Laboratorio de Métodos Numéricos - Primer Cuatrimestre 2013
Trabajo Práctico Número 1

Introducción

Numerosas aplicaciones utilizan la distribución Gamma Generalizada (DΓG) para modelar datos, cuya función de densidad se expresa como

$$f_\Theta(x) = \frac{\beta\, x^{\beta\lambda-1}}{\sigma^{\beta\lambda}\,\Gamma(\lambda)} \exp\left\{-\left(\frac{x}{\sigma}\right)^{\beta}\right\} \qquad \text{con } x \in \mathbb{R}_{>0}$$

donde $\Gamma(\cdot)$ es la función Gamma[1] definida como $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}\,dt$, y Θ representa a la tupla de parámetros $\Theta = (\sigma, \beta, \lambda)$. El primero se encuentra relacionado con la escala de la función f y los otros dos con la forma; todos son positivos. La DΓG engloba un conjunto amplio de distribuciones paramétricas, donde la distribución exponencial, Weibull o Gamma son casos especiales de ésta. En vez de utilizar (y almacenar) el histograma empírico de los datos, representarlos solamente con los parámetros de esta distribución resulta ser, en muchos casos, una opción más que conveniente. El problema que surge consiste en estimar de forma certera y eficiente los parámetros de la DΓG que mejor ajusta a los datos.

Estimación de parámetros

Sean n datos reales positivos $x_1, \ldots, x_n$. Estos datos pueden provenir de mediciones de un satélite, píxeles de imágenes, etc., y de los cuales asumimos que siguen una distribución DΓG cuyos parámetros queremos estimar. Como se vió en la materia *Proba*, existen métodos como '*estimadores de máxima verosimilitud*', donde a partir de las muestras se pueden estimar parámetros de distribuciones. Utilizando este método, llegamos a que se deben cumplir las siguientes condiciones:

$$\tilde{\sigma} = \left(\frac{\sum_{i=1}^n x_i^{\tilde{\beta}}}{n\tilde{\lambda}}\right)^{1/\tilde{\beta}} \tag{1}$$

$$\tilde{\lambda} = \left[\tilde{\beta}\left(\frac{\sum_{i=1}^n x_i^{\tilde{\beta}} \log x_i}{\sum_{i=1}^n x_i^{\tilde{\beta}}} - \frac{\sum_{i=1}^n \log x_i}{n}\right)\right]^{-1} \tag{2}$$

$$0 = \frac{\tilde{\beta}}{n}\sum_{i=1}^n \log x_i - \log \sum_{i=1}^n x_i^{\tilde{\beta}} + \log(n\tilde{\lambda}) - \psi(\tilde{\lambda}) \tag{3}$$

donde $\tilde{\Theta} = (\tilde{\sigma}, \tilde{\beta}, \tilde{\lambda})$ son las estimaciones a partir de las muestras de datos y $\psi(\cdot)$ es la 'conocida' función digamma[2] que se define como la derivada logarítmica de la función Gamma. Notar que una una vez estimado el parámetro β, los otros dos parámetros pueden obtenerse de las primeras dos ecuaciones. A partir del cálculo de los *momentos* de una variable aleatoria con función de densidad $f_\Theta(x)$ se pueden obtener las siguientes dos ecuaciones:

$$\log(\mathcal{M}(2\beta)) - 2\log(\mathcal{M}(\beta)) = \log(1 + \beta(\mathcal{R}(\beta) - \mathcal{R}(0))) \tag{4}$$

$$\frac{\mathcal{M}(2\beta)}{\mathcal{M}^2(\beta)} = 1 + \beta(\mathcal{R}(\beta) - \mathcal{R}(0))) \tag{5}$$

[1] La función Gamma, en el caso entero, es equivalente a la función factorial: $\Gamma(n) = n!$, $\forall n \in \mathbb{N}$

[2] $\psi(z) = \frac{\partial}{\partial z} \ln \Gamma(z) = \frac{\Gamma'(z)}{\Gamma(z)}$.

donde

$$\mathcal{M}(s) = \frac{1}{n}\sum_{i=1}^{n} x_i^s\,, \qquad \widehat{\mathcal{M}}(s) = \frac{1}{n}\sum_{i=1}^{n} x_i^s \log(x_i)\,, \qquad \mathcal{R}(s) = \frac{\widehat{\mathcal{M}}(s)}{\mathcal{M}(s)}$$

Estas ecuaciones tienen como ventaja que no dependen de los otros parámetros, utilizándose funciones más sencillas que sólo dependen de los datos. Hallando la solución de cualquiera de las ecuaciones (4) o (5), es posible estimar el parámetro β, y despejar el resto de los parámetros utilizando las ecuaciones (1) y (2).

Enunciado

El objetivo del trabajo práctico es implementar un programa que permita estimar los parámetros $\Theta = (\sigma, \beta, \lambda)$ a partir de un conjunto de n datos. Para ello, se deberá resolver la ecuaciones (4) o (5). Evaluando los distintos métodos vistos en clase que permitan resolver este problema, se deberá realizar una implementación cumpliendo lo siguiente:

1. Implementar al menos dos métodos (de los cuales uno de ellos debe ser el método de Newton) con aritmética binaria de punto flotante con t dígitos de precisión en la mantisa. El valor t debe ser un parámetro de la implementación, con $t < 52$.
2. Realizar experimentos numéricos con cada método implementado en el ítem anterior elegiendo varias instancias de prueba y en función de las cantidad de dígitos t de precisión en la mantisa (experimentar con al menos 3 valores distintos de t).
3. Para cada método implementado se deberán mostrar resultados obtenidos en cuanto a cantidad necesaria de iteraciones, tiempo de ejecución, precisión en el resultado, y cualquier otro parámetro que considere de interés evaluar.
4. Realizar el gráfico del histograma de los datos y el ajuste obtenido. Extraer conclusiones sobre la efectividad de cada método observando los resultados anteriores.

Formato de archivos de entrada

El programa debe tomar los datos desde un archivo de texto con el siguiente formato:

input_file.txt
n
x_1 x_2 ... x_n

El archivo contiene en la primer línea la cantidad de datos, y en la línea siguiente se encuentran los n datos (reales positivos) separados por espacio. En la web de la materia se publicarán varios archivos de prueba para realizar los primeros experimentos.

Fecha de entrega:

- *Formato electrónico:* domingo 14 de abril de 2013, hasta las 23:59 hs., enviando el trabajo (informe+código) a `metnum.lab@gmail.com`. El subject del email debe comenzar con el texto `[TP1]` seguido de la lista de apellidos de los integrantes del grupo.
- *Formato físico:* lunes 15 de abril de 2013, de 18 a 20hs (en la clase práctica).