

## A) Robustez.

1. Consideremos una v.a.  $X$  procedente de la distribución “contaminada”

$$F \sim (1 - \varepsilon)N(\mu, \sigma^2) + \varepsilon N(\mu, c^2\sigma^2),$$

es decir, con probabilidad  $1 - \varepsilon$ ,  $X$  se distribuye  $N(\mu, \sigma^2)$  y con probabilidad  $\varepsilon$  se distribuye  $N(\mu, c^2\sigma^2)$ .

- (a) Calcula la esperanza y la varianza de  $X$ .
- (b) Sea  $\bar{X}_1$  la media muestral calculada a partir de una m.a.s. de una distribución normal  $N(\mu, c^2\sigma^2)$  y  $\bar{X}_2$  la media muestral calculada a partir de una m.a.s. de la distribución contaminada. Calcula la eficiencia relativa entre las dos medias e interpreta intuitivamente el resultado obtenido.
- (c) Repite el ejercicio sustituyendo las medias muestrales por medianas y aplicando resultados asintóticos.
2. Sea  $\hat{\mu}(x_1, \dots, x_n)$  el M-estimador de posición dado por

$$\sum_{i=1}^n \psi((x_i - \hat{\mu})/\sigma) = 0$$

donde  $\sigma$  es conocido. Probar que tiene las siguientes propiedades

- (i)  $\hat{\mu}(x_1 + a, \dots, x_n + a) = \hat{\mu}(x_1, \dots, x_n) + a$
- (ii)  $\hat{\mu}(-x_1, \dots, -x_n) = -\hat{\mu}(x_1, \dots, x_n)$
3. Sea  $\hat{\sigma} = MAD(x_1, \dots, x_n)$  probar que

$$\hat{\sigma}(bx_1 + a, \dots, bx_n + a) = |b|\hat{\sigma}(x_1, \dots, x_n)$$

4. Hacer un programa que calcule el  $M$ -estimador para el modelo de posición cuando  $\sigma$  es desconocido.

## B) Ejercicios para hacer en la computadora

1. Generar muestras aleatorias de tamaño 20 de las siguientes distribuciones;

$C_0$ :  $N(0,1)$

$C_1$ : Igual que  $C_0$  cambiando la última observación por un outlier cuyo valor sea a)  $x = 2.5$ , b)  $x = 10$  y c)  $x = 100$

$C_2$ : Cauchy

y calcular los siguientes estimadores:

- a) media
- b) mediana
- c) media podada con  $\alpha = 0.1$  y  $\alpha = 0.25$
- d)  $M$ -estimador con  $\psi_k^H$ ,  $k = 1, 345$ , usando como estimador de dispersión el MAD.

Explicar el comportamiento en cada caso de los distintos estimadores.

2. Los siguientes datos son ppm (partes por millón) de DDT detectado en muestras de coliflores por 15 laboratorios diferentes.

2.79	2.93	3.22	3.78	3.22	3.38	3.18	3.33
3.34	3.06	3.07	3.56	3.08	4.62	3.34	

- (a) Realizar un box-plot de los datos. Detectar posibles outliers
- (b) Calcular los siguientes estimadores y comparar los diferentes resultados
  - Media
  - Mediana
  - Media  $\alpha$ -podada, con  $\alpha = 0.1$  y  $\alpha = 0.25$
- (c) Eliminar los outliers detectados en el box-plot y recalculer los estimadores del inciso anterior. Sacar conclusiones.

3. Los siguientes datos son porcentajes (sobre el peso total declarado) de camarones en latas de cóctel de camarones, medidos por 18 laboratorios diferentes.

32.2	33.0	30.8	33.8	32.2	33.3	31.7	35.7	32.4
31.2	26.6	30.7	32.5	30.7	31.2	30.3	32.3	31.7

- (a) Realizar un box-plot de los datos e identificar posibles outliers.
- (b) Calcular el desvío estándar y la MAD de los datos. Compararlos y sacar conclusiones.
- (c) Calcular la media y la mediana de los datos. Compararlas y sacar conclusiones.
- (d) Eliminar los outliers detectados en el box-plot y recalculer los estimadores de los incisos anteriores. Analizar los nuevos resultados.