PROBABILIDADES Y ESTADÍSTICA (C)

Práctica 7

1. Se analizaron 12 piezas de pan blanco de cierta marca elegidas al azar. 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 esperanza del porcentaje de carbohidratos contenido en una pieza de pan de esta marca.

   (b) Estimar la mediana del porcentaje de carbohidratos contenido en una pieza de pan de esta marca.

   (c) Estimar la probabilidad de que el porcentaje de carbohidratos de una pieza de pan de esta marca no exceda el 76.5%.

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

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

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

   (c) el par (μ, σ^2) simultáneamente.

3. (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

   Usando el método de MV, indique el valor estimado de la media y la varianza del peso neto de cada bolsa de caramelos envasada por esta máquina.

   (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 valores:

   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 conocer 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.

4. Se sabe que el tiempo de duración de una clase de lámparas tiene distribución $\mathcal{E}(\theta)$. Consideremos el siguiente experimento. Se seleccionan n lámparas al azar y se registran sus tiempos de duración: $X_1, \ldots, X_n$.

   (a) Calcular el estimador de momentos y de MV de θ y del percentil 90 (o cuantil 0.9) del tiempo de duración de una lámpara bajo el modelo propuesto.

   (b) Decidir si estos estimadores son insesgados o asintóticamente insesgados.

   (c) Decidir si estos estimadores son consistentes. Justificar

   (d) Se han probado 20 lámparas, obteniéndose los siguientes tiempos de duración (en días):

39.08	45.27	26.27	14.77	65.84	49.64	0.80	66.58	69.60	32.42
228.36	64.79	9.38	3.86	37.18	104.75	3.64	104.19	8.17	8.36

      i. Basándose en estas observaciones, estime θ y el percentil 90 (o cuantil 0.9) del tiempo de duración de una lámpara bajo el modelo propuesto usando los estimadores hallados en el ítem (a).

      ii. ¿Cómo haría para estimar el percentil solicitado si no conociera la distribución de las X_i?

   (e) En un nuevo experimento se registra la duración de 350 lámparas. Los valores observados se encuentran en el archivo `lamparas2.txt`. Repita el ítem (d) para este nuevo conjunto de datos.

5. El número de llamadas que recibe una central telefónica en un día es una variable aleatoria con distribución $\mathcal{P}(\theta)$. Se desea estimar la probabilidad de que en un determinado día se reciban exactamante 40 llamadas. Para ello se registrarán la cantidad de llamadas recibidas durante 48 días.

   (a) Basándose en la muestra aleatoria $X_1, \ldots, X_n$, hallar el estimador de momentos y de máxima verosimilitud de θ.

   (b) Deducir un estimador para la probabilidad de que en un determinado día se reciban 40 llamadas.

   (c) ¿Cómo haría para estimar la probabilidad deseada si no conociera la distribución de las X_i?

(d) Durante 48 días se registran las siguientes cantidades

40	46	46	45	42	44	50	31	41	42	50	34	62	32	46	50
39	42	44	41	47	42	41	50	34	47	38	40	44	45	35	51
38	41	39	34	48	35	40	40	43	36	40	49	45	47	34	45

Estimar θ. Estimar la probabilidad de que en un determinado día se reciban exactamente 40 llamadas utilizando los estimadores propuestos en los items b) y c).

(e) En un nuevo experimento se registra la cantidad de llamados en 200 días. Los valores observados se encuentran en el archivo `llamadas.txt`. Repita el item anterior para este nuevo conjunto de datos.

6. Se sabe que la longitud de los ejes que fabrica un establecimiento siderúrgico tiene densidad
$$f(x;\theta) = \frac{1}{\theta} x^{\left(\frac{1}{\theta}-1\right)} I_{(0,1)}(x).$$

(a) Hallar el estimador de MV, $\widehat{\theta}_{MV}$, y el estimador de momentos, $\widehat{\theta}_{M}$, de θ.

(b) Decidir si $\widehat{\theta}_{MV}$ es insesgado o asintóticamente insesgado.

(c) Estudiar la consistencia de ambos estimadores.

(d) 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 θ usando cada uno de los métodos propuestos.

7. 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 variable aleatoria). 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).

(a) ¿Qué distribución tiene X?

(b) Estimar θ por el método de máxima verosimiltud y por el método de los momentos.

(c) Estimar por máxima verosimilitud $P_\theta(X \geq 5)$.

(d) Estimar $P_\theta(X \geq 5)$ sin asumir que X tiene distribución geométrica. Comparar el resultado con el del ítem anterior.

hasta Ejer 7 solo estimadores sin chequear propiedades

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

(a) Probar que $M_n = \max\limits_{1\leq i\leq n} X_i$ 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.

9. Sea $X_1, \dots X_n$ una muestra aleatoria de una distribución Rayleigh, cuya densidad está dada por
$$f(x;\theta) = \frac{x}{\theta} e^{-\frac{x^2}{2\theta}} I_{[0,\infty)}(x).$$

(a) Hallar el estimador de máxima verosimilitud de θ.

(b) Decir si el estimador obtenido es insesgado o asintóticamente insesgado, y consistente. Justificar.
(Sugerencia: Hallar la distribución de la v.a. X_1^2).

10. Sea $X_1, \dots, X_n$ una muestra aleatoria de una distribución con densidad
$$f(x;\theta) = e^{-(x-\theta)} I_{[\theta,\infty)}(x).$$

(a) Probar que $m_n = \min\limits_{1 \le i \le n} X_i$ 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.

11. Sea $X_1, \dots, X_n$ 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?

12. La cantidad de días de internación de los pacientes que llegan a la guardia de un hospital es una variable aleatoria Y con distribución binomial negativa con media μ y parámetro de dispersión r. Esto quiere decir que
$$\mathrm{P}(Y = k) = \frac{\Gamma(r+k)}{k!\Gamma(r)} \left(\frac{r}{r+\mu}\right)^r \left(\frac{\mu}{r+\mu}\right)^k \text{ para } k = 0, 1, 2, \dots$$

Sabendo que $E(Y) = \mu$ y $V(Y) = \mu + \mu^2/r$, calcular el estimador de momentos de (μ, r). ¿Es posible calcular explícitamente el estimador de MV de (μ, r)?

Se toman 1000 historias clínicas al azar de pacientes que llegaron a la guardia del hospital y se registra la cantidad de días de internación de cada uno. Los valores obtenidos se encuentran en el archivo `dias.txt`. Indique el valor estimado de (μ, r) para cada uno de los estimadores propuestos. Para el estimador de MV, utilice el comando `fitdistr(x, densfun = "negative binomial")` del paquete `MASS`.

13. 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 $E(s^4) = \dfrac{n+1}{n-1}\sigma^4$).

14. En el Ejercicio 8, calcular el ECM de los estimadores calculados en (a) y (b) y compararlos. En función de esta comparación, ¿cuál de los dos estimadores usaría?

15. Sea $X_1, \ldots, X_n$ una muestra aleatoria con distribución de Bernoulli de parámetro p y consideremos el nuevo parámetro $\theta = p\,(1-p)$.

(a) Proponer un estimador de θ basado en el estimador de MV de p.

(b) Mostrar que el estimador de θ propuesto en a) es sesgado pero asintóticamente insesgado.

(c) Proponer un estimador insesgado de θ.

(Sugerencia: multiplique por una constante adecuada al estimador propuesto en a)

d) ¿Cuál de los dos estimadores tiene menor varianza?

16. **(Optativo)** Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución con $E(X_i) = \mu$ y $V(X_i) = \sigma^2$ y sea $Y_1, \ldots, Y_m$ otra muestra aleatoria independiente de la anterior con $E(Y_i) = \mu$ y $V(Y_i) = 4\sigma^2$.

(a) Mostrar que para todo $\delta \in (0,1)$, el estimador $\hat{\mu} = \delta\bar{X} + (1-\delta)\bar{Y}$ es un estimador insesgado de μ.

(b) Para valores fijos de n y m, calcular el ECM del estimador propuesto en a).

(c) Hallar el valor de δ que minimiza el ECM.