



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

TAREA 3

EJERCICIOS

1. Sea $\Theta = \{1, 2, \dots, N\}$ y una familia de distribuciones discretas sobre los números naturales cuyo representante para un $\theta \in \Theta$ se especifica así:

$$f_X(X = x; \theta) = \begin{cases} \frac{1}{\theta} & \text{si } x = 1, \dots, \theta \\ 0 & \text{en otro caso.} \end{cases} \quad (1)$$

- a) Demuestra (por ejemplo, por inducción) que si para todo θ se verifica $E_\theta[g(X)] = 0$, entonces $g(x) = 0$ (es decir, demuestra que una única observación X es un estadístico completo).
- b) Demuestra que $\hat{\theta}(X) = 2X - 1$ es un estimador insesgado de θ .
- c) Habida cuenta que $\hat{\theta}(X)$ depende sólo de X y X es completo además de (obviamente) suficiente, ¿que puedes decir de $\hat{\theta}(X)$ como estimador de θ ?
2. El modelo de regresión logística con un sólo regresor considera una variable aleatoria Y cuya probabilidad $P(Y = 1)$ verifica:

$$\log \left(\frac{P(Y = 1)}{1 - P(Y = 1)} \right) = \beta_0 + \beta_1 X. \quad (2)$$

- a) Escribe la función de probabilidad conjunta de una muestra Y_1, \dots, Y_n dados x_1, \dots, x_n .
- b) Reescribe en forma canónica la función de probabilidad anterior. Muestra que “encaja” en la forma de la familia exponencial.
- c) Por simple inspección, encuentra estadísticos conjuntamente suficientes para β_0 y β_1 .
3. Considera una tabla de contingencia de doble entrada, en cuya casilla ij se anota el total x_{ij} de veces que aparecen en conjunción los niveles i y j de dos caracteres.
- a) En el caso en que dicha tabla se suponga resultante de muestrear una población multinomial que atribuye a la casilla ij la probabilidad θ_{ij} (sin más restricción que $\sum_i \sum_j \theta_{ij} = 1$), ¿cuáles serían estadísticos suficientes para $\theta = (\theta_{11}, \dots, \theta_{IJ})$?
- b) En el caso de que dicha tabla procediera de una población en que hay independencia entre los dos caracteres considerados (es decir, $\theta_{ij} = \theta_{i+} \times \theta_{+j}$, en que θ_{i+} y θ_{+j} son las probabilidades marginales de los niveles i y j del primer y segundo carácter respectivamente), ¿cuáles serían estadísticos suficientes para $\theta = (\theta_{1+}, \dots, \theta_{I+}, \theta_{+1}, \dots, \theta_{+J})$?
- c) Obtén, en un caso como en otro, estimaciones insesgadas de mínima varianza para los parámetros correspondientes.

4. El estimador de mínima varianza insesgado basado en n observaciones para la media de una distribución exponencial,

$$f_X(x; \theta) = \theta^{-1} e^{-x/\theta} \quad (3)$$

fue ya obtenido en clase.

- a) Calcula la información de Fisher asociada a una observación.
- b) Encuentra la cota de Cramer-Rao para estimadores regulares insesgados de θ .
- c) Comprueba que el estimador obtenido en clase con ayuda del teorema de Rao-Blackwell alcanza la cota de Cramer-Rao.

Lectura recomendada. Todos los manuales que se citan a continuación son de interés: [1], [3], [5], [8], [9]. Hay problemas resueltos en todos ellos y en [7], [11] y [4].

Puede ayudarte a resolver el problema 1 el artículo [12] (las revistas de Estadística están en la hemeroteca, a la que, como estudiante de segundo ciclo, tienes acceso; los números muy atrasados han de pedirse en el mostrador).

El Ejercicio 3 hace referencia al caso más simple posible de una tabla de contingencia. Sobre análisis de datos categóricos y tablas de contingencia hay mucha bibliografía. Un pequeño manual muy legible es [6]. Mucho más avanzados son [2] o [10].

Referencias

- [1] P. J. Bickel and K. A. Doksum. *Mathematical Statistics*. Holden-Day, Inc., San Francisco, 1977.
- [2] Y.M.M. Bishop, S.E. Fienberg, and P.W. Holland. *Discrete Multivariate Analysis. Theory and Practice*. MIT Press, Cambridge, Mass., 1975.
- [3] D. R. Cox and D. V. Hinkley. *Theoretical Statistics*. Chapman and Hall, London, 1979 edition, 1974.
- [4] D. R. Cox and D. V. Hinkley. *Problems and Solutions in Theoretical Statistics*. Chapman and Hall, London, 1980 edition, 1980.
- [5] E.J. Dudewicz and S.N. Mishra. *Modern Mathematical Statistics*. Wiley, 1988.
- [6] S.E. Fienberg. *The Analysis of Cross-Classified Categorical Data*. MIT Press, Cambridge, Mass., 1980.
- [7] A. Garín and F. Tusell. *Problemas de Probabilidad e Inferencia Estadística*. Ed. Tébar-Flores, Madrid, 1991.
- [8] P.H. Garthwaite, I.T. Jolliffe, and B. Jones. *Statistical Inference*. Prentice Hall, London, 1995.
- [9] E. L. Lehmann. *Theory of Point Estimation*. Wiley, New York, 1983.
- [10] R.L. Plackett. *The Analysis of Categorical Data*. Griffin, London, 1974.
- [11] J. P. Romano and A. F. Siegel. *Counterexamples in Probability and Statistics*. Wadsworth and Brooks/Cole, Monterrey, California, 1986.
- [12] S.M. Stigler. Completeness and unbiased estimation. *American Statistician*, 26(2):28–29, 1972.