

Laboratorio 1: Monte Carlo y eficiencia de simulación

Indicaciones: Para este laboratorio puede usar todas las funciones de Python que estime necesarias, restringiéndose solamente al uso de los generadores básicos de números aleatorios uniformes en $[0,1]$ en lo que respecta a las simulaciones. En cada uno de los siguientes problemas, cuando se pidan comparaciones para distintos parámetros, se deberá entregar en las respuestas las tablas correspondientes y los códigos utilizados. Debe ser ordenado al programar, comentando cada método implementado para facilitar la corrección.

Problema 1

En *python* existen distintas librerías que permiten simular variables aleatorias. Entre ellas se destacan las siguientes:

- *Numpy*
- *Scipy*
- *Random*

1. Programe la función `uniforme`, que reciba un valor entero n y un método (*Numpy*, *Scipy* o *Random*), y retorne n simulaciones de una variable aleatoria uniforme en $[0, 1]$.
2. Genere 10^6 uniformes para cada librería y grafique en un histograma cada muestra generada, utilizando la librería *seaborn*.
3. Genere 100 muestras de 1000 uniformes en $[0,1]$ y utilizando la librería *time* calcule los tiempos de ejecución que toman generar cada muestra para cada librería. Grafique los tiempos encontrados para cada método y calcule la media y varianza del tiempo de ejecución de cada muestra generada para cada método.
4. En base a los resultados anteriores. ¿Cual es el mejor método a utilizar? Argumente.
5. Para cada librería genere funciones que permitan obtener una muestra de las siguientes variables:
 - $\text{Bernoulli}(p)$.
 - $\text{Binomial}(p, N)$.
 - $\text{Geométrica}(p)$.

Compare los tiempos de ejecución con el de los métodos de cada librería y comente cual es mejor.

Las librerías anteriores generan números **pseudoaleatorios** que se asemeja bastante a lo que se necesita.

6. Averigüe y explique en qué consisten los métodos para generar números pseudoaleatorios uniformes en $[0,1]$ disponibles en la versión que se usará de Python. Especifique: número de bits, período (de congruencias lineal utilizada o medida equivalente para el método que corresponda), posibilidad y manera de cambiar semilla.

Problema 2

Tomando en cuenta que

$$I = \frac{\pi}{4} = \int_0^1 \sqrt{1-x^2} dx = \int_0^1 \int_0^1 \mathbf{1}_{\{x^2+y^2 \leq 1\}} dx dy,$$

se considerarán dos métodos de Monte Carlo para calcular numéricamente I :

- Utilizando la variable aleatoria $X = \sqrt{1-U^2}$, con U v. a. uniforme en $[0, 1]$.
 - Utilizando la variable aleatoria $Z = \mathbf{1}_{\{U_1^2+U_2^2 \leq 1\}}$, con U_i v. a. uniforme en $[0, 1]$ e independientes.
1. Calcule las varianzas $\text{Var}(X)$ y $\text{Var}(Z)$ de forma teórica y de forma simulada con diferentes cantidades de réplicas n . Grafique. Estime una cantidad de réplicas necesarias para X y Z con tal de obtener una aproximación de la varianza con un error del orden del 1 %.
 2. Calcule la cantidad de réplicas necesarias para X y Z con tal de aproximar I con un error máximo de $\text{Err}_1 = 0,1$ y probabilidad $\text{Pr}_1 = 90\%$. Haga el mismo ejercicio con $\text{Err}_2 = 0,01$ y $\text{Pr}_2 = 95\%$, $\text{Err}_3 = 0,001$ y $\text{Pr}_3 = 99\%$.
 3. Aproxime las esperanzas $\mathbb{E}(X)$ y $\mathbb{E}(Z)$ de forma simulada con diferentes cantidades de réplicas n hasta llegar al n^* tal que se cumple Err_3 y Pr_3 . Grafique las aproximaciones en función de la cantidad de réplicas. Grafique el tiempo utilizado en aproximar las esperanzas en función de la cantidad de réplicas. Estime los costos de simular una réplica de X y una réplica de Z .
 4. Considerando Err_3 y Pr_3 calcule un intervalo de confianza para I utilizando X y Z . Mida el tiempo total utilizado por cada método para obtener dicha precisión y compare los errores de estimación. Compare los costos totales para cada método ¿Cuál método es más eficiente? ¿Cuál es el ratio de eficiencia entre ambos métodos (entendido como el cuociente entre los costos totales de ambos métodos, para obtener una precisión con un nivel de confianza dados)?
 5. Considerando Err_3 y Pr_3 calcule el costo teórico de estimar I utilizando X y Z , tomando como costo la cantidad de variables aleatorias uniformes necesarias ¿Cuál método es mas eficiente bajo este criterio? ¿Qué diferencia se observa entre comparar las eficiencias usando este criterio (número de uniformes) y el criterio anterior (costo total)? ¿Qué indica esa diferencia? ¿Cuál criterio debería preferirse en general?

Problema 3

1. Programe el método $\text{DiscreteQuantile}(f,u)$, que recibe como parámetros una función de masa discreta f y un vector $u \in [0, 1]^r$, y retorne el menor vector $n \in \mathbb{N}^r$ (coordenada a coordenada) tal que $\sum_{j=0}^{n_i} f(j) \geq u_i$.
2. Programe el método $\text{DiscreteQuantileF}(F,u)$, que recibe como parámetros una función de distribución F y un vector $u \in [0, 1]^r$, y retorne el menor vector $n \in \mathbb{N}^r$ tal que $F(n_i) \geq u_i$.
3. Programe el método $\text{ContinuousQuantile}(F,f,u)$, que recibe como parámetros una función de distribución F , su función de densidad f y un vector $u \in [0, 1]^r$, y aplique el método de Newton para calcular el vector $x \in \mathbb{R}^r$ tal que $|F(x_i) - u_i| \leq \text{error}$, donde error es un parámetro de la clase inicializado con $\text{error}=10^{-4}$.

Considere X una variable aleatoria discreta con

$$\mathbb{P}(X = j) = \left(\frac{1}{2}\right)^j, \quad j \geq 1.$$

4. Para $k = 1, \dots, 5$, simule $n = 10^k$ réplicas de X para cada uno de los 3 métodos implementados, usando las mismas n^k variables uniformes para cada uno de ellos (en el caso del método continuo utilice el comando *ceil* para redondear el resultado) ¿En que medida coinciden los resultados de los tres métodos y por qué? Grafique el tiempo de ejecución en función de la cantidad de réplicas, estime el costo por réplica de cada uno de los métodos y ordene los métodos según su eficiencia.
5. Usando el método más eficiente, para $k = 1, \dots, 5$ simule $n = 10^k$ réplicas de X , calcule las medias y varianzas muestrales, y luego compare los resultados con los valores teóricos.

Problema 4

Considere $Y_{\lambda,s}$ variable aleatoria discreta con

$$\mathbb{P}(Y_{\lambda,s} = k) = \frac{e^{-\lambda} \lambda^k / k!}{\sum_{j=0}^s e^{-\lambda} \lambda^j / j!}, \quad \text{para } k = 0, \dots, s.$$

Para las evaluaciones considere $\lambda = 1$ y $s = 8$.

1. Escriba dos métodos para simular n réplicas de $Y_{\lambda,s}$, el primero utilizando v.a. uniformes discretas de parámetro s y el segundo utilizando v.a. de Poisson de parámetro λ . Evalúe y analice la eficiencia teórica y numérica de ambos métodos. Para el segundo método la v.a. Poisson debe simularse utilizando variables uniformes.
2. Realice un análisis completo para estimar $\mathbb{E}(Y_{\lambda,s})$ con una precisión de 0,2 y nivel de confianza 95 %. El análisis debe considerar los dos métodos implementados, cantidad de réplicas, número de v.a. uniformes, tiempo de ejecución, varianza, media y gráficos.