eman ta zabal zazu

Universidad del País Vasco

Euskal Herriko Unibertsitatea

TAREA 5

EJERCICIOS[1]

Puedes a tu elección trabajar en R, S-PLUS ó MATLAB.

1. En un experimento, los animales de un colectivo pueden acabar perteneciendo a cuatro categorías, que de acuerdo con el modelo genético empleado tienen probabilidades: $\frac{1}{2} + \frac{\theta}{4}, \frac{1}{4} - \frac{\theta}{4}, \frac{1}{4} - \frac{\theta}{4}$ y $\frac{\theta}{4}$ respectivamente. Se realiza el experimento y los 197 animales ensayados se desglosan en 125, 18, 30 y 34 de las categorías respectivas.

   (a) Estímese θ por máxima verosimilitud "a mano".

   (b) Estímese θ por máxima verosimilitud haciendo uso del método de Newton o "scoring".

   (c) Estímese θ por máxima verosimilitud haciendo uso del algoritmo EM.

2. Programa, a tu elección, el método de generación de variables aleatorias normales de Box-Muller o Marsaglia. Puedes hacer uso del generador de números aleatorios uniformes en $(0, 1)$ de que dispongas (`runif` en R).

3. Continúa el ejercicio anterior programando vectores aleatorios normales multivariantes con matriz de covarianzas

$$\Sigma = \begin{pmatrix} 0.3 & 0.1 \\ 0.1 & 0.3 \end{pmatrix}$$

   y vector de medias $\boldsymbol{\mu} = (0.50.7)$.

4. Emplea el método de Monte Carlo para estimar; i) el valor de π. ii) El valor de $\int_0^2 x^2 dx$.

5. *(optativo: trivial conceptualmente, pero te exigirá pensar un poco en cómo programarlo).* Las llegadas por minuto al peaje de una autopista se distribuyen como Poisson con parámetro $\lambda = 23$. Un empleado atiende a un número aleatorio de clientes que se distribuye como Poisson con parámetro $\delta = 10$. Estima por Monte Carlo la espera media y la cola media en número de vehículos si empleas a tres, cuatro y cinco empleados. Simplifica como veas necesario (en particular, has de suponer que los vehículos son atendidos de inmediato por cualquier empleado que quede libre, aunque estén en la cola de otro).

[1]Copia de estos enunciados y, en general, de todo el material escrito repartido en clase, está disponible en `http://etdx01.bs.ehu.es`.

AYUDAS, SUGERENCIAS Y COMPLEMENTOS

1. Como se indicaba en la tarea anterior, el algoritmo de Newton (y también por consiguiente "scoring") requieren hacer uso de derivadas. Podemos calcularlas "a mano" o hacer uso de las capacidades de derivación simbólicas de R (pobres: véase las funciones `deriv` y `deriv3`), algo más extensas de MATLAB o de un programa especializado como MATHEMATICA. Si seguimos esta última vía, podemos hacer todo el trabajo con él o habremos de "empalmar" el resultado en nuestra función en R o MATLAB.

   Esta última operación resulta facilitada porque podemos exportar fórmulas generadas por MATHEMATICA a notación C o FORTRAN (también a TEX, si deseáramos incluirlas en un documento): Véase [4], pág. 182 en relación con este último punto.

2. En el primer ejercicio, el uso del método EM puede parecer injustificado. No hay datos faltantes. Hay un truco muy utilizado, que consiste en convertir un problema con datos completos en otro con datos faltantes cuya verosimilitud es fácil o trivial de calcular. En el caso que te dan, si escindes la primera categoría en dos con probabilidades respectivas $\frac{1}{2}$ y $\frac{\theta}{4}$ tienes un nuevo problema en que solo las (nuevas) categorías 2 a 5 contienen información acerca de θ.

   Si supieras cuantos animales van a la categoría artificial 2 la verosimilitud sería trivial de maximizar (compruébalo, compara con la de una distribución binomial). No sabes cuantos animales asignar a dicha categoría 2, pero aquí el algoritmo EM viene en tu ayuda.

3. Para evaluar π por el método de Monte Carlo, puedes recurrir a generar números aleatorios en el cuadrado $[0,1] \times [0,1]$ y ver que proporción de ellos "caen" en un cuarto de círculo inscrito en dicho cuadrado (de área $\pi/4$, por tanto). (En textos de Estadística o Análisis Numérico encontrarás estrategias diferentes, como la de Buffon.)

   Análogamente, para evaluar aproximadamente la integral que te piden deberías generar números en un recinto adecuado y ver la proporción de ellos que quedan "por debajo" de x^2.

4. El problema del peaje es un problema (muy sencillo) de teoría de colas, que admite solución analítica muy simple. Para simularlo, te convendrá trabajar con la distribución de los tiempos de llegada y de atención al cliente (exponenciales).

References

[1] G. Dahlquist and Åke Björck. *Numerical Methods*. Prentice Hall, Englewood Cliffs, N.J., 1974.

[2] K. Lange. *Numerical Analysis for Statisticians*. Springer, 1998. Signatura: 519.6 LAN.

[3] R.A. Thisted. *Elements of Statistical Computing*. Chapman & Hall, New York, 1988.

[4] S. Wolfram. *Mathematica: A System for doing Mathematics by Computer*. Addison-Wesley, 1991.