Capítulo 2

Métodos estadísticos

2.1. Simulación estadística

La simulación estadística consiste en generar realizaciones de variables aleatorias que siguen unas distribuciones concretas. Si esas variables son independientes las generaremos mediante simulación univariante, considerando únicamente su distribución individual. En cambio si dos o más variables aleatorias estan relacionadas por una distribución conjunta, necesitaremos de simulación multivariante.

Simulación univariante

La simulación univariante se basa en el teorema de transformación integral, expresado en el anejo de estadística. Este teorema nos afirma que si $F_X(x)$ es la función de distribución acumulada de X, y U es una variable aleatoria con distribución uniforme en $(0, 1)$. Entonces $F_X(u)^{-1}$, con u una realización de U, tiene la distribución de X.

Para realizar la simulación de las variables se han utilizado dos métodos, dependiendo de la existencia o no de la forma explícita de la inversa de la función de distribución acumulada. En todos los casos se conocía la expresión analítica de la función de distribución.

En el caso en que la función de distribución acumulada es conocida, la simulación se realiza en dos pasos.

1. Se realiza la simulación de la variable aleatoria $u \sim U(0, 1)$, utilizando la funciones internas del lenguaje de programación.

2. La variable aleatoria simulada x se genera como $x = F^{-1}(u)$.

Desafortunadamente algunas veces nos encontramos en el caso en que no conocemos la expresión analítica de la función de distribución acumulada. En esos casos, en que lo único que conocemos es la función de densidad de probabilidad, se debe realizar un tratamiento numérico para encontrar la antiimagen del número aleatorio generado. La secuencia completa es la siguiente. Como notación utilizaremos $F(x_i)$ como el valor de la distribución acumulada en el punto x_i

1. Se definen los límites en los que se va a calcular de forma numérica la función de distribución acumulada.

2. Se define el número de puntos base en los que se va a calcular el valor de la función de distribución acumulada.

3. Se definen los puntos utilizados para integrar numéricamente la función en los intervalos entre puntos base (en el programa se ha utilizado una cuadratura de Newton-Cotes de primer orden) y se calcula la integral de cada uno de los intervalos entre puntos base.

4. Se acumula el valor en los puntos base, obteniéndose $F(x_0), F(x_1), ..., F(x_n)$. Como en general la integral numérica de la función de densidad no será 1, es decir $F(x_n) \neq 1$ se reescala dividiendo el valor en todos los puntos base por $F(x_n)$.

5. Se realiza la simulación de la variable aleatoria $u \sim U(0, 1)$, utilizando las funciones internas del lenguaje de programación.

6. Se hallan x_i, x_{i+1} tales que $F(x_i) \leq u \leq F(x_{i+1})$

7. Se halla x utilizando interpolación (en el programa se ha utilizado interpolación lineal).

Simulación bivariante

La realización de una simulación bivariante (simulación en la que se deben generar dos variables aleatorias que se encuentran relacionadas por una distribución conjunta) es mucho más complicada que la realización de una simulación univariante, ya que no poseemos una herramienta tan potente como el Th. de transformación integral. No obstante podemos recurrir a métodos que bajo algunas restricciones nos permiten transformar la simulación bivariante en una serie de simulaciones univariantes. Como notación consideraremos que

las variables a simular son X e Y, con distribución conjunta $f_{XY}(x, y)$. Distribuciones condicionales $f_X(x|Y = y)$ y $f_Y(y|X = x)$. Si son independientes $f_X(x|Y = y) = f_X(x)$ y $f_Y(y|X = x) = f_Y(y)$.

Si las variables X, Y son variables aleatorias independientes, por definición podemos escribir $f_{XY}(x, y) = f_X(x)f_Y(y)$. En ese caso el esquema a seguir simplemente es:

1. Encontrar $f_X(x)$ y $f_Y(y)$ tales que $f_{XY}(x, y) = f_X(x)f_Y(y)$
2. Simular X e Y de forma separada siguiendo el procedimiento descrito para la simulación univariante

Variables dependientes que mediante una transformación se pueden obtener variables independientes

El esquema general es:

1. Transformar las variables aleatorias $\mathbf{X} = (X, Y)$ a $\mathbf{Z} = (Z_1, Z_2)$ tales que $f_{Z_1Z_2} = f_{Z_1}(Z_1)f_{Z_2}(Z_2)$, es decir, que sean independientes.
2. Simular utilizando el procedimiento anterior para generar Z_1 y Z_2.
3. Aplicar la transformación inversa a Z_1 y Z_2 para obtener X e Y.

Un ejemplo claro de este caso es la simulación de variables que siguen una distribución binormal. Se puede consultar el ejemplo en el apéndice de simulación.

Variables dependientes de las que poseemos las distribuciones una distribución marginal y una condicionada

En este caso podemos simular las dos variables por separado, siguiendo el siguiente esquema:

1. Simular la primera variable utilizando la distribución marginal
2. Hallar la distribución de la segunda variable condicionada al valor generado de la primera
3. Con la distribución encontrada en el paso anterior simular la segunda variable

2.2. Análisis espectral del oleaje

Espectro del oleaje

El uso de análisis espectral es un elemento importante en la forma de abordar el análisis del estado del mar. Se pasa de una descripción ola a ola del oleaje, a una descripción en la que se considera el oleaje como una perturbación aleatoria de la superficie libre entorno a un nivel medio, con algunas características de su distribución preestablecidas o conocidas. Esta visión hace posible el uso de la teoría de análisis de señales con todas las herramientas matemáticas de análisis de Fourier que esta conlleva.

Podemos definir el espectro del oleaje de manera informal como la distribución de la energia asociada al oleaje (potencia en realidad) con la frecuencia. De hecho si seguimos la teoría lineal y multiplicamos el espectro por ρg, coincide con la potencia suministrada por el oleaje. No obstante, la definición formal del espectro del oleaje es la transformada de Fourier de la función de autocovarianza de la elevación de la superfície libre. En esta misma tesina se incluye información sobre la función de autocovarianza en el apéndice de estadística y sobre transformadas de Fourier en el apéndice de análisis de Fourier. Como el muestreo de la superficie libre no se realiza de forma continua, si no que se realiza una medición cada cierto tiempo (periodo de muestreo) no podemos aplicar la Transformada de Fourier continua, sino la Transformada Finita de Fourier (TFF).

Modelo Autorregresivo

Se define un proceso autorregresivo como aquel en el que podemos predecir el estado en un instante a partir de los estados anteriores. En concreto decimos que un proceso autorregresivo es de orden p si el estado en el instante t se puede determinar en función de los p instantes anteriores.

$$X(t) = -\sum_{k=1}^{p} b_k X(t-k) + W(t), \quad W(t) \sim N(0, \sigma_W^2) \tag{2.1}$$

en donde:

- $X(i)$ Posición de la superficie libre en el instante i.

- $W(t)$ Ruido blanco a introducir. Se trata de una perturbación aleatoria que sigue una distribución de media 0 y varianza σ_W^2.

Partiendo del espectro del oleaje, se deben encontrar los coeficientes que regirán el proceso autorregresivo, así como la varianza del ruido blanco a introducir. A tal efecto se aplica sobre el espectro del oleaje el siguiente esquema. Las bases teóricas se encuentran expuestas en el apéndice sobre análisis de Fourier.

1. Discretización del espectro en el caso de ser continuo

2. Sobre el espectro discretizado aplicar la transformada discreta de Fourier inversa, obteniéndose la función de autocovarianza.

$$S(w_0), \ldots, S(w_n) \Rightarrow C(-(n-1)), \ldots, C(n-1) \tag{2.2}$$

3. Nos quedamos con un número limitado de términos de la función de autocovarianza, en este caso estaríamos ante un proceso autorregresivo de orden p. $C(-p), \ldots, C(0), \ldots, C(p)$

4. Planteamos las ecuaciones normales, como podemos encontrar en Brockwell & Davis (1987):

$$\begin{bmatrix} C(0) & C(-1) & \cdots & C(-(p-1)) & C(-p) \\ C(1) & C(0) & \cdots & C(-(p-2)) & C(-(p-1)) \\ \vdots & & \ddots & & \vdots \\ C(p) & C(-(p-1)) & \cdots & C(-1) & C(0) \end{bmatrix} \begin{bmatrix} 1 \\ b_1 \\ \vdots \\ b_p \end{bmatrix} = \begin{bmatrix} \sigma_W^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{2.3}$$

El sistema podemos reescribirlo como:

$$\begin{bmatrix} C(0) & \cdots & C(-(p-2)) & C(-(p-1)) \\ \vdots & \ddots & & \vdots \\ C(-(p-1)) & \cdots & C(-1) & C(0) \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_p \end{bmatrix} = \begin{bmatrix} -C(1) \\ \vdots \\ -C(p) \end{bmatrix} \tag{2.4}$$

con:

$$\sigma_W^2 = C(0) + b_1 C(-1) + \ldots + b_p C(-p) \tag{2.5}$$

5. resolvemos el sistema obteniendo $b_1, b_2, \ldots, b_p$ y por consiguiente σ_W^2

Supongamos un proceso autorregresivo de orden p. A partir de los coeficientes autorregresivos se puede reconstruir el espectro original del que derivan como:

$$S(\omega) = \frac{\sigma_W^2}{\left|\sum_{k=0}^{p} b_k \exp(-i\omega T k)\right|^2} \tag{2.6}$$

En donde:

- σ_W^2 es la varianza del ruido blanco utilizado en el modelo.
- b_k es el k-ésimo coeficiente autorregresivo
- T es el periodo de muestreo. Es decir el intervalo de tiempo entre una posición simulada y la siguiente. La máxima pulsación que se puede conseguir en la simulación es la pulsación de Nyquist: π/T.
- ω Es la pulsación angular, que equivale a 2π por la frecuencia, o lo que es lo mismo 2π entre el periodo.

Una vez obtenidos los coeficientes autorregresivos, la simulación de la elevación de la superficie libre es algo inmediato.

1. Se escogen los valores iniciales del proceso, generados aleatoriamente o por cualquier otro método. Si estamos en un proceso AR de orden p se necesitan p valores, estos serán $X(-p+1), \ldots, X(0)$.
2. Se genera la posición 1 a partir de la fórmula escrita anteriormente.
3. A partir de las posiciones $X(-p+2), \ldots, X(1)$ se genera la posición $X(2)$.
4. Se continua el procedimiento las veces necesarias.

El principal problema a resolver son las condiciones iniciales, que influyen de forma muy importante en el inicio del proceso, y su influencia se extiende hasta muchas simulaciones después. Para resolver este problema, lo que se ha hecho es generar unas condiciones iniciales de forma aleatoria, pero que respetan el proceso AR al que pertenecen. El método utilizado consiste en, partiendo de unas condiciones iniciales fijas, generar posiciones aleatorias de la superfície libre del mar un determinado número de veces. Después de esta generación se eligen los últimos p términos generados (si se han generado suficientes posiciones son aleatorios) y se utilizan como condición inicial del proceso. La influencia de las condiciones iniciales en una proceso AR se pueden observar en los gráficos de la página 16.

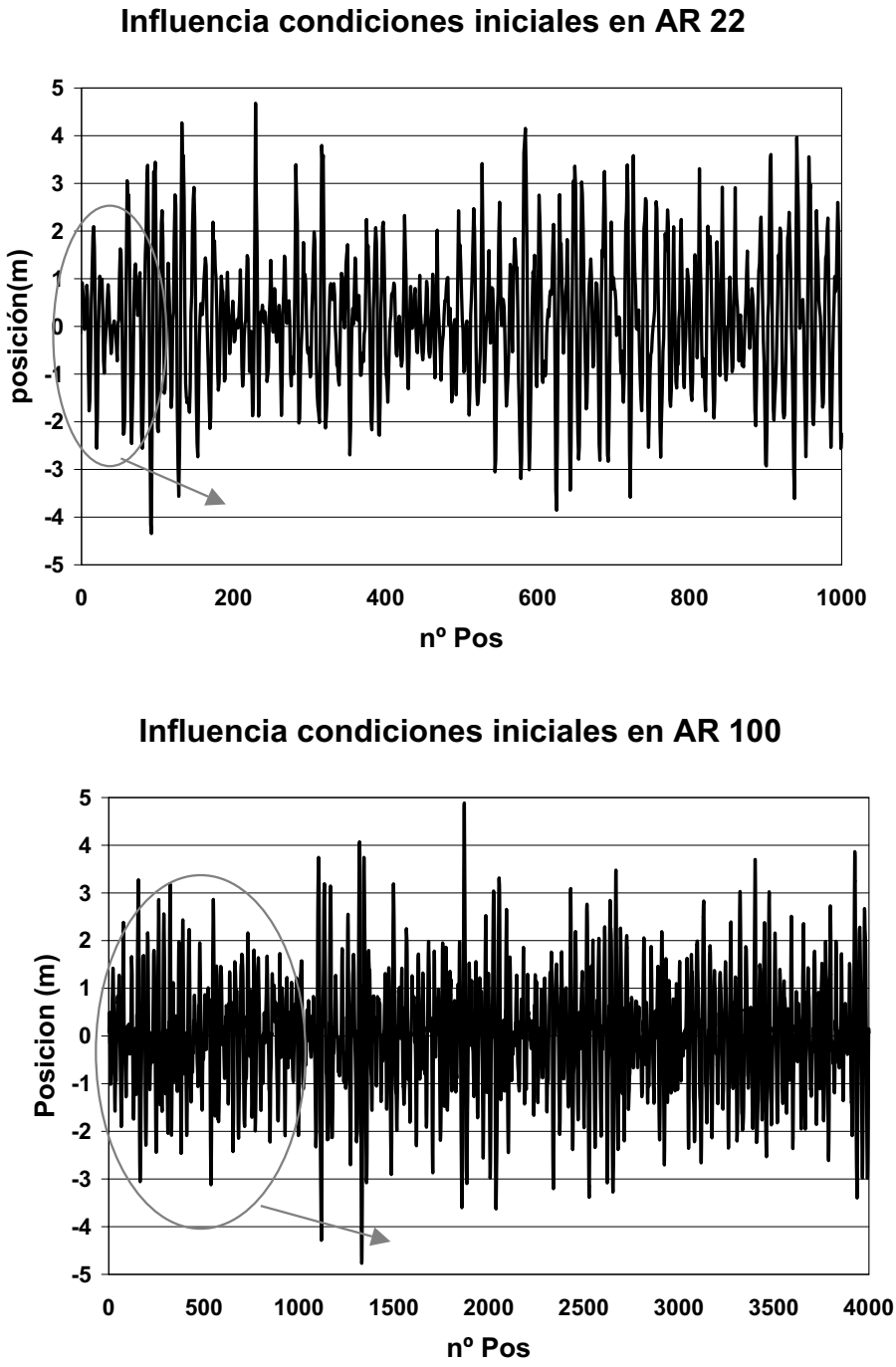

Figura 2.1: Influencia de las condiciones iniciales en la simulación