Estadística en Física Experimental

Guía de Problemas No.7 (1$^{\text{er}}$ cuatrimestre de 2011)

Estimadores: consistencia, sesgo, eficiencia, suficiencia. Máxima verosimilitud.

1. a. Muestre que $S^2 = \sum(x_i - \mu)^2/n$ es un estimador no sesgado de la varianza cuando la esperanza μ es conocida. Encuentre el error de S^2 cuando los x_i son gausianos.
   b. Muestre que $s^2 = \sum(x_i - \bar{x})^2/n$ es un estimador sesgado de la varianza, cuyo bias vale $-\sigma^2/n$, mientras que $s^2 = \sum(x_i - \bar{x})^2/(n-1)$ es no sesgado.

2. a. Usando la desigualdad de Chebychef, muestre que $S^2 = \sum(x_i - \mu)^2/n$ es un estimador consistente de la varianza cuando la esperanza μ es conocida, para el caso que los $\{x_i\}$ tienen distribución normal.
   b. Sin hacer más cuentas que en (a), establezca una condición suficiente para que S^2 sea un estimador consistente de la varianza, cuando $\{x_i\}$ tiene distribución arbitraria.

3. 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 $f_x(t)$ simétrica alrededor de $t = 0$.
   (b) Sea k una variable aleatoria con distribución binomial $B_k(n, p)$. Muestre que k/n es un estimador no sesgado de p, mientras que $(k/n)^2$ 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.

4. 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 $V(t)$?

5. Muestre que la aplicación de la desigualdad de Cramer-Rao al el 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 sur conclusiones aplicando Cramer-Rao al parámetro σ^2.

6. 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.

7. Compruebe que la distribución de Cauchy $f(x) = 1/[\pi(1 + (x - \mu)^2)]$ 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?

8. Muestre que la distribución de Poisson $P_k(\mu)$ satisface el teorema de Darmois para μ, e identifique un estimador suficiente.

9. 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 $\underline{\theta} = \{\mu, \sigma\}$. Encuentre $B_1(\mu, \sigma)$, $B_2(\mu, \sigma)$, $C_1(x)$, $C_2(x)$, $D(\mu, \sigma)$ y $E(x)$, e identifique el par de estimadores suficientes para (μ, σ) que surgen de $C_1(x)$ y $C_2(x)$.

10. 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 estimador suficiente. Note que en este caso no puede aplicar los teoremas de Cramer-Rao o Darmois (por qué?). Muestre asimismo que $\{x_{min}, x_{max}\}$ son un par de estimadores suficientes para $E(x)$.

11. Obtenga el estimador de máxima verosimilitud (MV) para: (a) $\hat{\lambda}$ en la distribución exponencial $f(x; \lambda) = \lambda e^{-\lambda x}$; (b) $\hat{\tau}$ en la distribución exponencial con parametrización $f(x; \tau) = e^{-x/\tau}/\tau$. Verifique que se satisface la invarianza ante transformación de parámetros de los estimadores MV. Muestre que $\hat{\lambda}$ es sesgado, mientras que $\hat{\tau}$ no lo es. Muestre asimismo que $\hat{\lambda}$ es asintóticamente no sesgado, como todo estimador MV. Halle las varianzas de $\hat{\lambda}$ y $\hat{\tau}$.

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

13. Encuentre la ecuación que debe satisfacer el estimador MV para el centro de una Cauchy descentrada $f(x) = [\pi(1+(x-\mu)^2)]^{-1}$. 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.

14. 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". 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 $\bar{\mu} = \sum x_i/n$ es también un estimador no sesgado de μ, pero de mayor varianza, como correponde a un estimador que no es de MV.
(d) Si los $\{x_i\}$, en vez de normal, tuvieran distribución uniforme $f(\mu, a_i) = 1/(2a_i)$, $\mu - a_i \le x \le \mu + a_i$, muestre que el estimador MV de $\hat{\mu}$ es el x_i con menor a_i.

15. 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 la matriz de información de Fisher.

16. Muestre que (a) la información de Fisher de dos experimentos independientes que determinan el mismo parámetro es la suma de las informaciones de cada experimento por separado; (b) la información de Fisher provista por un conjunto de estadísticas suficientes es igual a la información de la muestra entera; (c) la información de una estadística no suficiente es menor que la de la muestra entera.

17. Se desea estimar, usando el principio de máxima verosimilitud, los parámetros de una función lineal $y = a_1 x + a_2$ a partir de datos $\{x_i, y_i\}$, con x_i sin error e y_i variables aleatorias independientes.
a. Muestre que si los y_i tienen distribución gaussiana se obtiene el método de cuadrados mínimos.
b. En cambio, si y_i tiene distribución doble exponencial, se obtiene el método de "módulos mínimos".

18. 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, $\hat{\boldsymbol{\theta}} = (\mathbf{A}^\mathbf{T}\,\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{A}^\mathbf{T}\,\mathbf{V}^{-1}\,\mathbf{y}$, se reduce a la fórmula de "cuadrados mínimos":
$$\begin{aligned} \hat{a}_1 &= (\textstyle\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i)/\Delta \\ \hat{a}_2 &= (N \textstyle\sum x_i y_i - \sum x_i \sum y_i)/\Delta \\ \Delta &= N \textstyle\sum x_i^2 - (\sum x_i)^2 \end{aligned} \qquad \text{con} \qquad \mathrm{Cov}(\hat{a}_1, \hat{a}_2) = \frac{\sigma^2}{\Delta}\begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & N \end{pmatrix}$$

19. 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) (Ejercicio 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.

20. *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
$$S = \sum_{i=1}^{n}\left(\frac{y_i - f(x_i)}{\sigma_i}\right)^2 \qquad \text{con} \qquad (\sigma_i)^2 = (\sigma_i^y)^2 + \left(\sigma_i^x \left.\frac{\partial f}{\partial x}\right|_{x_i^o}\right)^2$$
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. 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.