Práctica 2

Métodos universales para la simulación de variables continuas

2.1. Método de inversión

1. La distribución doble exponencial (o distribución de Laplace) de parámetro λ tiene función de densidad:

$$f(x) = \frac{\lambda}{2} e^{-\lambda|x|},\ x \in \mathbb{R}$$

y función de distribución:

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \begin{cases} \frac{1}{2}e^{\lambda x} & \text{si } x < 0 \\ 1 - \frac{1}{2}e^{-\lambda x} & \text{si } x \geq 0 \end{cases}$$

   a) Escribir una función que permita generar, por el método de inversión, una muestra de n observaciones de esta distribución (NOTA: esta distribución puede generarse fácilmente simulando una distribución exponencial y otorgarle un signo positivo o negativo con equiprobabilidad; ver método de composición).

   b) Generar 10^4 valores de la distribución doble exponencial de parámetro $\lambda = 2$ y obtener el tiempo de CPU que tarda en generar la secuencia.

   c) Representar el histograma y compararlo con la densidad teórica.

2.2. Método de aceptación-rechazo

2. Desarrollar el código necesario para generar, por el método de aceptación-rechazo, una muestra de n observaciones de una distribución normal estándar:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},\ x \in \mathbb{R},$$

empleando como distribución auxiliar una doble exponencial. Puede verse que la elección de la densidad auxiliar óptima se corresponde con $\lambda = 1$ y que la cota optima es:

$$c_{\text{opt}} = \sqrt{\frac{2e}{\pi}} \simeq 1.3155.$$

Para esteblecer la condición de aceptación o rechazo se puede tener en cuenta que:

$$c \cdot U \cdot \frac{g(T)}{f(T)} = \sqrt{\frac{2e}{\pi}} U \sqrt{\frac{\pi}{2}} \exp\left(\frac{T^2}{2} - |T|\right) = U \cdot \exp\left(\frac{T^2}{2} - |T| + \frac{1}{2}\right).$$

a) Generar una muestra de 10^4 observaciones empleando este algoritmo. Obtener el tiempo de CPU y calcular el número medio de generaciones de la distribución auxiliar.

b) Representar el histograma y compararlo con la densidad teórica.

c) Aproximar la cota óptima numéricamente.

d) Aproximar el parámetro óptimo de la densidad auxiliar numéricamente (normalmente comenzaríamos por este paso).

3. Para la estimación Bayes de la media de una normal se suele utilizar como distribución a priori una Cauchy.

   a) Generar una muestra i.i.d. $X_i \sim N(\theta_0, 1)$ de tamaño $n = 10$ con $\theta_0 = 1$. Utilizar una $Cauchy(0, 1)$ (`rcauchy`) como distribución a priori y como densidad auxiliar para simular por aceptación-rechazo una muestra de la densidad a posteriori (emplear `dnorm` para construir la verosimilitud). Obtener el intervalo de probabilidad al 95 %.

   b) Repetir el apartado anterior con $n = 100$.

2.3. Ejercicios de fin de práctica

4. Escribir el código necesario para generar, por el método de inversión, una muestra de n observaciones de una distribución de Cauchy.

   a) Generar una muestra de 10^4 observaciones y obtener el tiempo de CPU.

   b) Representar el histograma (limitar el rango, e.g. `xlim = c(-10, 10)`) y compararlo con la densidad teórica (`dcauchy`).

   c) Obtener conclusiones sobre la existencia de una media teórica a partir de la media muestral aproximada por simulación (estudiar la convergencia de la media muestral). Suponiendo que el vector `x` contiene las simulaciones, estudiar la convergencia de la media muestral mediante el gráfico:

   ```
   plot(1:nsim, cumsum(x)/(1:nsim), type="l", ylab="Media muestral", xlab="Nº de simulaciones")
   ```

5. El tiempo de respuesta (en centésimas de segundo) de un servidor de bases de datos es una variable con función de densidad:

   $$f(x) = xe^{-x} \text{ si } x \geq 0.$$

   Escribir el código necesario para generar, por el método de aceptación-rechazo, una muestra de n observaciones de esta distribución empleando como densidad auxiliar una exponencial:

   $$g(x) = \lambda e^{-\lambda x} \text{ si } x \geq 0.$$

   a) Aproximar numéricamente el parámetro óptimo ($\lambda_{opt} < 1$) y la cota óptima (c_{opt}) de la densidad auxiliar y compararlos con los valores teóricos: $\lambda_{opt} = 1/2$ y $c_{opt} = 4/e$.

   b) Generar una muestra de 1000 observaciones de la distribución de interés tomando como semilla inicial los cuatro primeros dígitos del DNI. Obtener el tiempo de CPU que tarda en generar la secuencia y calcular el número medio de generaciones de la distribución auxiliar.

   c) Representar el histograma y compararlo con la densidad teórica.