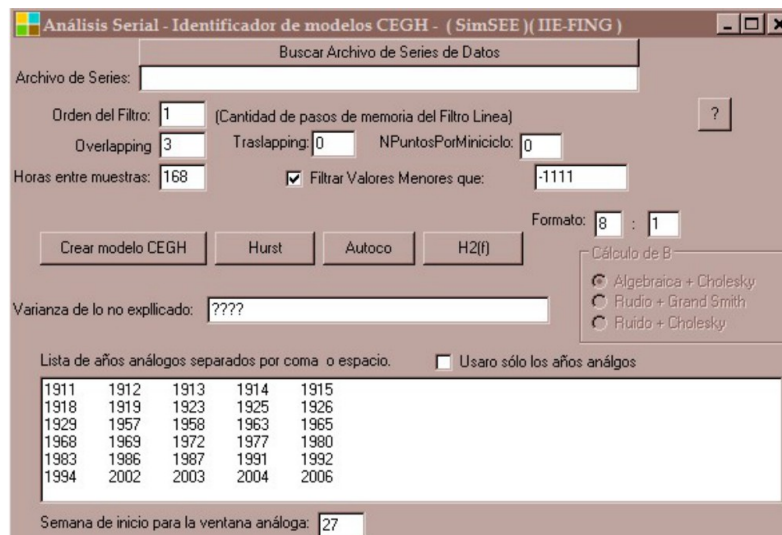


Fig. 2: Pantalla de AnalisisSerial.

Como se puede apreciar, en la parte superior está el botón “Buscar Archivo de Series de Datos” que permite seleccionar el archivo con las series de datos a procesar. Este archivo debe contener las series de datos con el formato especificado en la próxima sección.

Luego se debe especificar algunos parámetros (Orden del Filtro, Overlapping, etc.) y presionar el botón “Crear modelo CEGH” para que se ejecute el análisis de los datos de entrada y se escriban los resultados en el archivo de salida.

Una vez que termina el cálculo, se escribe el casillero “Varianza de lo no explicado”, que es una medida de lo no explicado por el filtro identificado. Si ese valor fuera 1, significaría que no se logró captar información útil de las series de entrada. Esto implicaría que las series de datos son ruido blanco y que no tiene sentido tratar de identificar un modelo CEGH. Esta medida de lo no explicado es útil para evaluar si al aumentar el orden del filtro mejora o no el modelo.

Para el cálculo de la matriz B del filtro, originalmente se permiten dos opciones. Una consistente en la descomposición por el método de Grand Smith de los vectores de ruido y la otra en la descomposición de Cholesky de la matriz de co-varianza de los vectores de ruidos. Recientemente se agregó la opción “Algebraica + Cholesky” que es la que está activa y se deshabilitó la posibilidad de que los usuarios elijan el método de cálculo de la matriz B. El método actual, calcula la matriz $\langle BB^T \rangle$ con operaciones algebraicas a partir de la matriz de covarianzas de las series de datos y de la matriz

A identificada para el filtro lineal por mínimos cuadrados y luego B es calculada mediante la descomposición de Cholesky. Esta forma de cálculo garantiza que las varianzas de las salidas del filtro lineal son unitarias. Los métodos anteriores utilizan los residuos de la identificación de A para la estimación de $\langle BB^T \rangle$ lo que no asegura

que la varianza fuera unitaria dado que dichos residuos no son 100% no correlacionados.

Las tres formas de cálculo logran una matriz B que descompone los vectores de ruidos en vectores estadísticamente independiente, pero no son iguales las matrices pues existen muchas soluciones al problema. Se optó por dejar deshabilitada la opción de cambiar el método para el usuario común (los usuarios avanzados pueden volver a habilitar la opción en los fuentes del proyecto de la aplicación y recompilar la misma).

4. Formato de los datos de entrada.

Las series de datos a analizar deben incluirse en un archivo de texto con el formato que se especifica a continuación. Como ejemplo de este tipo de archivo, se puede ver el archivo “C:\simsee\datos_comunes\sintetizadores\series_BPS.txt”, incluido en la distribución de SimSEE y que contiene las series de aportes medios semanales a las represas Bonete, Palmar y Salto Grande que son las tres más importantes del sistema hidro-eléctrico del Uruguay.

La Fig.3 muestra el inicio del referido archivo que servirá de ejemplo. El archivo es un simple archivo de texto. La separación entre columnas corresponde a tabuladores.

3	NSeries							
1909	1	1	0	0	0	// año mes día hora minuto segundo		
168.5723	//Período de muestreo en horas							
5216	NPuntos							
52	Puntos por ciclo							
	Bonete	Palmar	Salto					
0	55	4	2976					
1	50	3	1987					
2	34	7	1297					

Fig. 3: Formato del archivo de series de datos a analizar.

En la primera línea del archivo, se debe especificar la cantidad de series de datos incluidas. En este ejemplo son 3.

Las segunda línea contienen el instante temporal de comienzo de las series, esto es la fecha y hora de la primer muestra. El formato es: año, mes, día, hora, minuto, segundo (todos separados por tabuladores). El “// año mes día hora minuto segundo” es solo un comentario.

En la tercer línea se especifica el período de muestreo en horas. En este caso dice 168.5723 para indicar que las muestras son una por semana. Como los datos históricos en este caso están dados como 52 valores por año y son crónicas largas (100 años) la duración de las semanas en horas debe ser calculada para que no se produzca un desfase en el tiempo (no son semanas de 168 horas).

La cuarta línea dice la cantidad de muestras en cada serie. En este caso dice 5216. Esto es, hay para cada serie 5216 datos comenzando el primero el 1/1/1909 00:00:00.

La quinta línea dice la cantidad de puntos por ciclo que sería recomendable utilizar en la identificación de las funciones deformantes. En este ejemplo dice 52 e indica que los datos tienen una estacionalidad que se repita cada 52 muestras.

La sexta línea tiene un tabulador (primer columna en blanco) y luego los nombres con que se identifican las series de datos. En esta caso son “Bonete”, “Palmar” y “Salto”.

De la séptima línea en adelante se tiene en la primer columna el ordinal que identifica el paso de tiempo. En el ejemplo se muestran solamente los tres primeros pasos de tiempo teniendo por tanto en la primer columna 0, 1, 2. En el archivo completo de este ejemplo las líneas siguen hasta el paso de tiempo cuyo ordinal es 5215. En las columnas correspondientes a cada serie de datos se encuentra el valor verificado para cada paso de tiempo. El ordinal (primer columna) es solo para referencia del usuario, no es utilizado en los cálculos por AnalisisSerial.

5. Parámetros.

En la Fig.1 se puede ver los parámetros que se deben fijar antes de presionar el botón “Crear modelo CEGH”.

Estos parámetros son:

- **Orden del Filtro.** Este parámetro acepta valores enteros 1, 2, 3, ... N y especifica la memoria del filtro recursivo. El valor 1 indica que el filtro guarda la información de su salida en el paso anterior. Si el valor es 2, se almacenan las dos últimas salidas, 3 las tres últimas y así sucesivamente. Incrementar este valor implica incrementar la memoria del filtro y por lo tanto la cantidad de parámetros que se deben estimar y también la dimensión de la variable de estado del sistema lineal. El mejor valor de este parámetro debe surgir de un análisis de los resultados.
- **Overlapping.** Este parámetro es entero y puede ser 0, 1, 2, ... N. El Overlapping indica la cantidad de pasos adyacentes a cada muestra en la que debe ser considerada la muestra. Por ejemplo, si fijamos un valor de 2, estamos indicando que cada secuencia de muestras debe ser considerada para el instante al que corresponde cada muestra y para los siguientes 2 pasos y para los 2 pasos anteriores. De alguna forma, el Overlapping relativiza la información del tiempo en el que ocurre una muestra re-utilizándola en los pasos previos y anteriores. Esto tiene un impacto fuerte en la definición de las funciones deformantes dado que aporta mayor cantidad de muestras a la formación de los histogramas y además relativiza la ocurrencia de eventos extremos pocos probables permitiendo que los mismos ocurran en un entorno del tiempo en donde ocurrieron en las series de datos histórica. Por ejemplo, en la serie de datos de caudales de aportes hidráulicos a las represas existen unos eventos extremos en una semana de setiembre que si pusiera Overlapping=0 forzaría a las funciones deformantes a permitir la ocurrencia de ese extremo solamente en esa semana específica en que se verificó en los datos históricos, cuando es razonable suponer que ese mismo evento podría haber ocurrido en un entorno de mas menos 2 semanas de la que se verificó en la serie histórica (ver sec.5.1).
- **Traslapping.** Este parámetro es un entero 0, 1, 2, ...N. El traslapping es similar al Overlapping, en cuanto a que hace que una misma muestra sea utilizada en más de una posición temporal. A diferencia del overlapping, en lugar de ser posiciones temporales ayacentes a la original, son posiciones obtenidas de la original por “saltos” de NPuntosPorMiniCiclo (ver sec.5.1).
- **NPuntosPorMiniCiclo.** Fija la cantidad de pasos de tiempo que se utiliza por el mecanismo de Traslapping para identificar los instantes temporales en que se aplica la información de una muestra (ver sec.5.1).
- **Horas entre muestras.** Este parámetro no tiene consecuencias en la identificación, pero sirve para que al escribir los resultados se escriba este valor. Es importante esta información para utilizar el sintetizador en SimSEE. Según sea el paso de tiempo de simulación-optimización en SimSEE la fuente aleatoria que se crea asociada a un sintetizador será manejada con muestreo síncrono si el paso de tiempo coincide con el de la fuente o con sub-muestreo o sobre-muestreo según el paso de tiempo de la simulación sea superior o inferior al de la fuente.

- **Filtrar Valores Menores que.** Este parámetro tiene un check box que si se marca se aplica este filtro a las muestras y si no está marcado no se utiliza. Si está marcado, el filtro consiste en descartar las secuencias de muestras que contiene algún valor inferior al especificado en el casillero (con valor -11110 en la fig. 1). Esto sirve para señalar en las serie de datos los datos inválidos y así no tenerlos en cuenta al identificar el filtro. Por ejemplo, en un sistema de adquisición de velocidades de viento, cuando detectamos que el anemómetro está roto ponemos -999999 para que esos valores se filtren. Si se eliminaran directamente los valores se estaría perdiendo el sincronismo de los datos respecto de la época del año.

5.1. Más sobre *Overlapping* y *Traslapping*.

Los parámetros *Overlapping* y *Traslapping* permiten indicar que el efecto de una misma muestra se asigna a los pasos laterales y "saltando" NPuntosPorMiniciclos pasos a la derecha o a la izquierda (futuro y pasado respectivamente).

Si *Overlapping* = 3 y *Traslapping* = 0, una misma muestra será considerada en el instante de tiempo al que está asociada, en los 3 anteriores y en los 3 siguientes.

Si *Overlapping*= 3 y *Traslapping* = 2 con NPuntosPorMiniciclo = 10, la misma muestra se considerará en su paso de tiempo (casillero k), en los 3 anteriores y en los 3 siguientes, y ese grupo de 7 muestras (3+1+3) será considerado centrado en casillero k al que está asociado la muestra y desplazado hacia ambos lados de ese casillero con desplazamientos -20, -10, +10 y +20.

Usando ambos parámetros, una misma muestra será considerada en $(2 \textit{Overlapping} + 1)(2 \textit{Traslapping} + 1)$ casilleros.

Ejemplos de uso:

- En la identificación del CEGH de aportes semanales medios a las represas: *Overlapping* = 3 y *Traslapping* = 0 pueden ser valores razonables.
- En la identificación de velocidades de viento horarias: *Overlapping* = 2 y *Traslapping* = 7 con NPuntosPorMiniciclo = 24 son valores razonables.

La siguiente Fig.4 intenta clarificar el uso de estos parámetros.

En el primer cuadro se muestra la serie original en la que se ha marcado una muestra con un "1" (casillero amarillo).

En el segundo cuadro se muestra el efecto de *Overlapping*=2, *Traslapping*=0. La serie original es expandida en 5 series desplazando la original hacia adelante y hacia atrás en 2 casilleros.

El tercer cuadro muestra el efecto de *Overlapping*=2, *Traslapping*=1, NPuntosPorMiniciclo=7. Como se aprecia el grupo de 5 series antes expandido por el *Overlapping*=2 es a su vez expandido en dos nuevos grupos desplazando dicho grupo en 7 casilleros (cantidad de NPuntosPorMiniciclo) una vez hacia adelante y otra hacia atrás por haber impuesto *Traslapping*=1.

6. Formato del archivo de resultados.

El archivo de resultados se guarda en la carpeta “c:\simsee\bin” y tiene el nombre: “SintetizadorCEGH.txt”.

Este archivo contiene la descripción de las funciones deformantes necesarias para construir las funciones no lineales para pasar del espacio gaussiano al real y la descripción del filtro lineal. Primero están las funciones de deformación y luego la descripción del filtro lineal.

El siguiente cuadro muestra un ejemplo del comienzo de la descripción de las funciones deformantes.

<+FUNCIONES DEFORMANTES>		
NSS	3	Número de Series de Salida
NPP	52	Número de Puntos por Período
NPFD	200	Número de Puntos por Función Deformante
DurPasoSorteo	168	

El significado de los parámetros es el siguiente:

- NSS = Número de Series de Salida. Indica que el sintetizador en cuestión genera un vector de NSS series. En el ejemplo del cuadro anterior, corresponde a la salida de la identificación del sintetizador de caudales para las represas de Bonete, Palmar y Salto y por eso NSS es 3.
- NPP = Número de Puntos por Período. Las funciones deformantes se pueden calcular de forma de captar la estacionalidad que tengan los datos originales en las mismas. Dependiendo del tipo de datos, dependerá la estacionalidad que valga la pena intentar captar en las mismas funciones de deformación. En el ejemplo, las series temporales de datos correspondían a los caudales medios semanales, a las represas de Bonete, Palmar y Salto disponiendo de 52 valores por año. Es claro que los histogramas de las series presentan una marcada estacionalidad. Para lograr reproducir los histogramas con la misma estacionalidad, las funciones deformantes se realizaron considerando una distinta para cada semana del año y por eso en el ejemplo NPP = 52.
- NPFD = Número de Puntos de las funciones deformantes. Las funciones deformantes están descritas por la inversa de la curva de probabilidad acumulada. Dicha función está descrita mediante los valores que toma en una discretización realizada del intervalo (0,1). NPFD es la cantidad de puntos considerados en dicha discretización.
- DurPasoSorteo = es la duración del paso de sorteo en horas de los datos originales. Este valor es necesario para determinar el comportamiento del modelo en SimSEE. Por ejemplo, si el paso de sorteo es 168 horas como en el ejemplo, y se utiliza en una corrida de SimSEE con paso de simulación horario, el sintetizador generará valores cada 168 pasos de simulación. En los pasos de simulación, SimSEE considera la interpolación de los valores sintetizados. También puede ocurrir a la inversa, es decir que el paso de simulación de la corrida que se está realizando en SimSEE sea superior al paso de sorteo de una fuente aleatoria (en este caso nuestro modelo CEGH). SimSEE administra

automáticamente las situaciones de sub-muestreo, sobre-muestreo o muestreo síncrono, pero para hacerlo necesita saber el paso de tiempo válido para cada sub-modelo. Este es el propósito de este parámetro.

Continuando con la descripción del contenido del archivo de resultados, a continuación del encabezado mostrado en el cuadro anterior viene la descripción de las funciones deformantes para cada una de las series. En el ejemplo tendremos entonces, 3 bloques (uno para cada serie) con la descripción de las 52 funciones deformantes (una para cada punto del período) de cada series. Cada función deformante está descrita en el ejemplo por los 200 valores de la inversa de la función de probabilidad acumulada.

La Fig.5 muestra el inicio de la descripción de las funciones deformantes de la primera de las series (Bonete en el ejemplo).

serie1	Bonete	0.50%	1.00%	1.50%	2.00%	2.50%	3.00%	3.50%	...
paso:	1	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	...
paso:	2	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	...
paso:	3	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	...
paso:	4	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	...
paso:	5	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	...
paso:	6	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	...
paso:	7	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	...
paso:	8	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	...
paso:	9	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	...
paso:	10	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	...
paso:	11	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	...

Fig. 5: Inicio de descripción de funciones deformantes.

El archivo de resultado es texto plano con los valores separados por tabuladores por lo que si es abierto con una aplicación como Excel se despliega como se muestra en la Fig.5.

En la primer fila, primer columna está la palabra serie seguida del ordinal (1, 2 ... NSS) a la que corresponde la descripción. En la segunda columna el nombre que identifica la serie. Este nombre es el que aparece como “nombre de borne” cuando el modelo es usado en SimSEE.

En la segunda fila, están libres las dos primeras columnas (dos tabs) y luego están la probabilidad a la que corresponde cada una de las columnas. Hay tantas columnas como NPPD.

Luego, para cada NPP hay una fila, con la palabra “paso” en la primer columna y el ordinal del punto dentro del período al que corresponde la función deformante contenida en la misma fila. En las siguientes columnas están los valores de la inversa de la función densidad de probabilidad acumulada para cada una de las probabilidades enunciadas en la fila 2.

Al final del bloque correspondiente a una serie, se deja una fila en blanco y comienza la descripción de las funciones deformantes de la siguiente serie y así sucesivamente hasta haber descrito todas las series.

Luego de la descripción de las funciones deformantes, sigue la descripción del filtro lineal y de posibles reducciones del estado a aplicar en SimSEE.

La Fig.6 muestra el final del archivo usado de ejemplo, con la descripción del filtro lineal y una reducción de las variables de estado de 3 a 2 variables.

<+FILTRO LINEAL>									
NFRBG	3								
NSS	3								
NCOLSA	3								
Filtro A									
			S1-1	S2-1	S3-1		u1	u2	u3
serie:	1		7.61E-01	2.49E-02	7.64E-02		3.80E-01	-1.81E-01	-3.97E-01
serie:	2		1.58E-01	6.26E-01	9.50E-03		6.14E-01	2.53E-01	1.74E-01
serie:	3		1.21E-01	-3.31E-02	7.80E-01		1.82E-01	-4.77E-01	2.43E-01
nVE	2								
nd1	5	Estado_H1	0.5	0	0.5	Estadoinicial	0		
probs	0.2	0.2	0.2	0.2	0.2				
nd2	5	Estado_H2	0	1	0	Estadoinicial	0		
probs	0.2	0.2	0.2	0.2	0.2				

Fig. 6: Descripción del filtro lineal y de la reducción de estado.

Al comienzo en la Fig.6, se tiene en la primer columna el texto “<+FILTRO INEAL>” que indica que comienza la descripción del filtro lineal. Las siguientes tres filas definen los valores de NFRBG, NSS y NCOLSA que en el ejemplo son 3, 3 y 3 y cuyos significados son:

- NFRBG = número de fuentes de ruido blanco gaussiano. En la actualidad por construcción este número coincide con el número de series, pero podría ser diferente si más adelante se cambia el algoritmo de identificación.
- NSS = Número de series. (salidas del modelo)
- NCOLSA = Número de columnas de la matriz principal del filtro lineal. Este número es igual a NSS multiplicado por la cantidad de retardos en el tiempo que tenga en cuenta el filtro lineal. En el ejemplo $NCOLSA = 3 = NSS$ quiere decir que el filtro en cuestión solamente considera 1 (un) retardo de las señales en el tiempo.

Luego de este encabezado, va una línea en blanco y otra que comienza con la palabra “Filtro A” anunciando que vienen los coeficientes del filtro lineal.

El filtro lineal está identificado por dos matrices A y B cada una de NSS filas.

La matriz A tiene NCOLSA columnas y la matriz B tiene NFRBG columnas. En la tabla, la matriz A está encabezada por los textos: “S1-1”, “S2-1” y “S3-1” que indican que sus coeficientes multiplican el valor de la serie 1, 2 y 3 en el paso anterior “-1”.

En el ejemplo, la matriz B está encabezada por “u1”, “u2” y “u3” y multiplican a los valores de las fuentes de ruido blanco gaussiano.

Cada fila de las matrices está precedida de un identificador de fila del tipo “serie: 1 |” (primeras tres columnas de la tabla) seguida por los coeficientes.

Cada fila describe la forma de calcular la siguiente salida del filtro para cada una de las series. A modo de ejemplo, usando los coeficientes de la fila k podemos escribir:

$$s_k(t) = \sum_{i=1}^{i=NSS} \sum_{j=1}^{j=NRT} A[k, (i-1) * NRT + j] * s_k(t - jT) + \sum_{h=1}^{h=NFRBG} B[k, h] * u_h(t)$$

En donde:

- $s_k(t)$, es la salida del filtro para la serie k .
- NRT es la cantidad de retardos del tiempo considerados y se puede calcular como: $NRT = NCOLSA / NSS$. (es 1 en el ejemplo).
- $u_h(t)$, son las entradas de ruido blanco gaussiano de varianza 1.

Continuando con la descripción del archivo de resultados, luego de las filas correspondientes a las matrices del filtro, viene una línea en blanco y comienza la descripción de las variables de estado a ser utilizadas si el modelo CEGH es utilizado para una Optimización Dinámica Estocástica. Por defecto, ANALISSERIAL escribe la descripción de las variables de estado como una variable para cada columna de la matriz A, indicando una discretización de 5 para cada variable y asignando a cada discretización el 20% de probabilidad. En ese caso, el estado del filtro lineal tiene igual dimensión en la Optimización que en la Simulación.

En la práctica, generalmente esta parte del archivo hay que editarla a mano para reflejar la reducción del espacio de estados que se quiere realizar en la etapa de Optimización en SimSEE como forma de luchar contra la Maldición de la Dimensionalidad de Bellman. Por ejemplo, en la Fig.6 se muestra que en lugar de 3 variables de estado, se utilizarán 2 para la optimización, que la primera de esas variables se discretiza en 5 puntos en el espacio de estado, se la llama “Estado_H1” y se define como: Estado_H1 = 0.5 * Bonete + 0.0 * Palmar + 0.5 * Salto (siempre operando sobre las señales en el espacio gaussiano).

La variable Estado_H2, también será discretizada en 5 puntos y está definida como: Estado_H2 = 0.0 * Bonete + 1.0 * Palmar + 0.0 * Salto

Por defecto AnalisisSerial escribe una descripción de las variables de estado para optimización que coincide con el estado del filtro lineal imponiendo 5 (cinco) puntos de discretización para cada variable. Por ejemplo, en el modelo del ejemplo de la Fig.6, en la salida del programa AnalisisSerial diría nVE=3 y a continuación vendría la descripción de las tres variables con igual nombre que las series, con una 5 puntos de discretización cada una con probabilidades 0.2 para cada punto y los coeficientes de la matriz reductora serían tal que la misma es la identidad.

Las variables de estado tienen por construcción distribución normal $N(0,1)$. La partición del rango de cada variable en “puntos” para la discretización del espacio de estado implica que cada punto es el “centro” de una banda de probabilidad. El valor indicado en el modelo, es el área (probabilidad acumulada) asociada a la banda de probabilidad y el punto discretización se considera en el valor de la variable que deja igual área dentro de la banda a cada lado del punto.

Para fijar ideas, en la Fig.7 se muestra la curva correspondiente a la densidad de probabilidad acumulada (CDF) de una variable x con distribución Normal de media cero y varianza unidad. Las líneas verticales punteadas (verde) separan las bandas de 0.2 de probabilidad acumulada. Los círculos con las flechas verticales identifican los puntos centrales en cada banda que corresponden a los valores considerados como puntos de discretización de la variable cuando sean usados en el algoritmo de Programación Dinámica Estocástica de SimSEE.

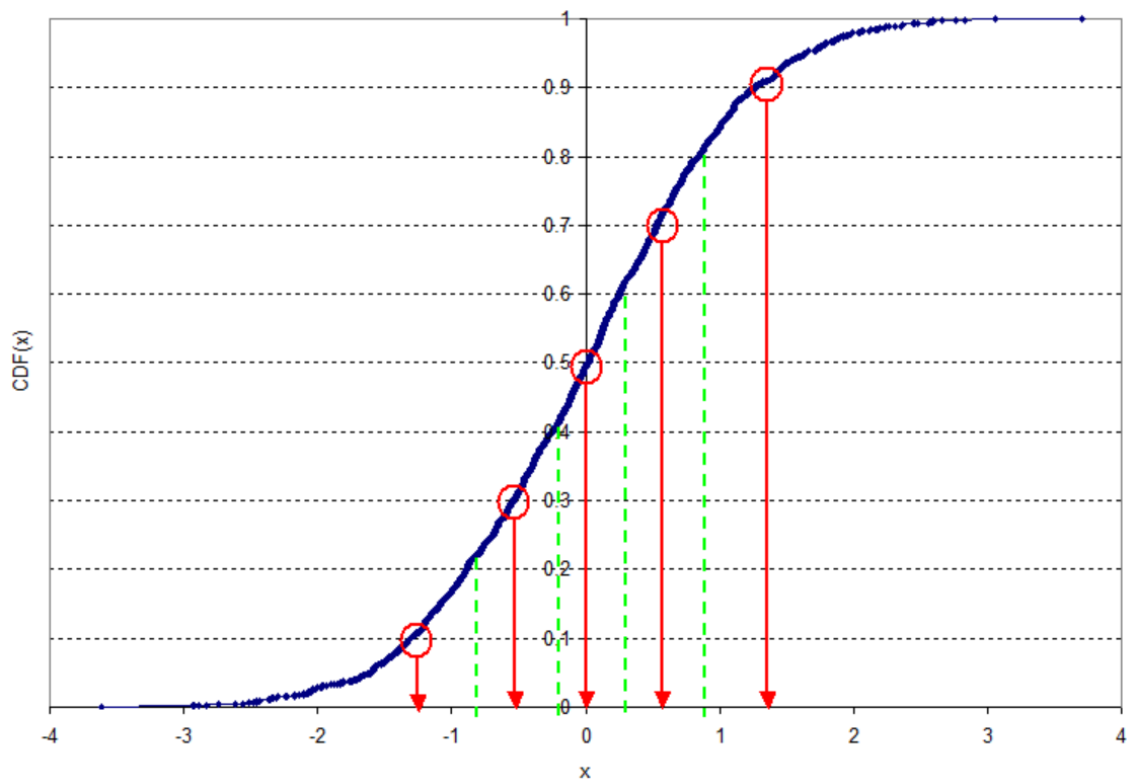


Fig. 7: Ejemplo de partición en 5 bandas equiprobables.

Bibliografía

1: Ruben Chaer, Fundamentos de modelo CEGH de procesos estocásticos multivariados., 2011