ESTADÍSTICA I
Grado en Matemáticas (2023/24)

Tema 4: Intervalos de confianza

4.1. Consideramos la v.a. X = cantidad de contaminación por mercurio (en p.p.m.) en los peces capturados en los ríos norteamericanos Lumber y Wacamaw (fichero `Datos-mercurio.txt`).

a) Representa un estimador de la función de densidad de X. Compara esta densidad estimada con la densidad normal de igual media y desviación típica (representada en la misma gráfica). En vista de las dos funciones ¿dirías que la función de densidad de X es aproximadamente normal?
b) Calcula un intervalo de confianza de nivel 0.95 para la media de X. ¿Se puede considerar fiable este intervalo a pesar de la posible no-normalidad de X?
c) ¿Qué tamaño muestral habría que tomar para estimar la contaminación media con un margen de error máximo de 0.06?
d) Calcula ahora un intervalo de confianza de nivel 0.99 para la media de X. Compáralo con el intervalo obtenido en (b).

4.2. En un estudio sobre el efecto del hábito de fumar en la agregación de plaquetas en la sangre (que puede dar lugar a la formación de coágulos) se extrajeron muestras de sangre de 11 individuos antes y después de fumar un cigarrillo, y se midió el máximo porcentaje de plaquetas agregadas. Los resultados fueron:

Individuo	1	2	3	4	5	6	7	8	9	10	11
Antes	25	25	27	44	30	67	53	53	52	60	28
Después	27	29	37	56	46	82	57	80	61	59	43

Bajo hipótesis de normalidad, calcula un intervalo de confianza de nivel 0.95 para la diferencia media μ del máximo porcentaje de plaquetas agregadas antes y después de fumar un cigarrillo. ¿Qué conclusión se obtiene del resultado obtenido?
Indicación: Se trata de *datos emparejados*. Un procedimiento habitual en este caso consiste en suponer que la variable D, la diferencia entre el máximo porcentaje de plaquetas agregadas antes y después de fumar un cigarrillo, sigue una distribución $N(\mu, \sigma)$, con σ desconocido.

4.3. a) Utilizando el fichero `Datos-lipidos.txt`, estima, mediante un intervalo de confianza de nivel 0.95, la proporción de pacientes que tienen una concentración de colesterol superior o igual a 220 mg/dl. ¿Qué tamaño muestral habría que usar para tener una probabilidad aproximada de 0.95 de no cometer un error mayor que 0.01 en la estimación de esta proporción?.
b) Suponiendo que la distribución de la variable "concentración de colesterol" fuese aproximadamente normal, obtener un estimador puntual para la proporción indicada en (a).

4.4. Sea una v.a. con función de densidad $f(x;\theta) = \theta x^{-(\theta+1)} \mathbb{1}_{[1,\infty)}(x)$.

a) Obtener el estimador de máxima verosimilitud de θ.
b) Obtener su distribución asintótica.
c) Calcular una cantidad pivotal aproximada y, a partir de ella, un intervalo de confianza de nivel aproximado $1-\alpha$ para θ.

4.5. Sea $X_1, \ldots, X_n$ una muestra aleatoria simple de una distribución uniforme en el intervalo $(\theta, 1)$, donde $\theta < 1$.

a) Calcula el estimador de máxima verosimilitud de θ.
b) Modifica el estimador de (a) para obtener una cantidad pivotal y un intervalo de confianza de nivel $1-\alpha$ para θ.

4.6. Construye tres intervalos de confianza asintóticos diferentes para el parámetro λ de una distribución de Poisson usando los tres métodos siguientes:

a) Utiliza el comportamiento asintótico de la media muestral $[\sqrt{n}(\bar{X}-\lambda) \xrightarrow{d} N(0,\sqrt{\lambda})]$, estima de forma consistente la varianza y aplica el teorema de Slutsky.
b) Igual que el anterior, pero sin estimar la varianza.
c) Aplicando el método delta para "estabilizar la varianza", es decir, buscando una función g tal que $\sqrt{n}(g(\bar{X})-g(\lambda)) \xrightarrow{d} N(0,1)$

4.7. Sean $\theta > 0$ y X una variable aleatoria con función de densidad

$$f(x) = \frac{2(\theta - x)}{\theta^2}, \qquad \text{para } 0 \leq x \leq \theta.$$

a) Calcula la función de distribución de X/θ y comprueba que X/θ es una cantidad pivotal.
b) Utiliza esta cantidad pivotal para encontrar un intervalo de confianza de nivel $1-\alpha$ para θ basado en la única observación X.
c) Determina un estimador insesgado de θ basado en la única observación X.

4.8. Se desea estudiar la influencia de la hipertensión en los padres sobre la presión sanguínea de los hijos. Para ello se seleccionan dos grupos de niños, unos con padres de presión sanguínea normal (grupo 1) y otros con uno de sus padres hipertenso (grupo 2), obteniéndose las siguientes presiones sistólicas:

Grupo 1	104	88	100	98	102	92	96	100	96	96
Grupo 2	100	102	96	106	110	110	120	112	112	90

Halla un intervalo de confianza para la diferencia de medias, suponiendo que la distribución de la presión sistólica es normal y que las varianzas en las dos poblaciones de niños son iguales.

4.9. Sea $X_1, \ldots, X_n$ una muestra de una distribución normal de media μ y varianza θ. Estamos interesados en estimar θ. Sea $S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)$ la varianza muestral insesgada. Calcular $V(S^2)$ y compararla con la cota de Fréchet-Cramer-Rao.

4.10. a) Sea X_n una v.a. con distribución χ^2_n. Determina hacia dónde convergen en distribución las sucesiones $\sqrt{n/2}(X_n/n - 1)$ y $\sqrt{2X_n} - \sqrt{2n}$.
b) Sea X una v.a. con distribución $\gamma(5, 10)$. Calcular $\mathbb{P}\{X \leq 3\}$ usando las tablas de la distribución χ^2.
Indicación: Comprobar que, si ξ es una v.a. $\gamma(a, p)$ con densidad

$$g(t) = \frac{a^p}{\Gamma(p)} e^{-at} t^{p-1}, \qquad \text{para } t > 0,$$

y $c > 0$ es una constante, entonces $c\xi$ es una v.a. $\gamma(a/c, p)$.
c) Sea $Y \sim \chi^2_{100}$. Calcular aproximadamente $\mathbb{P}\{Y \leq 90\}$ usando las tablas de la $N(0,1)$.
Indicación: Expresar Y como suma de v.a.i.i.d.
d) Sea $\bar{X}_n$ la media muestral calculada con las n primeras observaciones de una sucesión de v.a.i.i.d. procedentes de una población con esperanza μ y varianza σ^2. Determina hacia dónde converge en distribución la sucesión $n(\bar{X} - \mu)^2$.