

Simulación Estocástica

Tarea 3

Fecha de entrega: 14 de septiembre

1. Considera una variable aleatoria X con distribución gamma de parámetros $(\alpha, 1)$ suponiendo $\alpha \geq 1$ y realiza lo siguiente:
 - (a) Muestra que $\frac{X}{\beta}$ se distribuye gamma de parámetros (α, β) . ¿Qué significa esto para efectos de la simulación de variables gamma?
 - (b) Para el caso $\alpha < 1$ muestra que si X es gamma $(1 + \alpha, 1)$ independiente de $U \sim U(0, 1)$ entonces $XU^{\frac{1}{\alpha}}$ se distribuye gamma $(\alpha, 1)$.
 - (c) Implementa un algoritmo de aceptación y rechazo para simular de una distribución gamma de parámetros $(\alpha, 1)$ (con $\alpha \geq 1$) utilizando como distribución instrumental una gamma de parámetros $(\lfloor \alpha \rfloor, \frac{\lfloor \alpha \rfloor}{\alpha})$.
 - (d) Muestra que si X se distribuye gamma $(\alpha, 1)$ y Y se distribuye gamma $(\beta, 1)$, entonces $\frac{X}{X+Y}$ se distribuye beta de parámetros (α, β) .
 - (e) Utiliza los resultados anteriores y tu código para simular una muestra de tamaño $n = 100, 1000, 10000$ de una beta de parámetros $(2.2, 4.7)$. Para cada tamaño de muestra analiza la eficacia del algoritmo en términos del número total de muestras propuestas y el número total de muestras rechazadas. ¿Qué tan eficiente es el algoritmo?
2. En este ejercicio se estudiará una variante para simular de una distribución normal estándar usando el método polar de Marsaglia, cuyo algoritmo es:
 1. Simular $U, V \sim U(-1, 1)$.
 2. Si $0 < S = U^2 + V^2 < 1$ hacer

$$\begin{aligned} X &= U \sqrt{-\frac{2 \log(S)}{S}} \\ Y &= V \sqrt{-\frac{2 \log(S)}{S}}. \end{aligned}$$

3. Repetir en caso contrario.

De esta forma al aceptar se tiene que X y Y son variables aleatorias independientes con distribución normal estándar.

- (a) Implementa el algoritmo de Marsaglia y genera una muestra de tamaño $n = 10, 100, 1000, 10000$ de una distribución normal estándar. Muestra que la probabilidad teórica de aceptación del algoritmo es $\pi/4$ y compara este valor con la proporción de propuestas aceptadas obtenida en tus simulaciones.
 - (b) Compara tu algoritmo contra el método de Box-Muller discutido en clase y el algoritmo por default de R para los mismos tamaños de muestra $n = 10, 100, 1000, 10000$. ¿Cuál es más eficiente para cada caso?
3. (1 punto) Considera una urna con bolas de K diferentes colores, donde al tiempo 0, hay α_i bolas del i -ésimo color. Al tiempo n , se saca una bola y se regresa a la urna junto con una nueva bola de ese color y se repite el proceso N veces. Si $N \rightarrow \infty$, se puede demostrar que el vector de proporciones de bolas en la urna converge casi seguramente a un vector aleatorio con distribución Dirichlet.
- (a) Considerando $K = 3$, $\boldsymbol{\alpha} = (2, 5, 1)$ y N “suficientemente grande”, implementa este algoritmo y obtén una muestra de tamaño $n = 10, 50, 100, 1000$. Para diferentes valores de N (por ejemplo: 100, 1000, 10000, 100000) compara la media, el segundo momento y la covarianza muestral contra los momentos teóricos de la distribución.
 - (b) Realiza el mismo proceso pero ahora a partir de la generación de variables Dirichlet utilizando la construcción a partir de las variables gamma vista en clase.
 - (c) Compara los resultados obtenidos mediante ambos algoritmos. ¿Cómo afecta N a la aproximación obtenida mediante la urna de Pólya? ¿Qué diferencias observas entre ambos métodos en términos de precisión y costo computacional?