Silvia Calatrava
Santiago Dobón

PRÁCTICA 3: SIMULACIÓN, PRECISIÓN Y REDUCCIÓN DE LA VARIANZA

En esta práctica realizaremos dos ejercicios para comprender las técnicas de reducción de la varianza y la precisión y su relación con los intervalos de confianza.

Generación de muestras de tamaño 1000 obtenida por el método de Box-Muller usando tanto la fórmula del seno como la del coseno, y sus respectivas antitéticas.

Con el objetivo de aumentar la precisión en simulación, podemos reducir la varianza de las variables generadas. Para reducir la varianza de un estimador existen diversas técnicas: desde la más simple consistente en tomar más muestras hasta otras relacionadas con el muestreo o el uso de variables antitéticas.

En esta práctica vamos a centrarnos en el uso de variables antitéticas Esta técnica se basa en inducir una correlación negativa entre las distintas muestras obtenidas en sucesivas simulaciones, de tal manera que la varianza de la variable obtenida se reduzca al ser su covarianza negativa:

$$V(\overline{X}) = \tfrac{1}{4}(V(X_1) + V(X_2) + 2COV(X_1, X_2))$$

Veamos un ejemplo simulado de esta reducción:

Utilizamos el generador de Box-Muller para obtener valores correspondientes a dos distribuciones normales con media 0 y varianza 1. El generador de Box-Muller utilizado usa valores aleatorios producidos mediante un generador congruencial multiplicativo. Estos valores pertenecen a una distribución uniforme [0, 1].

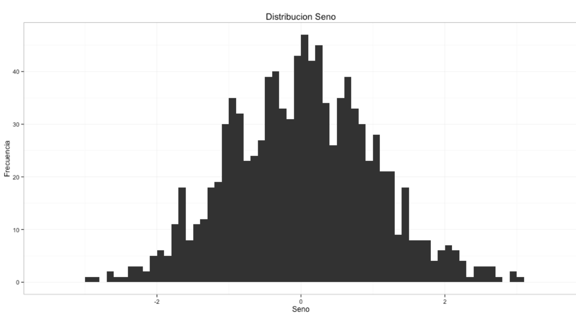
Realizamos también una modificación del generador de Box-Muller para generar los valores complementarios de la distribución normal tanto en la fórmula del seno como en la del coseno, Para ello, restamos 1 a los valores producidos por el generador congruencial en cada iteración, generando las variables antitéticas de las anteriores.

Calculamos los estadísticos de las distribuciones obtenidas:

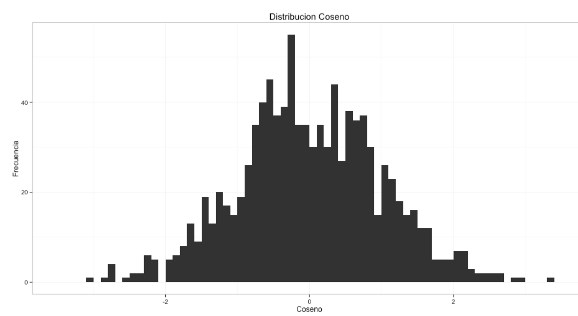
		Media Muestral	**Cuasivarianza**	**Varianza media muestral**	**Intervalo de confianza 95%**
Normales	**Seno**	0.0215871	1.005168	0.001005168	[-0.04055349, 0.08372769]
	Coseno	0.005818751	0.9983748	0.0009983748	[-0.05611150, 0.06774901]
Antitéticas	**Seno**	-0.04075334	0.9424225	0.0009424225	[-0.10092318, 0.01941651]
	Coseno	-0.01259807	1.017283	0.001017283	[-0.07511201, 0.04991588]

Mostramos a continuación los histogramas de las distribuciones obtenidas:

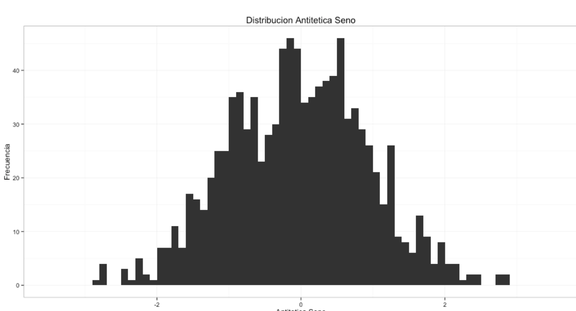
Normal Box-Muller seno



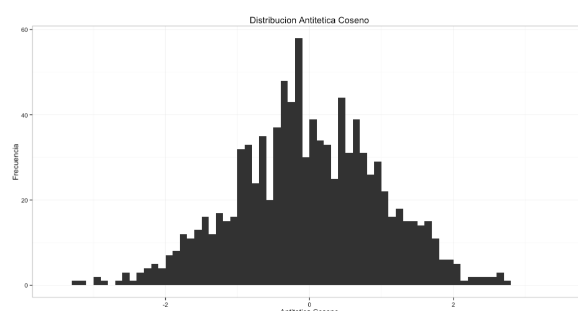
Normal Box-Muller coseno



Normal Box-Muller antitetica seno



Normal Box-Muller antitetica coseno



Y finalmente calculamos la correlación entre las distribuciones obtenidas. Usamos el coeficiente de correlación:

Distribución A	**Distribución B**	**Coeficiente de Correlación**
Normal Seno	Normal Coseno	-0.0009144869
Antitñetica Seno	Antitética Coseno	0.04509202
Normal Seno	Antitética Seno	**-0.5981391**
Normal Seno	Antitética Coseno	-0.01701175
Normal Coseno	Antitética Seno	0.01825988
Normal Coseno	Antitética Coseno	**0.5817548**

Observamos que es posible reducir la varianza si simulamos 2 veces, una utilizando la distribución normal obtenida con la fórmula del seno y otra utilizando función Box-Muller en su versión antitética. Ocurre lo contrario si utilizamos la fórmula del coseno.

Generación de 100 grupos de muestras de tamaño 100 obtenida por el método de Box-Muller usando tanto la fórmula del seno y análisis de su precisión

Utilizamos el generador de Box-Muller para obtener valores correspondientes a 100 distribuciones normales con media 0 y varianza 1 usando la fórmula del seno de Box-Muller.

Para cada distribución obtenida calculamos sus intervalos de confianza para la media muestral con distintos grados de confianza. Contabilizamos luego en cuántos de estos intervalos está incluido el valor 0, es decir el valor de la media que queríamos obtener en nuestra distribución simulada.

Grado de confianza	**Intervalos que no contienen el 0**
90	10
95	7
99	2

Los resultados obtenidos coinciden con los esperados, ya que un intervalo de confianza de cierto grado de confianza α indica precisamente que si tomamos un número de intervalos, en $100(1 - \alpha)\%$ de ellos, sí encontraremos incluido nuestro estimador. No es posible por tanto asegurar la inclusión del parámetro para un único caso.