Estadística en Física Experimental (1er Cuatrimestre 2023)

Guía de Problemas N° 7 | Estimación puntual de parámetros

Consistencia y Sesgo

1. Considere el estadístico $S^2 = \sum_i^n (x_i - \mu)^2/n$, donde los x_i son variables aleatorias independientes e idénticamente distribuidas y la esperanza $E(x_i) = \mu$ es conocida.
   (a) ¿Puede S^2 ser considerado un *estadístico* dado que no solo es función de las observaciones sino también de los parámetros μ y n?
   (b) Mostrar que es un estimador no sesgado de la varianza de X.
   (c) Encuentre el error de S^2 cuando los x_i son gaussianos.
   (d) ¿Cuánto vale la varianza de S^2 al usar la fórmula de propagación de errores? ¿Porqué falla? [Rta: “Var(S^2)”=0]

2. Muestre que $s^2 = \sum_i^n (x_i - \bar{x})^2/n$ es un estimador sesgado de Var(X), cuyo bias vale $-\sigma^2/n$, mientras que $\tilde{s}^2 = \sum_i^n (x_i - \bar{x})^2/(n-1)$ es no sesgado.

3. Usando la desigualdad de Tshebycheff, muestre que $S^2 = \sum_i^n (x_i - \mu)^2/n$ es un estimador consistente de la varianza cuando la esperanza μ es conocida, para el caso que los $\{x_i\}$ tienen distribución normal.

4. Establezca una condición suficiente para que $S^2 = \sum_i^n (x_i - \mu)^2/n$ sea un estimador consistente de la varianza, cuando los $\{x_i\}$ tienen distribución arbitraria. Intente no hacer más cuentas que en el ejercicio anterior.

5. En general, que t sea un estimador no sesgado de θ, no implica que t^2 sea no sesgado para θ^2.
   (a) Convénzase intuitivamente que ésto es cierto, sin hacer cuentas, para el caso que $\theta = E(x)$ y $t = \bar{x}$, con $f_X(x)$ simétrica alrededor de $x = 0$.
   (b) Sea k una variable aleatoria con distribución binomial $B_k(n, p)$. Muestre que $t = k/n$ es un estimador no sesgado de p, mientras que $t' = (k/n)^2$ no lo es para p^2. Halle el bias de $(k/n)^2$, y a partir de éste encuentre un estimador no sesgado de p^2.

Eficiencia y Mínima varianza

6. Escriba la función verosimilitud para un experimento binomial $B_k(n, p)$, y aplicando la relación de Cramer-Rao muestre que $t = k/n$ es un estimador 100% eficiente de p. ¿Cuánto vale $V(t)$?

7. Muestre que la aplicación de la desigualdad de Cramer-Rao al parámetro σ de la distribución normal $N(\mu, \sigma)$ establece que existe una única función de σ con estimador 100% eficiente, y permite encontrar su estimador, sesgo y varianza. Verifique sus conclusiones aplicando Cramer-Rao al parámetro σ^2.

8. Sea la distribución exponencial, $f(x) = \lambda \exp(-\lambda x)$. Encuentre para que función $h(\lambda)$ existe un estimador 100% eficiente. Muestre que Cramer-Rao permite extraer directamente su sesgo y su varianza.

9. Compruebe que la distribución de Cauchy descentrada $f(x) = 1/[\pi(1 + (x - \mu)^2)]$ no posee un estimador eficiente para μ. ¿Cuál es la cota mínima para un estimador de μ no sesgado, con una muestra de tamaño n?
   Ayuda: $\int_0^\infty t^2/(1 + t^2)^3 dt = \pi/16$

Suficiencia

10. Por definición un estadístico $t(\underline{x})$ es suficiente para un parámetro desconocido θ si la probabilidad condicional de obtener dicha muestra, dado que se conoce $t(\underline{x})$: $P(x_1, \ldots, x_n | t)$, no depende de θ. Suponga una secuencia de mediciones $\{x_1, \ldots, x_n\}$ donde cada observación x_i proviene de una densidad de probabilidad $f_X(x; \theta)$.
    (a) Convencerse de que el estadístico *vectorial* $t(\underline{x}) = \underline{x} = \{x_1, \ldots, x_n\}$ es un estadístico suficiente para θ.
    (b) Si factoriza $f(\underline{x}; \theta) = G(t|\theta)H(\underline{x}|t)$, donde H no depende de θ, ¿cuánto valen en este caso G y H?
    (c) ¿Qué gran problema presenta este estadístico?

11. Considere una muestra $x_1, \ldots, x_n$ de variables aleatorias independientes tomadas de una distribución de Poisson, $P_k(\lambda)$
    (a) Considere el estadístico $t(\underline{x}) = \sum x_i$. Mostrar que la probabilidad condicional de obtener dicha muestra, dado que se conoce t: $P(x_1, \ldots, x_n|t)$, no depende de λ. Es decir, t es un estadístico suficiente de λ. ¿Es no sesgado? ¿Es consistente?
    (b) Mostrar que otro estadístico que sea función de t: $t' = g(t)$, es también un estadístico suficiente de λ.
    (c) Muestre que $P_k(\lambda)$ satisface el teorema de Darmois para λ, e identifique un estadístico suficiente. Comentario: el estadístico que se desprende de Darmois es un estadístico escalar (es decir su dimensión es siempre 1) independientemente del tamaño de la muestra.

12. Muestre que la distribución normal $N(\mu, \sigma)$, satisface la condición de Darmois para muchos parámetros

$$f(x, \underline{\theta}) = \exp\Big(\sum_{j=1}^{2} B_j(\underline{\theta}) C_j(x) + D(\underline{\theta}) + E(x)\Big)$$

para el caso de $\underline{\theta} = \{\mu, \sigma\}$.
    (a) Encuentre $B_1(\mu, \sigma)$, $B_2(\mu, \sigma)$, $C_1(x)$, $C_2(x)$, $D(\mu, \sigma)$ y $E(x)$, e identifique un par de estimadores suficientes para (μ, σ) que surgen de $C_1(x)$ y $C_2(x)$. ¿Es alguno de estos estimadores no sesgado para μ o para σ^2?
    (b) Suponga ahora que μ es conocido pero σ^2 no lo es. Redefina las funciones $B_j(\underline{\theta})$, $C_j(x)$, $D(\underline{\theta})$ y $E(x)$, y demuestre que $t(\underline{x}) = \sum -\frac{1}{2}x_i^2 + \mu x_i$, es un estadístico suficiente para σ^2, pero que es sesgado. A partir de t, encuentre una transformación $t' = g(t)$ tal que t' sea no-sesgado. Muestre que otra posible definición de $B_j(\underline{\theta})$, $C_j(x)$, $D(\underline{\theta})$ y $E(x)$ hubiera permitido encontrar directamente t'.

    Moraleja: los estimadores suficientes pueden no ser estimadores de los parámetros que queremos, pero una función de ellos sí. Además, los estadísticos suficientes no son únicos, y se pueden transformar sin perder su carácter de suficientes.

13. Considere una muestra $\{x_i\}$ extraída de $U[x; a]$, la distribución uniforme en $[a,a{+}1]$, con a real. Muestre que si bien $\bar{x}$ es un estimador consistente y no sesgado de $E(x)$, no es un estadístico suficiente. Note que en este caso no puede aplicar los teoremas de Cramer-Rao o Darmois (¿por qué?). Muestre asimismo que $\{x_{min}, x_{max}\}$ conforman un estadístico suficiente (de dimensión 2) para $E(x)$.

Máxima verosimilitud

14. (a) Obtenga el estimador de máxima verosimilitud (MV) para:
        i. $\hat{\lambda}$ en la distribución exponencial $f(x; \lambda) = \lambda e^{-\lambda x}$;
        ii. $\hat{\tau}$ en la distribución exponencial con parametrización $f(x; \tau) = e^{-x/\tau}/\tau$.
    (b) Verifique que se satisface la invarianza ante transformación de parámetros de los estimadores MV.
    (c) Muestre que $\hat{\lambda}$ es sesgado, mientras que $\hat{\tau}$ no lo es. Ayuda: Notar que la distribución exponencial es un caso particular de la Gamma y usar las propiedades de esta última.
    (d) Muestre asimismo que $\hat{\lambda}$ es asintóticamente no sesgado, como todo estimador MV.
    (e) Halle las varianzas de $\hat{\lambda}$ y $\hat{\tau}$.

15. Sea $\{x_i\}$ una muestra tomada de una cierta distribución f. Muestre que el estimador MV para $E(x)$ es el que se detalla a continuación, y muestre además que es no sesgado:
    (a) $\bar{x}$, si f es gaussiano;
    (b) $(x_{max} + x_{min})/2$ si f es uniforme en [a, b].;
    (c) la mediana si f es la doble exponencial $f(x) = (\lambda/2)\exp(-\lambda|x - \mu|)$.

16. Encuentre la ecuación que debe satisfacer el estimador MV para el centro de una Cauchy descentrada $f(x) = 1/[\pi(1+(x-\mu)^2)]$. Note que ésta no puede resolverse en una forma analítica cerrada, requiriendo una solución numérica. Muestre que este estimador satisface las condiciones para tender a distribución gaussiana para muestras grandes, y analice porque no hay contradicción con el hecho que la suma de variables aleatorias con distribución de Cauchy no tiende a una gaussiana para n grande.

17. Se realizan n mediciones $\{x_i\}$ cada una con distribución $N(\mu, \sigma_i)$ (o sea con distintos errores cada una).
    (a) Muestre que el estimador MV de μ es $\hat{\mu} = (\sum x_i/\sigma_i^2)/(\sum 1/\sigma_i^2)$, el llamado "promedio pesado" o "promedio ponderado". Interprete físicamente este resultado y obtenga su varianza. Verifique que si todos los σ_i son iguales, $\hat{\mu}$ corresponde al promedio de la muestra, como esperado.
    (b) Muestre que $\bar{x} = \sum x_i/n$ es también un estimador no sesgado de μ, pero de mayor varianza, como correponde a un estimador que no es de MV.

18. Encuentre los estimadores de máxima verosimilitud conjuntos para la esperanza y la varianza de una gaussiana y obtenga su matriz de covarianza a partir de Cramer-Rao (la matriz de información de Fisher). ¿Son sesgados estos estimadores? ¿Qué condiciones son necesarias para calcular la matriz de covarianza a partir de Cramer-Rao?

    Comentario: los estimadores MV $\hat{\mu}$ y $\hat{\sigma}^2$, son función de las estadísticas suficientes de la distribución gaussiana: $t_1 = \sum_i^n x_i$ y $t_2 = \sum_i^n x_i^2$ (mirar el ejercicio 12), como se espera para cualquier estimador MV cuando las estadísticas suficientes existen. Por lo tanto, $\hat{\mu}$ y $\hat{\sigma}^2$ son también estadísticas suficientes.

Cuadrados mínimos

19. Considere la aplicación del principio de máxima verosimilitud, al ajuste de una función $y = f(x, \vec{a})$ sobre los puntos $\{x_i, y_i\}$.
    (a) Muestre que si los y_i tienen distribución gaussiana respecto de $f(x_i, \vec{a})$ se obtiene el método de "cuadrados mínimos".
    (b) En cambio, si y_i tiene distribución doble exponencial, se obtiene el método de "módulos mínimos".
    (c) ¿Como modificaría cuadrados mínimos del punto (a) si en vez de datos (x_i, y_i) tiene que ajustar $y = f(x, \vec{a})$ a un histograma (x_i, n_i) siendo x_i el centro del bin i-*ésimo* y n_i su numero de entradas?
    (d) ¿Cual es la expresión a minimizar si se resuelve el ítem (c) por máxima verosimilitud?

    Sugerencia: por simplicidad considere que todas las mediciones en (a) tienen el mismo error σ y en (b) tienen el mismo parámetro λ

20. Muestre que al ajustar una recta $y{=}a_1{+}a_2x$ a un conjunto de datos no correlacionados $y_i \pm \sigma$, la expresión general de regresión lineal $\hat{\underline{\theta}} = \left(\mathbb{A}^T\, \mathbb{V}^{-1} \mathbb{A}\right)^{-1} \mathbb{A}^T\, \mathbb{V}^{-1}\, \underline{y}$, se reduce a la fórmula de "cuadrados mínimos", ecuación 1 del problema 10 de la guía 5.

21. (a) Haga el ajuste de una parábola, $y{=}a_1{+}a_2x{+}a_3x^2$, a los datos $\{x_i, y_i \pm \sigma_i\}$: (-0.6,5±2), (-0.2,3±1), (0.2,5±1) y (0.6,8±2).
    (ayuda: el ejercicio está resuelto en la sección 10.2.5 del Frodesen)
    (b) Repita el ejercicio suponiendo todos los errores iguales $\sigma_i = \sigma$, y estime σ de los datos. [Rta: $\hat{\sigma} = 0.67$]

22. *Ajuste de datos con errores en ambas variables, "cuadrados mínimos con errores en x e y".*
    Se realiza un conjunto de n mediciones $\{x_i, y_i\}$ con errores gaussianos independientes σ_i^x, σ_i^y, para ajustar una función $y = f(x; a_k)$ que depende de $m < n$ parámetros a_k, k=1,m.
    (a) Muestre que $\hat{a}_k$ y $\hat{x}_i^o$, los estimadores de máxima verosimilitud de a_k y $x_i^o \equiv E(x_i)$, son aquellos que minimizan la función
    $$S(a_k, x_i^o) = \sum_{i=1}^{n} \left[\left(\frac{x_i - x_i^o}{\sigma_i^x} \right)^2 + \left(\frac{y_i - f(x_i^o)}{\sigma_i^y} \right)^2 \right]$$
    Este es el denominado *criterio de Deming.* ¿Por qué consideramos estimadores $E(x_i)$ y no los de $E(y_i)$?
    (b) Compruebe que bajo la aproximación $f(x_i) = f(x_i^o) + (x_i{-}x_i^o)\, \partial f/\partial x|_i$, en un entorno alrededor de cada x_i^o, el criterio de Deming se reduce al método de varianza efectiva, en que se minimiza
    $$S = \sum_{i=1}^{n} \left(\frac{y_i - f(x_i)}{\sigma_i} \right)^2 \qquad \text{con} \quad (\sigma_i)^2 = (\sigma_i^y)^2 + \left(\sigma_i^x\, \frac{\partial f}{\partial x}\bigg|_{x_i^o} \right)^2$$
    así llamado pues es formalmente similar al caso de cuadrados mínimos ordinarios, pero remplazando σ_i^y por σ_i, un error efectivo en y más grande. Además del caso lineal $f(x) = a_1 + a_2\, x$, ¿Cuándo será válida esta aproximación? Analice porqué éste es un problema no lineal que requiere solución iterativa aún cuando la función a ajustar sea una recta.