eman ta zabal zazu

Universidad del País Vasco — Euskal Herriko Unibertsitatea

TAREA 4

EJERCICIOS

1. Obtén el estimador MV para θ en el caso de una muestra $X_1, \ldots, X_n \sim U(0, \theta)$. (Nota que también a efectos de estimación MV es la distribución uniforme entre 0 y θ especial; para comenzar, no es derivable, por lo que no puede obtenerse $\hat{\theta}_{MV}$ igualando la primera de rivada de la verosimilitud a cero). ¿Es insesgado el estimador que obtienes? ¿Depende de un estadístico suficiente?

2. Un cierto tipo de planta puede presentarse en cuatro variedades, cuyas probabilidades respectivas de acuerdo a una teoría genética son: $\frac{1}{2} + \theta/4$, $(1-\theta)/4$, $(1-\theta)/4$, y $\theta/4$, en que θ es un parámetro, no especificado por la teoría, con valor entre 0 y 1. Una muestra de n plantas proporcionó $n_0, \ldots, n_3$ ejemplares respectivamente de las cuatro variedades.

   a) Obtén estadísticos suficientes para θ.

   b) Obtén el estimador MV de θ y un estimador de su varianza.

   c) En este caso es posible la obtención analítica del estimador MV (que se reduce a resolver una ecuación de sgundo grado). Emplea no obstante el método de Newton-Raphson para resolver la ecuación de verosimilitud y llegar al mismo resultado numéricamente, cuando $n_0 = 5$, $n_1 = 4$, $n_2 = 3$ y $n_3 = 1$.

   d) No hay aquí datos perdidos, y en consecuencia parecería que el uso del algoritmo EM está fuera de lugar. Observa, sin embargo, que si imaginamos la primera celda (la de probabilidad $\frac{1}{2} + \theta/4$) como compuesta por otras dos de las que hemos perdido los valores n_{01} y n_{02} y sólo conocemos $n_0 = n_{01} + n_{02}$, el problema se simplifica. Escribe la verosimilitud para comprobarlo.

3. Considera $X_1, \ldots, X_n$ iid con densidad $f_X(x; \theta) = \theta e^{-\theta x}$ para $x \geq 0$ (es decir, con distribución exponencial: nota sin embargo que a veces la misma distribución se parametriza asi: $f_X(x; \lambda) = \lambda^{-1} e^{-x/\lambda}$).

   a) Obtén el estimador máximo verosímil de θ.

   b) La distribución anterior es buen modelo matemático en multitud de situaciones relacionadas con la fiabilidad y la supervivencia. En tales contextos, frecuentemente no se pueden observar los tiempos de fallo (o supervivencia) de las n unidades ensayadas, sino que al finalizar el ensayo en el momento T quedan “vivas” o sin fallar $m \leq n$ unidades. Se habla en este caso de un ensayo *censurado por la derecha* y se dice que T es el tiempo de censura.

Obtén el estimador máximo verosímil de θ si $n - m$ variables X_i han sido observadas y las restantes m han sido censuradas (es decir, seguían vivas en el momento T de finalizar el ensayo, pero no sabemos aún por cuanto tiempo).

c) Una alternativa para la estimación de θ consistiría en considerar las unidades que aún no han fallado en T como observaciones perdidas, y omitirlas de la muestra. Explica por qué esta alternativa es claramente indeseable, y en qué dirección sesgaría la estimación de θ.

AYUDAS, SUGERENCIAS Y COMPLEMENTOS

Te serán de utilidad todos y cualquiera de los manuales que hemos venido utilizando: [4], [2], [5], [3], [1], [6], por citar sólo algunos.

Referencias

[1] P. J. Bickel and K. A. Doksum. *Mathematical Statistics*. Holden-Day, Inc., San Francisco, 1977.

[2] P.H. Garthwaite, I.T. Jolliffe, and B. Jones. *Statistical Inference*. Prentice Hall, London, 1995.

[3] J. C. Kiefer. *Introduction to Statistical Inference*. Springer-Verlag, New York, 1987 edition, 1983.

[4] E. L. Lehmann. *Theory of Point Estimation*. Wiley, New York, 1983.

[5] A. Stuart, K. Ord, and S. Arnold. *Kendall’s Advanced Theory of Statistics*, volume 2A. Arnold, 6 edition, 1999.

[6] G.A. Young and R.L. Smith. *Essentials of Statistical Inference*. Cambridge Univ. Press, 2005.