Simulación de variables aleatorias y estadística

Estadística.

1. Hallar el estimador de máxima verosimilitud de μ y de σ en forma simultánea en el caso normal.

2. Hallar los estimadores de máxima verosimilitud de a y b a partir de una m.a.s. de variables uniformes en $[a, b]$

3. Hallar el estimador de máxima verosimilitud de α a partir de una m.a.s. de variables exponenciales de parámetro α.

4. Calcular el estimador de máxima verosimilitud de λ a partir de una m.a.s. de variables con distribución de Poisson de parámetro λ.

5. *Distribución empírica de una muestra.* Dada una muestra aleatoria simple $X_1, \dots, X_n$ de v.a.i.i.d. con distribución común F, construimos su función empírica mediante la fórmula

$$F_n(x) = \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{\{X_k \leq x\}}.$$

(a) Verificar que la distribución F_n corresponde a una variable aleatoria discreta que toma los valores $X_1, \dots, X_n$ (que se suponen fijos en el razonamiento) con equiprobabilidad.

(b) Demostrar, que para x fijo, $F_n(x)$ converge en probabilidad a $F(x)$. Para esto se puede utilizar el Teorema de Bernoulli, identificando un esquema de Bernoulli subyacente.

6. *Test de Kolmogorof.* Se define el estadístico de Kolmogorov para la distribución empírica de una m.a.s, en el caso en que F sea continua, mediante la fórmula

$$D_n = \sup_{x \in \mathbf{R}} |F_n(x) - F(x)|$$

Este estadístico se utiliza para testear si una determinada muestra proviene de una distribución.

(a) Demostrar que el supremo es en realidad un máximo, que se toma en uno de $2n$ valores:

$$D_n = \max_{k=1,\dots,n} \max\{|F_n(X_k-) - F(X_k-)|, |F_n(X_k) - F(X_k)|\},$$

donde $F(x-) = \lim_{y\uparrow x} F(y)$ (límite por la izquierda).

(b) Se puede demostrar que la distribución de la variable aleatoria D_n no depende de la distribución F (esto es un resultado notable).

(c) Como no conocemos la distribución de D_n, tomamos $n = 100$ y, tomando 1000 muestras de 100 variables uniformes en $[0, 1]$, hacemos un histograma de la variable D_n. De allí calculamos el valor de d_0 tal que

$$\mathbf{P}(D_n \geq d_0) \sim 0,05.$$

Simulación.

7. (a) Hacer un algoritmo para simular una variable aleatoria uniforme en el intervalo $[0, 1]$.

(b) Comparar en un mismo gráfico la distribución teórica de una variable uniforme $[0, 1]$ con la distibución empírica de la simulación de (a) para una muestra de 100 datos. Calcular el estadístico de Kolmogorov D_n y hacer un test de hipótesis con la hipótesis nula de que la distribución de la muestra en uniforme.

8. *Histrograma de una muestra aleatoria simple* Un histograma de una muestra aleatoria simple de una variable continua, es una función que parte el eje real en intervalos, y en cada intervalo toma un valor constante, cuya area es la proporción de datos que caen en el intervalo. De esa forma el area total de un histograma es uno, y las alturas indican las zonas de mayor probabilidad. Un histograma es un estimador (muy elemental) de la densidad de una variable aleatoria continua. El ejercicio es explorar en R los comandos que construyen histogramas, y utilizarlo en tres situaciones diferentes. Se pueden también explorar otras formas de estimar densidades, tomando intervalos equiespaciados, u otras formas mas elaboradas, como `plot(density())` o el paquete `kdensity`.

9. En este ejercicio vamos a simular variables aleatorias con distintas distribuciones. En cada caso vamos a hacer un histograma y a superponerlo con la densidad teórica de la variable aleatoria. El único comando de R que vamos a usar es `runif(n)`. En cada caso tomamos una muestra de tamaño $n = 10000$.

(a) Simular variables aleatorias uniformes en un intervalo $[-1, 1]$.

(b) Simular variables aleatorias exponenciales de parámetro $\alpha = 1/2$ y $\beta = 2$.

(c) Simular variables aleatorias con distibución de Cauchy.

(d) Simular variables aleatorias con distribución de Pareto en $[1, \infty)$ de parámetro $a = b = 1$ (ver ejercicio 9 del práctico 4)

(e) Simular variables aleatorias con distribución triangular.

(b) Simular variables aleatorias con distribución normal estándar.