

## PRÁCTICAS DE ORDENADOR V

Para los ejercicios que siguen, es conveniente **familiarizarse** primero con el contenido de los programas `normalR.txt`, `readnormalR.R` y `cltR.txt`, que se encuentran en `Software5.zip`.

[1] Obtener una **muestra aleatoria** de tamaño  $n = 250$  de las siguientes distribuciones  $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ :

a)  $p = 1$ ;  $\boldsymbol{\mu} = 1$ ;  $\sigma^2 = 2$ .

b)  $p = 2$ ;  $\boldsymbol{\mu} = (1, 2)^T$ ;  $\boldsymbol{\Sigma} = \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}$ .

c)  $p = 3$ ;  $\boldsymbol{\mu} = (0, 3, -1)^T$ ;  $\boldsymbol{\Sigma} = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 4 & 0 \\ 1 & 0 & 5 \end{pmatrix}$ .

En los apartados a), b) y c), construir una **representación gráfica** de los datos para comprobar las **características elípticas** de la muestra obtenida. Dibujar también las **marginales** en los casos b) y c). ¿**Depende** del **tamaño muestral** la percepción de las **propiedades de simetría** de la muestra? Responder a esta última cuestión analizando el **comportamiento del caso bidimensional** b), para **diferentes valores** de  $n$ .

[2] Usando los **datos** obtenidos del índice Dow Jones en el apartado [3] de la **PRÁCTICA II** con el programa `dowjones10R.txt`, discutir la **normalidad multivariante** de los retornos diarios asociados a los activos  $I = c('MO', 'KO', 'EK', 'HWP', 'INTC', 'MSFT', 'IBM', 'MCD', 'WMT', 'DIS')$ . Justificar **gráficamente** que, si bien la **parte central** de la distribución es simétrica, las **colas no lo son**, y por tanto los **retornos no pueden ser normales**. Una distribución más apropiada sería una **mixtura** de la forma

$$f(\mathbf{x}) = \pi N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) + (1 - \pi) N_p(\mathbf{m}, \boldsymbol{\Gamma}), \quad (1)$$

donde  $0 < \pi < 1$ ,  $\pi \cong 1$ , es la **proporción** de mezclado, que es **desconocida** y que por tanto hay que **estimar**. El **tratamiento analítico** completo de la distribución de (1) es **complejo**, y queda fuera de los propósitos de este curso.

[3] Como **alternativa** al modelo de (1), el programa `curveR.txt` de `Software5.zip` contiene un análisis de la **distribución** de los **retornos diarios de Microsoft** ('MSFT'), comparando el **modelo usual**  $N(\mu, \sigma^2)$  con uno de **posición** y **escala** de la forma

$$Y = \sigma X + \mu,$$

donde  $X \sim t_k$  sigue una **distribución t de Student** con  $k$  grados de libertad. **Comparar brevemente** las propiedades de ambos modelos.

[4] El paquete ISLR de **R** contiene el conjunto de datos `Smarket`, que se refiere a los **retornos diarios en porcentaje** del índice bursátil S&P 500 entre 2001 y 2005. Comprobar **empíricamente** que los datos del vector aleatorio

$$\mathbf{X} = (\text{Lag1}, \text{Lag2}, \text{Lag3}, \text{Lag4}, \text{Lag5}, \text{Today})^T \quad (2)$$

tienen, aproximadamente, las características de una distribución  $N_6(\mathbf{0}, a\mathbf{I}_6)$ , donde  $a > 1$ . ¿Está esta propiedad relacionada con la **eficiencia** del **mercado**? Razonar la respuesta. Comparar estos resultados con los obtenidos al analizar los datos del vector aleatorio de (2) relativos al conjunto de datos `Weekly`, que se refiere a los retornos **semanales** en porcentaje del índice S&P 500 entre los años 1990 y 2010.