PROBABILIDADES Y ESTADÍSTICA (C)

Práctica 6

1. Consideremos $x_1, \ldots, x_n$ una muestra de una población cualquiera. Sean $\bar{x}$ y $\tilde{x}$ la media y la mediana muestral, respectivamente.

   a) Si se suma una constante c a cada uno de los x_i de la muestra, obteniéndose $y_i = x_i + c$, ¿cómo se relacionan $\bar{x}$ con $\bar{y}$ y $\tilde{x}$ con $\tilde{y}$?

   b) Si cada x_i es multiplicado por una constante c, obteniéndose $y_i = cx_i$, responder a la pregunta planteada en (a).

2. Sea s_X^2 la varianza muestral correspondiente a la muestra $x_1, \ldots, x_n$. Demostrar que:

   a) $s_X^2 = \frac{1}{n-1} \sum_{i=1}^n x_i^2 - \frac{n}{n-1} \bar{x}^2$.

   b) Si $y_i = x_i + c$, con c constante, entonces $s_Y^2 = s_X^2$.

   c) Si $y_i = cx_i$, con c constante, entonces $s_Y^2 = c^2 s_X^2$.

3. Sea $x_1, \ldots, x_n$ una muestra de una población con media μ.

   (a) ¿Para qué valores de c se minimiza $\sum_{i=1}^n (x_i - c)^2$?
   (Sugerencia: derivar con respecto a c).

   (b) Usando (a) decidir cuál de estas dos cantidades es más pequeña:

   $$\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \qquad \text{o} \qquad \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 .$$

4. Dada una muestra $x_1, ..., x_n$, se define la siguiente medida de dispersión:

   $$\mathrm{MAD}\,(x_1, \ldots, x_n) = \mathrm{med}_{1 \le i \le n} |x_i - \tilde{x}|$$

   a) Sea c una constante. Si definimos $y_i = x_i + c$, ¿cuál es la relación entre $\mathrm{MAD}\,(x_1, \ldots, x_n)$ y $\mathrm{MAD}\,(y_1, \ldots, y_n)$?

   b) Responder (a) para $y_i = cx_i$.

5. A partir de una muestra $x_1, ..., x_n$ se calculan la media y el desvío estándar muestrales, $\bar{x}$ y s_X respectivamente. Si definimos $y_i = (x_i - \bar{x}) / s_X$, ¿cuánto valen $\bar{y}$ y s_Y? Interpretar este resultado.

6. Se analizó una muestra de 12 piezas de pan blanco de cierta marca y se determinó el porcentaje de carbohidratos contenido en cada una de las piezas, obteniéndose los siguientes valores:

   76.93 76.88 77.07 76.68 76.39 75.09 77.67 76.88 78.15 76.50 77.16 76.42

   a) Estimar la media del porcentaje de carbohidratos contenido de las piezas de pan de esta marca.

b) Estimar la probabilidad de que el porcentaje de carbohidratos de una piezas de pan de esta marca no excede el 76.5%.

7. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución $N(\mu, \sigma^2)$. Obtener los estimadores de máxima verosimilitud (EMV) de

   a) μ y σ^2.

   b) μ, siendo $\sigma^2 = \sigma_0^2$ conocida.

   c) σ^2, siendo $\mu = \mu_0$ conocida.

8. a) Una máquina envasa caramelos, siendo el peso neto (en gramos) de cada bolsa una v.a. con distribución normal. Los siguientes datos corresponden al peso de 15 bolsas elegidas al azar:

210	197	187	217	194	208	220	199	193	203	181	212	188	196	185

   Hallar los EMV de la media y la varianza del peso neto.

   b) Con cierto instrumento se realizan 20 mediciones de una magnitud física μ. Cada observación es de la forma $X = \mu + \epsilon$, donde ϵ es el error de medición (aleatorio). Se obtuvieron los siguientes datos:

25.11	25.02	25.16	24.98	24.83	25.05	24.94	25.04	24.99	24.96
25.03	24.97	24.93	25.12	25.01	25.12	24.90	24.98	25.10	24.96

   Suponiendo que los errores de medición tienen distribución normal con media cero y varianza 0.01, estimar μ. ¿Cuál es la varianza del estimador de μ?

   c) Para controlar la precisión de un sistema de medición se mide 24 veces una magnitud conocida $\mu_0 = 12$, obteniéndose los siguientes valores

12.51	11.66	11.91	12.25	11.54	11.36	12.40	12.19	12.88	12.16	12.69	12.91
12.12	11.02	12.53	11.77	12.72	10.56	11.52	11.66	12.25	12.09	11.48	12.36

   Estimar la precisión (es decir, la varianza del error de medición), suponiendo que los errores están normalmente distribuídos con media cero.

9. Consideremos muestras aleatorias $X_1, \ldots, X_n$ para cada una de las siguientes distribuciones:

   i) Poisson de parámetro θ.

   ii) exponencial de parámetro θ.

   iii) con densidad

$$f(x;\theta) = \frac{1}{\theta} x^{\left(\frac{1}{\theta}-1\right)} I_{[0,1]}(x), \quad 0 < \theta < 1.$$

   iv) geométrica de parámetro θ.

   a) Encontrar en cada caso el estimador de máxima verosimilitud de θ y el estimador obtenido por el método de momentos y ver si coinciden.

b) Decidir si los estimadores obtenidos en i) y ii) son insesgados o asintóticamente insesgados.

c) Demostrar que los estimadores obtenidos en i) a iv) son consistentes.

10. a) Durante 20 días se ha registrado el número de llamadas en una central telefónica, obteniéndose los siguientes valores:

35	41	38	40	34	36	41	48	42	39	57	41	35	37	38	41	43	44	46	47

Suponiendo que el número de llamadas diarias sigue una distribución $\mathcal{P}(\theta)$, estimar el promedio diario de llamadas.

b) Se sabe que el tiempo de duración de una clase de lámparas tiene distribución $\mathcal{E}(\theta)$. Se han probado 20 lámparas, obteniéndose los siguientes tiempos de duración (en días):

45	53	50	61	39	40	45	47	38	53	54	60	34	46	34	50	42	60	62	50

Estimar el tiempo esperado de la duración de una lámpara.

c) Se sabe que la longitud de los ejes que fabrica un establecimiento siderúrgico tiene densidad

$$f(x;\theta) = \frac{1}{\theta} x^{\left(\frac{1}{\theta}-1\right)} I_{[0,1]}(x).$$

Se eligen al azar 20 ejes, cuyas longitudes son:

0.5	0.7	0.8	0.95	0.9	0.6	0.2	0.85	0.3	0.2
0.76	0.55	0.48	0.8	0.76	0.13	0.15	0.67	0.9	0.95

Estimar el valor del parámetro θ.

d) Un estado tiene varios distritos. Supongamos que cada distrito tiene igual proporción θ de personas que están a favor de una propuesta de control de armas. En cada uno de 8 distritos elegidos al azar, se cuenta la cantidad de personas que hay que encuestar hasta encontrar alguna de acuerdo con la propuesta (llamemos X a esta cantidad). Los resultados son: 3, 8, 9, 6, 4, 5, 3, 2 (i.e.: en el primer distrito las dos primeras personas encuestadas estaban en contra y la tercera a favor). Basándose en estos datos, calcular el EMV de $P_\theta(X \geq 5)$.

11. Sea $X_1, \ldots X_n$ una muestra aleatoria de una distribución $\mathcal{U}[0,\theta]$.

a) Probar que $T = \max\limits_{1\leq i\leq n} X_i$ es el estimador de máxima verosimilitud de θ.

b) Calcular el estimador de θ basado en el primer momento.

c) Decir si los estimadores obtenidos son insesgados o asintóticamente insesgados, y consistentes. Justificar.

12. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución con densidad

$$f(x;\theta) = e^{-(x-\theta)} I_{[\theta,\infty)}(x).$$

a) Probar que $T = \min_{1 \leq i \leq n} X_i$ es el estimador de máxima verosimilitud de θ.

b) Calcular el estimador de θ basado en el primer momento.

c) Decir si los estimadores obtenidos son insesgados o asintóticamente insesgados, y consistentes. Justificar.

13. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución con media μ y varianza σ^2.

    a) Probar que $\bar{X}^2$ no es un estimador insesgado de μ^2. ¿Es asintóticamente insesgado?¿Es consistente?

    b) ¿Para qué valores de k es $\widehat{\mu^2} = \left(\bar{X}^2 - ks^2\right)$ un estimador insesgado de μ^2?

14. Se define el *error cuadrático medio* de un estimador $\hat{\theta}$ como

$$\mathrm{ECM}(\hat{\theta}) = E\left[\left(\hat{\theta} - \theta\right)^2\right]$$

    a) Verificar que $\mathrm{ECM}(\hat{\theta}) = V(\hat{\theta}) + \left(\mathrm{sesgo}(\hat{\theta})\right)^2$, donde $\mathrm{sesgo}(\hat{\theta}) = E(\hat{\theta}) - \theta$.

    b) ¿Cuánto vale $\mathrm{ECM}(\hat{\theta})$ si $\hat{\theta}$ es un estimador insesgado de θ?

    c) Consideremos un estimador de la varianza de la forma $\hat{\sigma}^2 = ks^2$, siendo s^2 la varianza muestral. Hallar el valor de k que minimiza $\mathrm{ECM}(\hat{\sigma}^2)$.

    (Sugerencia: Usar que $E(s^4) = \dfrac{n+1}{n-1}\sigma^4$).

15. En el Ejercicio 11, calcular el ECM de los estimadores calculados en (a) y (b) y compararlos. En función de esta comparación, ¿cuál de los dos estimadores usaría?

16. Sea $X_1, \ldots, X_n$ una muestra aleatoria con distribución de Bernoulli de parámetro p y sea $T_n = \sum_{i=1}^{n} X_i$. Consideremos el nuevo parámetro $\theta = p\,(1-p)$.

    a) Mostrar que $T_n\,(n - T_n)\,/\,(n(n-1))$ es un estimador insesgado de θ.

    b) Hallar el EMV de θ.

    c) Mostrar que el EMV de θ es sesgado, pero asintóticamente insesgado.

    d) Mostrar que el estimador insesgado dado tiene mayor varianza que el EMV para θ.