

APRENDIZAJE: CONCEPTOS GENERALES

Jesús Manuel de la Cruz García y Gonzalo Pajares Martinsanz

A1.1 INTRODUCCIÓN

En este apéndice se muestra el procedimiento seguido para obtener algunos de los resultados y figuras del capítulo 1.

La figura 1.6 y los datos para construirla se encuentran en el archivo en formato Excel con el nombre *histogramas.xls*. En la hoja Excel correspondiente aparece una tabla de la forma que se indica también en la tabla 1.3, con seis valores posibles de la variable X y cuatro de la variable Y, así como las ocurrencias de los datos (x_i, y_j) utilizados para construir la figura. También se muestra la tabla de probabilidades que se obtiene al dividir los valores en la anterior por el número de datos u observaciones.

El ejemplo de la sección 1.4.2 de regresión, y la figura 1.9 se obtienen con Matlab <http://www.mathworks.com/> utilizando el archivo *sc_regr.m*. En este archivo se obtienen los parámetros de la familia de funciones (1.46), que mejor aproximan a los puntos observados, mediante la minimización del riesgo empírico (1.47). La optimización se realiza mediante la función de Matlab *fminsearch*. La función utiliza el archivo *fun6b.m* para evaluar el riesgo empírico. Comentada en *sc_regr.m* aparece una sentencia que permite calcular el vector de parámetros utilizando la función *lsqnonlin*. Esta función permite obtener una mejor estima de los parámetros, si bien dicha función forma parte del toolbox de *Optimización de Matlab*, por lo que si no se dispone de ese toolbox no se puede ejecutar. En el caso de que se disponga de él, se debe comentar la sentencia que realiza la optimización mediante la función *fminsearch* y usar la sentencia con *lsqnonlin*. En este caso la función *fun6* se utiliza para calcular sólo los residuos o diferencia entre los

valores observados y los predichos por el modelo, ya que la propia función *lsqnonlin* se encarga de elevar los residuos al cuadrado y sumarlos. Notar que dado que los valores observados de la variable *y* tienen una componente aleatoria, cada ejecución de la función *sc_regr* genera resultados diferentes. El promedio de los resultados obtenidos con muchas observaciones sería la mejor estima para el vector de parámetros.

El ejemplo de la sección 1.4.3 y la figura 1.10 se obtienen con Matlab con el archivo *sc_fdp.m*. El archivo genera un conjunto de observaciones mediante una función densidad de probabilidad Normal o Gaussiana y determina el valor medio y la desviación estándar de la *fdp* mediante minimización del riesgo empírico (1.50) que realiza a través de la función de optimización de Matlab *fminsearch*. Esta función utiliza *fun7.m* para evaluar el valor del riesgo empírico. Al igual que en el ejemplo anterior, al ser las muestras aleatorias, cada ejecución de *sc_fdp* da una *fdp* distinta. La mejor estima sería la media obtenida tras un número elevado de ejecuciones, de forma que cuanto mayor mejor.