Estimación puntual

A) Estimadores basados en los momentos.

1. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución

   - $Bi\,(1, \theta)\,,\ 0 < \theta < 1.$
   - $\mathcal{P}\,(\theta)\,,\ \theta > 0.$
   - $\mathcal{E}\,(\theta)\,,\ \theta > 0.$
   - $\mathcal{G}\,(\theta)\,,\ 0 < \theta < 1.$

   En cada uno de estos casos, encontrar:

   (a) Un estimador de θ basado únicamente en el primer momento.

   (b) Un estimador de θ basado únicamente en el segundo momento.

   Verificar que los respectivos estimadores cumplan las restricciones impuestas sobre el parámetro. Mas precisamente, si $\delta_n(X_1, \ldots, X_n)$ es el estimador resultante, verificar que $\delta_n(X_1, \ldots, X_n) \in \bar{\Theta}$

2. (Para hacer en R) El número de partículas que emite una fuente radiactiva por segundo, constituye un proceso de Poisson de intensidad λ.
   Se recogieron durante 15 minutos los siguientes valores de emisiones:

   $$\begin{array}{l} 500 - 488 - 426 - 510 - 450 - 368 - 508 - 514 - 426 \\ 476 - 512 - 526 - 444 - 524 - 236 \end{array}$$

   Estimar el valor del parámetro λ basado en la muestra utilizando un estimador de momentos basado en el momento de orden 1 y otro basado en el momento de orden 2. Mediante un *bootstrap* paramétrico, obtenga una aproximación de la distribución de los estimadores obtenidos (grafique la función empírica de distribución o bien obtenga un histograma con las realizaciones del estimador). Utilizando también *bootstrap* estime la varianza del estimador.

3. Sea $X_1, \ldots, X_n$ una m.a. de una distribución $\mathcal{E}\,(\theta)$.

   (a) Hallar un estimador de los momentos de $q\,(\theta) = P_\theta\,(X_1 \geq 1)\,.$

   (b) Deducir de (a) un estimador para θ.

4. Sea $X_1, \ldots, X_n$ una m.a. de una distribución $N\,(0, \sigma^2)\,,$ con $\sigma > 0$.

   (a) Encontrar un estimador de σ basado en el segundo momento.

   (b) Encontrar otro estimador de σ a partir de $E_\sigma\,(|X_1|)\,.$

5. Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes e idénticamente distribuidas Hallar el estimador de los momentos de μ si

   (a) la densidad de X_1 está dada por:

   $$f\,(x) = \frac{1}{\mu^4} x I_{[0,\mu^2]}\,(x) + \frac{1}{\mu^4}(2\mu^2 - x) I_{[\mu^2, 2\mu^2]}\,(x) \qquad \mu > 0$$

   (b) la función de probabilidad de X_1 está dada por:

k	1	2	4
$p\,(k)$	μ^2	$2\mu\,(1-\mu)$	$(1-\mu)^2$

   $0 < \mu < 1$

B) Estimadores de máxima verosimilitud.

1. Dada una muestra aleatoria $X_1, \ldots, X_n$ para cada una de las distribuciones consideradas en el ejercicio A.1, encontrar los respectivos estimadores de máxima verosimilitud (EMV) de θ.

2. Sea $X_1, \ldots, X_n$ una m.a. de una distribución $N(\mu, \sigma^2)$. Consideremos el parámetro bivariado $\theta = (\mu, \sigma^2)$.

   (a) Encontrar el EMV de θ.

   (b) Dado $\xi \in I\!R$, encontrar el EMV de $p = P_\theta(X_1 > \xi)$.

   (c) Dado $p \in (0,1)$, encontrar el EMV del valor ξ tal que $P_\theta(X_1 > \xi) = p$.

   (d) Hallar el EMV de μ cuando σ^2 es conocido. ¿Es razonable que no dependa de σ^2?

   (e) Hallar el EMV de σ^2 cuando μ es conocido. ¿Es razonable que dependa de μ?

   (f) Se supone que la distribución de un índice de colesterol en cierta población es $N(\mu, \sigma^2)$, donde μ y σ^2 son parámetros desconocidos. Se hacen análisis a 25 personas elegidas al azar de la población y se obtienen los siguientes datos:

1.53	1.65	1.72	1.83	1.62	1.75	1.72	1.68	1.65	1.61
1.70	1.60	1.73	1.61	1.52	1.81	1.72	1.50	1.51	1.65
1.58	1.82	1.65	1.72	1.65					

   i) Estimar μ y σ^2 por máxima verosimilitud basados en la muestra dada.

   ii) Se considera que el índice es normal si es menor que 1.73. Estimar la proporción de la población con un índice anormal.

   iii) (Para hacer en R) Como se ha hecho en el ejercicio 2 de de la sección de momentos, mediante un *bootstrap* paramétrico estime la distribución y la varianza del estimador obtenido.

3. Si $X_1, \ldots, X_n$ es una m.a. de una distribución $\mathcal{U}[0, \theta]$

   (a) Hallar el EMV de θ.

   (b) Hallar el estimador de los momentos de θ.

4. Sea $X_1, \ldots, X_n$ una m.a. de una distribución exponencial desplazada, cuya densidad es

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

   (a) Encontrar el EMV de θ.

   (b) Encontrar el estimador de los momentos de θ.

5. Si $X_1, \ldots, X_n$ es una m.a. de una distribución $\mathcal{U}[\theta - 1/2, \theta + 1/2]$, mostrar que cualquier T tal que $X_{(n)} - 1/2 \leq T \leq X_{(1)} + 1/2$ es un EMV de θ.

6. Sean $X_1, \ldots, X_n$ una m.a. de una distribución $N(\mu_1, \sigma^2)$ e $Y_1, \ldots, Y_m$ una m.a. de una distribución $N(\mu_2, \sigma^2)$, independientes entre sí.

   (a) Hallar el EMV de $\theta = (\mu_1, \mu_2, \sigma^2)$.

   (b) Hallar el EMV de $\alpha = \mu_1 - \mu_2$.

7. Se tienen observaciones independientes $X_1, \ldots, X_n$ de poblaciones normales con la misma media μ pero con varianzas $\sigma_1^2, \ldots, \sigma_n^2$ respectivamente.

   (a) ¿Es posible estimar todos los parámetros por máxima verosimilitud?

   (b) Suponiendo $\sigma_1^2, \ldots, \sigma_n^2$ conocidos, hallar el EMV de μ. Interpretar.

8. El número de microorganismos que se encuentran por grupo sobre una superficie sigue la siguiente distribución

$$p(x) = \theta I_{\{1\}}(x) + \frac{1-\theta}{k-1} I_{\{2,3,\ldots,k\}}(x)$$

donde $0 < \theta < 1$. Supongamos que se examinan en forma independiente n grupos y se cuentan el número de microorganismos $X_1, \ldots, X_n$ en cada uno de ellos. Calcular el estimador de máxima verosimilitud de θ.

C) Estimadores de mínimos cuadrados.

En todos los ejercicios de esta sección se supondrá que $\varepsilon_1, \ldots, \varepsilon_n$ son v.a.i.i.d. con media cero.

1. Demostrar que la media muestral $\bar{X}$ es el estimador de mínimos cuadrados (EMC) de θ en el modelo de posición: $X_i = \theta + \varepsilon_i$, $1 \leq i \leq n$.

2. Consideremos el modelo de regresión lineal simple: $X_i = \theta_1 t_i + \theta_2 + \varepsilon_i$, $1 \leq i \leq n$, donde t_i son constantes conocidas. Hallar los EMC de θ_1 y θ_2.

3. Sea el modelo $X_i = S_i(\theta) + \epsilon_i$, $1 \leq i \leq n$, $\theta \in I\!R^k$. Si $\varepsilon_i \sim N(0, \sigma^2)$, mostrar que el EMC y el EMV de θ coinciden.

   Sugerencia: Notar que $X_i \sim\sim N(S_i(\theta), \sigma^2)$.

4. Consideremos nuevamente el modelo $X_i = S_i(\theta) + \epsilon_i$, $1 \leq i \leq n$, $\theta \in I\!R^k$. Se define el estimador de *mínimos valores absolutos* (EMVA) como aquel valor $\widehat{\theta}$ que minimiza

$$\sum_{i=1}^{n} |X_i - S_i(\theta)|$$

   (a) Mostrar que el EMVA coincide con el EMV si los ε_i tienen distribución doble exponencial, cuya densidad es

$$f(t) = \frac{1}{2} e^{-|t|}$$

   Sugerencia: Hallar la distribución de X_i para $1 \leq i \leq n$.

   (b) Mostrar que la mediana de $X_1, \ldots, X_n$ es el EMVA en el modelo de posición. *Sugerencia:* Considere primero el caso n impar.

5. (Para hacer en R) Tenemos el modelo real $Y_i = X_i + \epsilon_i$, con $X_i \sim U(0,1)$ y $\epsilon_i \sim N(0,1)$. El modelo tentativo será $Y_i = \beta X_i + \epsilon_i$. Fije $n = 20$, genere los datos siguiendo el modelo real, grafíquelos y mediante una regresión de mínimos cuadrados sobre el modelo tentativo estime el valor de β. Usando regresión por medianas, efectúe también una estimación de β. Sea $\widehat{Y}_i = \widehat{\beta} X_i$, grafique $(X_i, \widehat{Y}_i)$ en ambos casos.

   Ahora, genere nuevamente usando el modelo real los (X_i, Y_i) pero vamos a introducir una contaminación en la primer observación. Cambiaremos el valor de X_1 por 2, 10, 50 y 100 (niveles de contaminación) e $Y_1 = -30$. Para cada valor nuevo para X_1, regenere nuevamente la estimación usando ambos métodos y nuevamente grafique $(X_i, \widehat{Y}_i)$. Cómo se compara el comportamiento de ambos métodos frente a los distintos valores de contaminación?