

Capítulo 6

Método Monte Carlo y reducción de varianza.

En este capítulo introduciremos un método importante en la simulación estocástica que permite dar solución a problemas matemáticos que, como ya se mencionó en el primer capítulo, resultan costosos o imposibles resolver analíticamente. Además se presenta la importancia de la reducción de varianza en la implementación de algoritmos de simulación y algunas técnicas para conseguir este objetivo.

6.1. Método Monte Carlo

La simulación por el método de Monte Carlo (clásico o crudo) emplea números aleatorios para resolver ciertos problemas deterministas donde el transcurrir del tiempo no juega un papel sustancial.

El método se llamó así en referencia al Casino de Monte Carlo. Se originó y desarrolló durante la segunda guerra mundial cuando esta metodología fue aplicada a problemas relacionados al desarrollo de la bomba atómica.

Citando a uno de los grandes impulsores del desarrollo de este método [6] decimos que “el método de Monte Carlo consiste en representar la solución analítica de un problema como una propiedad de una población (hipotética) y hacer la estimación de esta propiedad (por ejemplo, a través de un parámetro) a partir de una muestra aleatoria”.

En general podemos decir que la aplicación del método de Monte Carlo implica seguir los siguientes pasos:

1. Formular analíticamente el problema.
2. Diseñar un modelo probabilista adecuado al problema.
3. Simular dicho modelo para generar una muestra aleatoria que permita realizar una estimación paramétrica.

Existen dos problemas que se abordan con frecuencia: integración y optimización. Empecemos examinando

algunos ejemplos donde es importante la aplicación del método de Monte Carlo.

Ejemplo 6.1.1. *Integración*

Supongamos que tenemos el siguiente problema de integración. Sean $f : \mathbb{R}^m \rightarrow [0, \infty)$ y h una función real integrable. El problema consiste en computar

$$\Lambda = \int_A h(x)f(x)dx, \quad A \subset \mathbb{R}^m.$$

Supongamos que la integral existe y es finita y que f es una función de densidad. También supondremos que el soporte de f , dado por el conjunto

$$\text{sup}(f) := \{x \in \mathbb{R} : f(x) > 0\},$$

es decir el conjunto de valores en donde f no se anula, está contenido dentro del conjunto A . Entonces ocurre que Λ es el valor esperado de h . Así, por la ley fuerte de los grandes números, tenemos que

$$\mathbb{E}(h(X)) = \int_A h(x)f(x)dx \approx \frac{1}{n} \sum_{i=1}^n h(X_i) =: \hat{\Lambda},$$

para una muestra aleatoria $\{X_1, \dots, X_n\}$ y una variable aleatoria X con densidad $f(x)$. Al valor $\hat{\Lambda}$ se le conoce como estimador de Monte Carlo de Λ .

Con este estimador, que se genera a partir de la simulación de variables aleatorias, podemos tener una solución a un problema que no es viable de resolver analíticamente. Tenemos las siguientes observaciones importantes:

1. $\mathbb{E}_f(\hat{\Lambda}) = \Lambda$.
2. $\text{Var}_f(\hat{\Lambda}) = \frac{1}{n} \int_A (h(x) - \Lambda)^2 f(x)dx,$

es decir, el estimador de Monte Carlo es un estimador insesgado y la varianza de dicho estimador también se puede estimar por el método de Monte Carlo.

A la desviación estándar de $\hat{\Lambda}$

$$\sigma_f(\hat{\Lambda}) = \sqrt{\text{Var}_f(\hat{\Lambda})}$$

se le conoce como error estándar. Sabemos que la ley fuerte de los grandes números implica que $\hat{\Lambda} \approx \Lambda$ cuando n crece a infinito, y la desigualdad de Chebyshev nos da una cota general para la probabilidad de que Λ esté a distancia $\epsilon > 0$ de su estimador :

$$\mathbb{P}(|\hat{\Lambda} - \Lambda| \geq \epsilon) \leq \frac{\text{Var}(\hat{\Lambda})}{\epsilon^2}.$$

Así, si elegimos

$$\epsilon = \sqrt{\text{Var}(\hat{\Lambda})/\delta},$$

ocurre que

$$|\hat{\Lambda} - \Lambda| \leq \sqrt{\text{Var}(\hat{\Lambda})/\delta}$$

con probabilidad $1 - \delta$. Como $\text{Var}(\hat{\Lambda}) = o(1/n)$, se sigue que la estimación de Λ por medio de $\hat{\Lambda}$ tiene un error de orden $n^{-1/2}$.

Ahora veamos un ejemplo concreto donde se utiliza el método Montecarlo para resolver un problema de integración.

Supongamos que queremos evaluar la integral

$$I = \int_a^b g(x)dx$$

donde $g(x)$ es una función que toma valores reales y que no es analíticamente integrable.

Para usar la simulación Monte Carlo, definimos $Y = (b - a)g(X)$ donde $X \sim U(a, b)$. Entonces ocurre que

$$\mathbb{E}(Y) = (b - a) \int_a^b g(x)f_X(x)dx = (b - a) \frac{\int_a^b g(x)dx}{(b - a)} = I$$

donde $f_X(x) = 1/(b - a)$ es la función de densidad de X . El problema de evaluar la integral se reduce a estimar el valor esperado $E(Y)$, que podemos hacer utilizando la media muestral de variables Y_1, \dots, Y_n independientes

$$\hat{I} = \frac{\sum_{i=1}^n Y_i}{n} = (b - a) \frac{\sum_{i=1}^n g(X_i)}{n},$$

donde X_1, X_2, \dots, X_n son v.a.i.i.d con distribución $U(a, b)$.

A continuación describimos otro ejemplo en donde podemos utilizar el método de Monte Carlo para variables aleatorias con un soporte infinito.

Ejemplo 6.1.2. Para $\alpha > 0$ queremos calcular

$$I_\alpha = \int_0^\infty x^{\alpha-1} e^{-x} dx.$$

Sea $f(x) = e^{-x}$ la densidad de X , es decir $X \sim \text{exp}(1)$. Notemos que

$$I_\alpha = E_f(X^{\alpha-1})$$

así I_α puede ser estimada por

$$\hat{I}_\alpha = \frac{1}{n} \sum_{i=1}^n X_i^{\alpha-1}$$

donde X_1, \dots, X_n es una muestra aleatorias con densidad f . A continuación se presentan resultados numéricos para el caso en que $\alpha = 1.9$. Recordemos que podemos generar de forma sencilla a las variables X_i 's, definiendo

$$X_i = -\ln(U_i)$$

donde las variables $U_i \sim U(0, 1)$, $i = 1, 2, \dots, n$ y son independientes. Presentamos los resultados obtenidos en la Tabla 6.1 a partir de la función EMC.GAMMA (ver B.3.15).

n	10	100	1,000	Tablas
$I_{1.9}$	0.8788	0.9312	0.9569	0.96177
Error Estándar	0.2302	0.0794	0.0277	

Tabla 6.1: Estimación para $\alpha = 1.9$

Ahora consideramos un ejemplo de optimización donde puede resultar de gran utilidad el método de Monte Carlo.

Ejemplo 6.1.3. Sea $h : [-1, 1] \times [-1, 1] \rightarrow \mathbb{R}$ definida por

$$h(x, y) = (x \operatorname{sen} 20y + y \operatorname{sen} 20x)^2 \operatorname{cosh}(x \operatorname{sen} 10x) + (x \operatorname{cos} 10y - y \operatorname{sen} 10x)^2 \operatorname{cosh}(y \operatorname{cos} 20y).$$

Utilizando el método de Monte Carlo encontraremos las posiciones de los valores máximo y mínimo y los valores máximo y mínimo alcanzados por la función h .

En este caso el método de Monte Carlo crudo se traduce en generar números aleatorios $\{u_1, \dots, u_n\}$ en $[-1, 1] \times [-1, 1]$ y calcular $\{h(u_1), \dots, h(u_n)\}$. De estos últimos, seleccionemos las parejas $(u^{max}, h(u^{max}))$ y $(u^{min}, h(u^{min}))$ tales que $h(u^{max}) = \max \{h(u_1), \dots, h(u_n)\}$ y $h(u^{min}) = \min \{h(u_1), \dots, h(u_n)\}$. En la Tabla 6.2 se presentan resultados numéricos para diferentes tamaños de muestra.

n	$h(u^{min})$	u^{min}	$h(u^{max})$	u^{max}
10	0.22×10^{-3}	(-0.0286, 0.0439)	2.061	(0.749, 0.875)
100	3.27×10^{-6}	(-0.0003, -0.0291)	5.301	(-0.083, 0.091)
1000	6.46×10^{-8}	(-0.00061, -0.017)	5,218	(-0.072, 0.099)
10000	4.08×10^{-8}	(-0.00012, -0.0055)	5.430	(-0.085, 0.088)
100000	1.10×10^{-9}	(-0.00004, -0.0009)	5.739	(0.099, -0.098)

Tabla 6.2: Estimadores de máximos y mínimos de h .

Queremos remarcar que este método de optimización, por medio de Monte Carlo crudo, es poco eficiente: estamos haciendo una búsqueda aleatoria de forma independiente. Esto tiene algunas similitudes a realizar fuerza bruta computacional, que es realizar una búsqueda de diccionario (enlistar todos los elementos) para encontrar los valores deseados. Sin embargo su implementación es muy sencilla; el único requisito es poder evaluar la función f que estamos intentando optimizar.

6.2. Reducción de la varianza

En esta sección nos enfocaremos en algunos aspectos teóricos y prácticos de las técnicas de reducción de varianza, motivados en encontrar estimadores de Monte Carlo de varianza mínima. El objetivo es utilizar de mejor forma nuestro poder computacional (y por lo tanto nuestra precisión) al utilizar estimadores más precisos que los dados por el estimador de Monte Carlo crudo.

Definición 6.1. Podemos definir a una técnica de reducción de la varianza como un medio para obtener estimadores más precisos utilizando información conocida de nuestro modelo.

Por lo general, entre mayor sea la información con la que contemos sobre nuestro modelo, más efectiva será la reducción de la varianza. Las técnicas principales y más efectivas para la reducción de la varianza

son el muestreo por importancia y el método de Monte Carlo condicional; ambos serán discutidos en esta sección. Otras técnicas que pueden ayudar a la reducción de la varianza son el uso de variables antitéticas, de control y la estratificación.

Como motivación de la importancia del uso de las técnicas de reducción de varianza, comenzaremos considerando el siguiente ejemplo, tomado de [22], que ilustra la importancia de este concepto en el método de Monte Carlo.

Ejemplo 6.2.1. *Suponga que $X \sim \text{Cauchy}(0, 1)$ y que estamos interesados en evaluar $P[X \geq 2]$, es decir*

$$\theta = \mathbb{P}(X \geq 2) = \int_2^{\infty} \frac{dx}{\pi(1+x^2)}. \quad (6.1)$$

Como esta integral tiene solución analítica, encontraremos dicha solución y nos servirá de referencia para evaluar qué tan buenos son los estimadores que calcularemos. Para encontrar la solución, hacemos el cambio de variable

$$x(\phi) = \tan \phi.$$

Entonces

$$1 + x^2 = 1 + \tan^2 \phi = \sec^2 \phi$$

y

$$dx = \sec^2 \phi d\phi.$$

Por otra parte, como $\phi(x) = \tan^{-1} x$, ocurre que $\phi(\infty) = \pi/2$, y se concluye

$$\theta = \int_2^{\infty} \frac{dx}{\pi(1+x^2)} = \frac{1}{\pi} \int_{\tan^{-1} 2}^{\pi/2} d\phi = \frac{1}{2} - \frac{1}{\pi} \tan^{-1} 2 \approx 0.147584.$$

Proponemos un primer estimador dado por

$$\tilde{\theta}_1 = \frac{1}{n} \sum_{i=1}^n I_{[2, \infty)}(X_i).$$

donde la muestra $\{X_1, \dots, X_n\}$ tiene distribución *Cauchy*(0, 1). Es claro que $n\tilde{\theta}_1$ es una variable aleatoria Binomial con parámetros (n, θ) y entonces

$$\text{Var}(\tilde{\theta}_1) = \frac{\theta(1-\theta)}{n} \approx \frac{0.125803}{n}.$$

La varianza no parece tan pequeña.

Utilizando la simetría de la densidad Cauchy respecto al cero, podemos calcular θ como

$$2\theta = \mathbb{P}(|X| \geq 2) = 1 - \int_{-2}^2 \frac{dx}{\pi(1+x^2)}$$

por lo que

$$\theta = \frac{1}{2} \int_{\mathbb{R} \setminus (-2,2)} \frac{dx}{\pi(1+x^2)}$$

y entonces proponemos un segundo estimador de θ por

$$\tilde{\theta}_2 = \frac{1}{2n} \sum_{i=1}^n I_{\mathbb{R} \setminus (-2,2)}(X_i)$$

con una muestra $\{X_1, \dots, X_n\}$ con distribución *Cauchy*(0, 1). Como $2n\tilde{\theta}_2 \sim \text{Bin}(n, 2\theta)$, la varianza de $\tilde{\theta}_2$ es aproximadamente:

$$\text{Var}(\tilde{\theta}_2) \approx \frac{0.0520109}{n},$$

que es menor que la de $\tilde{\theta}_1$.

Hasta ahora hemos propuesto dos estimadores y tenemos un criterio de decisión saber cual es mejor, ambos están basados en la simulación de variables aleatorias Cauchy, pero un buen punto a resaltar es que no tenemos la certeza de haber encontrado el estimador de menor varianza.

Continuando en la búsqueda de dicho estimador, notamos que

$$1 - 2\theta = \int_{-2}^2 \frac{dx}{\pi(1+x^2)} = 2 \int_0^2 \frac{dx}{\pi(1+x^2)}.$$

Entonces podemos generar números aleatorios $X_i \sim U(0, 2)$ y proponer un tercer estimador dado por:

$$\tilde{\theta}_3 = \frac{1}{2} - \frac{2}{n} \sum_{i=1}^n \frac{1}{\pi(1+X_i^2)}$$

cuya varianza está dada por

$$\text{Var}(\tilde{\theta}_3) = \frac{4}{n} \left\{ \int_0^2 \frac{1}{2} \left(\frac{1}{\pi(1+x^2)} \right)^2 dx - \left(\int_0^2 \frac{1}{2\pi(1+x^2)} dx \right)^2 \right\}$$

que resulta ser

$$\text{Var}(\tilde{\theta}_3) \approx \frac{0.0285088}{n},$$

entonces tenemos un estimador de varianza menor que la de los dos anteriores que propusimos.

Finalmente, si hacemos $y = 1/x$ tenemos que

$$\theta = \int_2^{\infty} \frac{dx}{\pi(1+x^2)} = \int_0^{1/2} \frac{y^{-2}dy}{\pi(1+y^{-2})} = \int_0^{1/2} \frac{dy}{\pi(1+y^2)}.$$

Entonces, si generamos $Y_i \sim U(0, 1/2)$, tenemos un nuevo estimador

$$\tilde{\theta}_4 = \frac{1}{2\pi n} \sum_{i=1}^n \frac{1}{1+Y_i^2},$$

cuya varianza aproximada es

$$\text{Var}(\tilde{\theta}_4) \approx \frac{0.0000955253}{n}.$$

En este caso, el estimador $\tilde{\theta}_4$ construido con una sola observación tiene aproximadamente la misma precisión que $\tilde{\theta}_1$ construido con 1000 observaciones.

Este ejemplo pone en evidencia el hecho de que no es una tarea fácil encontrar el estimador de menor varianza simplemente proponiendo diferentes estimadores, motivando la necesidad de contar con técnicas y metodología que nos permita reducir la varianza de los estimadores.

6.2.1. Muestreo por importancia.

Regresemos al problema de resolver la integral dada por (6.1) y escribámosla de la siguiente manera:

$$\theta = \int_2^{\infty} \frac{dx}{\pi(1+x^2)} = \int_{-\infty}^{\infty} \psi(x)f(x)dx$$

donde $\psi(x) = I_{[2,\infty)}(x)$ y f es la densidad dada por $f(x) = [\pi(1+x^2)]^{-1}$. Supongamos que g es una función de densidad fácil de muestrear. Entonces

$$\theta = \mathbb{E}_f(\psi(X)) = \int_{-\infty}^{\infty} \frac{f(x)\psi(x)}{g(x)}g(x)dx = \int_{-\infty}^{\infty} \phi(x)g(x)dx = \mathbb{E}_g(\phi(X))$$

donde

$$\phi(x) = \frac{f(x)\psi(x)}{g(x)}$$

y la notación \mathbb{E}_f significa que dentro de tal operador, la variable aleatoria X tiene la densidad f . Por lo tanto, un estimador posible para θ es

$$\tilde{\theta}_5 = \frac{1}{n} \sum_{i=1}^n \phi(X_i)$$

con $\{X_1, \dots, X_n\}$ una muestra aleatoria con densidad g . La varianza de este estimador está dada por

$$\text{Var}(\tilde{\theta}_5) = \frac{1}{n} \mathbb{E} \left[\left(\tilde{\theta}_5 - \theta \right)^2 \right] = \frac{1}{n} \int_{-\infty}^{\infty} \left(\frac{f(x)\psi(x)}{g(x)} - \theta \right)^2 g(x)dx,$$

la cual será mínima entre más similar sea g al integrando.

En el ejemplo de la densidad Cauchy, debemos buscar una función, $h(x)$, que tenga una forma similar y que sea fácil de simular. Para conseguir dicho objetivo, primero observemos que la función decrece de manera similar a $\frac{1}{x^2}$, entonces, es natural proponer a $h(x) = 1/ax^2$, para algún $a > 0$. Tomando h de la forma propuesta se tiene que

$$\int_2^{\infty} h(x)dx = \frac{1}{2a},$$

entonces

$$g(x) = \frac{2}{x^2}I_{(2,\infty)}(x),$$

es una función de densidad cuya función de distribución está dada por

$$G(x) = \int_2^x g(t)dt = 1 - \frac{2}{x}.$$

Ahora, utilizando el método de la transformada inversa, tenemos que si $U \sim U(0,1)$, entonces $X = 2/(1 - U)$ es una variable aleatoria con densidad $g(x)$. Por ello podemos estimar a θ por medio de

$$\tilde{\theta}_6 = \frac{1}{n} \sum_{j=1}^n \frac{X_j^2}{2\pi(1 + X_j^2)} I_{[2,\infty)}(X_j),$$

donde $\{X_1, \dots, X_n\}$ es una muestra aleatoria con densidad $g(x)$. Podemos calcular explícitamente la varianza de $\tilde{\theta}_6$:

$$\text{Var} [\tilde{\theta}_6] = \frac{1}{4n\pi^2} \left\{ \int_2^{\infty} \left(\frac{x^2}{1+x^2} \right)^2 \frac{2}{x^2} dx - \left[\int_2^{\infty} \left(\frac{x^2}{1+x^2} \right) \frac{2}{x^2} dx \right]^2 \right\} = \frac{0.0003821}{n},$$

que representa una reducción en 329 veces de la varianza del estimador de Monte Carlo crudo. Es decir los tamaños de muestra correspondientes n_1 y n_6 de cada estimador deben satisfacer la relación $n_6 = 0.003n_1$ para tener la misma varianza.

Extrapolando lo anterior, podemos ver que el método de muestreo de importancia consiste en escribir un valor por estimar θ , de la forma

$$\theta = \mathbb{E}_f(\psi(X)) = \int_{-\infty}^{\infty} \frac{f(x)\psi(x)}{g(x)} g(x)dx = \int_{-\infty}^{\infty} \phi(x)g(x)dx = \mathbb{E}_g(\phi(X)).$$

A la función g se le llama **función de muestreo por importancia**.

En resumidas cuentas cuando se desea estimar θ por esta técnica se debe encontrar una función de densidad g cuyo soporte contenga al soporte de $\psi \cdot f$ y que sea fácil de simular. En tal caso

$$\theta = \mathbb{E}_g \left(\psi(X) \frac{f(X)}{g(X)} \right).$$

Por lo tanto, si $\{X_1, \dots, X_n\}$ es una muestra aleatoria con densidad g entonces

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n \psi(X_i) \frac{f(X_i)}{g(X_i)}$$

es un estimador insesgado de θ , llamado **estimador de muestreo por importancia**. El cociente de densidades

$$\frac{f(x)}{g(x)}$$

es conocido como la razón o **cociente de verosimilitud**. Es importante hacer notar que una gran virtud de este procedimiento es que una función g dada puede ser usada para encontrar el valor esperado de cualquier función ψ .

Por otra parte, aunque la condición

$$\int_{-\infty}^{\infty} \frac{\psi(x)^2 f(x)^2}{g(x)} dx < \infty$$

no es una condición necesaria para la convergencia del estimador de muestreo de importancia a θ , sí es necesaria para la existencia de la varianza del estimador. Por lo tanto, podemos concluir que el muestreo por importancia tiene poca eficiencia cuando (6.2.1) no se cumple, pues produce inestabilidad en la generación de los números aleatorios y en consecuencia la estimación no es confiable, por lo que funciones g que no cumplan esta condición no son recomendables. Esto implica que g deberá tener colas más pesadas que f .

Minimización de varianza

Debido a la importancia que tiene la elección de la densidad g en el método de muestreo por importancia del estimador dado por (6.2.1), se puede plantear el problema de minimización siguiente:

$$\min_g \text{Var}_g \left(\psi(X) \frac{f(X)}{g(X)} \right). \quad (6.2)$$

Se puede probar que la solución a este problema está dada por

$$g^*(x) = \frac{|\psi(x)|f(x)}{\int_{-\infty}^{\infty} |\psi(x)|f(x)dx}, \quad (6.3)$$

es decir g^* es la función $|\psi(x)|f(x)$ normalizada para ser una densidad.

Nota 6.1. En el caso en que $\psi(x) \geq 0$ entonces

$$g^*(x) = \frac{\psi(x)f(x)}{\theta}, \quad (6.4)$$

y

$$\text{Var}_{g^*}(\hat{\theta}) = \text{Var}_{g^*} \left(\frac{\psi(X)f(X)}{g^*(X)} \right) = \text{Var}_{g^*}(\theta) = 0.$$

A la densidad g^* de las expresiones (6.3) y (6.4) se le llama **densidad óptima de muestro por importancia**.

Ejemplo 6.2.2. Sea $X \sim \text{Exp}(\lambda)$ y $\psi(x) = I_{\{x \geq \tau\}}$ para algún $\tau > 0$. Sea f la función de densidad de X y supongamos que queremos estimar:

$$\theta = \mathbb{E}_f(\psi(X)).$$

Ocurre que

$$g^*(x) = I_{\{x \geq \tau\}} \lambda e^{-(x-\tau)\lambda}.$$

6.2.2. Monte Carlo condicional

Comenzaremos recordando algunas propiedades de esperanza condicional que servirán de motivación para este método de reducción de varianza, llamado Monte Carlo condicional que también se conoce como **Rao-Blackwellization**.

Consideremos a dos variables aleatorias cualesquiera, X e Y . Recordemos que la (variable aleatoria) variancia condicional $\text{Var}(X|Y)$ está definida por

$$\text{Var}(X|Y) := \mathbb{E}((X - E(X|Y))^2|Y),$$

y que también se puede computar de la siguiente manera

$$\text{Var}(X|Y) = \mathbb{E}(X^2|Y) - \mathbb{E}(X|Y)^2.$$

Así que al tomar esperanza en la ecuación anterior ocurre que

$$\begin{aligned} \mathbb{E}(\text{Var}(X|Y)) &= \mathbb{E}(\mathbb{E}(X^2|Y)) - \mathbb{E}(\mathbb{E}(X|Y)^2) \\ &= \mathbb{E}(X^2) - \mathbb{E}(\mathbb{E}(X|Y)^2). \end{aligned}$$

Por otro lado,

$$\begin{aligned} \text{Var}(\mathbb{E}(X|Y)) &= \mathbb{E}(\mathbb{E}(X|Y)^2) - (\mathbb{E}(\mathbb{E}(X|Y)))^2 \\ &= \mathbb{E}(\mathbb{E}(X|Y)^2) - \mathbb{E}(X)^2. \end{aligned}$$

Sumando ambas expresiones se tiene que

$$\mathbb{E}(\text{Var}(X|Y)) + \text{Var}(\mathbb{E}(X|Y)) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \text{Var}(X),$$

de lo que se sigue que

$$\text{Var}(\mathbb{E}(X|Y)) \leq \text{Var}(X),$$

pues la variable aleatoria $\text{Var}(X|Y)$ es no negativa.

La desigualdad obtenida nos muestra que condicionar siempre conduce a la reducción de la varianza. Utilizaremos esta idea para proponer un método de reducción de varianza.

Supongamos que deseamos estimar

$$l = \mathbb{E}(h(X)) = \int h(x)f(x)dx,$$

donde X es una variable con función de densidad f y h es una función integrable. Supongamos que existe una variable aleatoria Y con función de densidad g (fácil de simular) donde $\mathbb{E}(h(X)|Y = y)$ es una expresión se puede calcular explícitamente. Entonces, debido a que

$$l = \mathbb{E}(h(X)) = \mathbb{E}(\mathbb{E}(h(X)|Y)),$$

se tiene que $\mathbb{E}(h(X)|Y)$ es un estimador insesgado de l y, por lo visto al inicio de esta sección, ocurre que

$$\text{Var}(\mathbb{E}(h(X)|Y)) \leq \text{Var}(h(X)).$$

Esto nos dice que usar como estimador a la variable aleatoria $\mathbb{E}(h(X)|Y)$, en lugar de la variable $h(X)$, conduce a la reducción de la varianza.

Ahora estamos listos para proponer el algoritmo formalmente.

Algoritmo 47 : Monte Carlo condicional

- 1: Generar una muestra Y_1, \dots, Y_n de la densidad g .
 - 2: Calcular $\mathbb{E}(h(X)|Y_k)$, $k = 1, \dots, n$ analíticamente.
 - 3: Estimar $l = \mathbb{E}(h(X))$ mediante $\hat{l} = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(h(X)|Y_k)$.
-

Como podemos ver, este algoritmo requiere encontrar una variable aleatoria Y con las siguientes características:

- Debe ser fácil de simular.
- Debe poderse calcular analíticamente $\mathbb{E}(h(X)|Y)$.
- Es deseable que el valor $\mathbb{E}(\text{Var}(h(X)|Y))$ sea lo más grande posible en comparación con $\text{Var}(\mathbb{E}(h(X)|Y))$.

Veamos un ejemplo del uso del Algoritmo que acabamos de presentar.

Ejemplo 6.2.3. Consideremos la siguiente integral

$$G = \int_0^1 \int_0^1 g(x, y) dx dy,$$

donde

$$g(x, y) = \begin{cases} 1 & \text{si } x^2 + y^2 \leq 1 \\ 0 & \text{si } x^2 + y^2 > 1 \end{cases} \quad (6.5)$$

Si muestreamos uniformemente y de manera independiente X e Y en $[0, 1]$, es decir, $(X, Y) \sim \text{Unif}([0, 1]^2)$ podemos definir al estimador de G (que es el estimador de Monte Carlo crudo con una sola observación) de la siguiente manera:

$$\hat{G} = g(X, Y) = \mathbf{1}_{\{x^2+y^2\}}(X, Y).$$

Ocurre que

$$E(\hat{G}) = G = \frac{\pi}{4}$$

y se puede calcular analíticamente su varianza

$$\text{Var}(\hat{G}) = \frac{\pi}{4} - \left(\frac{\pi}{4}\right)^2 = 0.168.$$

Utilizando propiedades de la esperanza condicional tenemos que:

$$\mathbb{E}(g(X, Y)) = \mathbb{E}(\mathbb{E}(g(X, Y)|Y)).$$

Dado y, podemos computar:

$$\mathbb{E}(g(X, Y)|Y = y) = \int_0^1 \mathbb{1}_{\{x \leq \sqrt{1-y^2}\}} f_{X|Y}(x|y) dx.$$

Como X es independiente de Y y X tiene distribución $Unif([0, 1])$ se tiene que,

$$f_{X|Y}(x|y) = f_X(x) = \mathbb{1}_{[0,1]}(x)$$

por lo que sustituyendo en la expresión para $\mathbb{E}(g(X, Y)|Y = y)$, se sigue que

$$\mathbb{E}(g(X, Y)|Y = y) = \int_0^1 \mathbb{1}_{\{x \leq \sqrt{1-y^2}\}} dx = \int_0^{\sqrt{1-y^2}} dx = \sqrt{1-y^2}.$$

Concluimos que

$$\mathbb{E}(g(X, Y)|Y) = \sqrt{1-Y^2}$$

y

$$\mathbb{E}(g(X, Y)) = \mathbb{E}(\sqrt{1-Y^2}).$$

Por lo tanto, un estimador de Monte Carlo condicional (de una única observación) está dado por $\sqrt{1-Y^2}$, donde Y tiene distribución uniforme en $[0, 1]$. La varianza de este nuevo estimador está dada por

$$\int_0^1 (1-y^2) dy - \left(\frac{\pi}{4}\right)^2 = 0.050.$$

Obtuvimos una reducción de la varianza cercana a un tercio de la del estimador de Monte Carlo crudo.

Presentamos otro ejemplo de Monte Carlo condicional.

Ejemplo 6.2.4. Sea $\{X_i\}_{i=1}^\infty$ una sucesión de variables aleatorias independientes e idénticamente distribuidas tales que X_i tiene distribución F y una variable aleatoria M , que toma valores en los naturales, independiente de la sucesión $\{X_i\}_{i=1}^\infty$. Definamos $S_M = \sum_{i=1}^M X_i$. Queremos estimar $l = \mathbb{P}(S_M \leq x)$, para un x dado.

Denotemos por F^m a la función de distribución de $S_m = \sum_{i=1}^m X_i$, para un número natural m . Entonces tenemos que

$$F^m(x) = \mathbb{P}\left(\sum_{i=1}^m X_i \leq x\right) = F\left(x - \sum_{i=2}^m X_i\right).$$

Por otro lado, como

$$l = \mathbb{E}(I_{\{S_M \leq x\}}) = \mathbb{E}\left[\mathbb{E}\left(I_{\{S_M \leq x\}} \mid \sum_{i=2}^M X_i\right)\right] = \mathbb{E}\left[F\left(x - \sum_{i=2}^M X_i\right)\right],$$

podemos estimar a esta esperanza por medio de

$$\hat{l} = \frac{1}{n} \sum_{i=1}^n F\left(x - \sum_{j=2}^M X_j^i\right).$$

6.2.3. Variables de control

Supongamos que deseamos estimar $\theta := \mathbb{E}(Y)$ donde $Y = h(X)$ es el resultado de un experimento de simulación. Además supongamos que Z es otro resultado (que posiblemente resultado del proceso de simular) que puede ser fácilmente calculado. Finalmente asumamos que conocemos $\mathbb{E}[Z]$. Con esto en consideración podemos construir varios estimadores insesgados de θ :

$$\hat{\theta}_c = Y + c(Z - \mathbb{E}[Z]),$$

donde $c \in \mathbb{R}$ (en el caso en que $c = 0$, recuperamos el estimador de Monte Carlo crudo). Es claro que $\mathbb{E}[\hat{\theta}_c] = \theta$. ¿Cuándo ocurre que $\hat{\theta}_c$ tiene menor varianza que $\hat{\theta}$?

Notemos que:

$$\text{Var}(\hat{\theta}_c) = \text{Var}(Y) + c^2 \text{Var}(Z) + 2c \text{Cov}(Y, Z).$$

Como esta expresión es válida para cualquier c , podemos minimizar esta varianza como una función de c . Obtenemos el mínimo de la varianza cuando:

$$c^* = -\frac{\text{Cov}(Y, Z)}{\text{Var}(Z)}.$$

Así, ocurre que

$$\text{Var}(\hat{\theta}_{c^*}) = \text{Var}(Y) - \frac{\text{Cov}(Y, Z)^2}{\text{Var}(Z)} = \text{Var}(\hat{\theta}) - \frac{\text{Cov}(Y, Z)^2}{\text{Var}(Z)},$$

de modo que para reducir la varianza basta que el valor $\text{Cov}(Y, Z)$ sea distinto de cero.

Lo anterior nos sugiere el uso del siguiente estimador:

$$\hat{\theta}_{c^*} = \frac{\sum_{i=1}^n (Y_i + c^*(Z_i - \mathbb{E}[Z]))}{n},$$

donde Z es conocida como la **variable de control del estimador**. Debido a que necesitamos conocer el valor c^* para implementar tal estimador, lo que podemos hacerlo es estimarlo, por medio de estimadores estadísticos usuales:

$$\hat{\text{Cov}}(Y, Z) = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(Z_i - \mathbb{E}(Z))}{n-1},$$

$$\hat{\text{Var}}(Z) = \frac{\sum_{i=1}^n (Z_i - \mathbb{E}(Z))^2}{n-1},$$

para aproximar c^* por

$$\hat{c}^* = -\frac{\hat{\text{Cov}}(Y, Z)}{\hat{\text{Var}}(Z)}.$$

Ejemplo 6.2.5. Supongamos que deseamos estimar $\theta = \mathbb{E}[e^{(U+W)^2}]$ donde U y $W \sim U(0, 1)$ son variables aleatorias $Unif(0, 1)$ independientes. Utilizemos a $Z := (U + W)^2$ como variable de control. Notemos que en este caso

$$\mathbb{E}(Z) = \mathbb{E}(U^2 + 2UW + W^2) = \frac{1}{3} + \frac{1}{2} + \frac{1}{3} = \frac{7}{6}.$$

6.2.4. Variables antitéticas

En esta subsección abordaremos un método para reducir la varianza de los estimadores de Monte Carlo en donde se busca no añadir esfuerzo computacional, y que puede ser útil en diversas situaciones.

Suponga que

$$I = \int_{-\infty}^{\infty} f(x)g(x)dx,$$

donde f es una función de densidad, y donde tenemos dos estimadores insesgados de I basados en las muestras $\{x_1, \dots, x_n\}$ e $\{y_1, \dots, y_m\}$, identificados por $I_1 = I_1(x_1, \dots, x_n)$ e $I_2 = I_2(y_1, \dots, y_m)$ respectivamente. Además, supongamos que las respectivas varianzas son $\text{Var}(I_1)$ y $\text{Var}(I_2)$. Entonces

$$\hat{I} := \frac{I_1 + I_2}{2}$$

es un estimador insesgado de I con varianza dada por

$$\text{Var}(\hat{I}) = \frac{1}{4}\text{Var}(I_1) + \frac{1}{4}\text{Var}(I_2) + \frac{1}{2}\text{Cov}(I_1, I_2).$$

Si $\{x_1, \dots, x_n\}$ e $\{y_1, \dots, y_m\}$ son independientes y provienen de la misma distribución, entonces habremos reducido la varianza a la mitad del promedio de las varianzas originales. Entonces, uno podría pensar que en este proceso redujimos la varianza. Esto es sin embargo una ilusión: por ejemplo, supongamos que $n = m$ y que las muestras son independientes. Entonces el estimador \hat{I} utiliza $2n$ observaciones para reducir la varianza a la mitad, que es la varianza correspondiente al estimador de Monte Carlo crudo con $2n$ observaciones.

Pero, ¿qué pasa si las muestras $\{x_1, \dots, x_n\}$ y $\{y_1, \dots, y_m\}$ no son independientes? , en este caso, $\text{Cov}(I_1, I_2)$ no es cero y su valor puede ser positivo o negativo. Si las muestras son correlacionadas negativamente, la varianza del estimador \hat{I} se reducirá. Concluimos que en la búsqueda de la reducción de la varianza es importante utilizar muestras correlacionadas negativamente.

Para fijar ideas, supongamos que $\text{Var}(I_1) = \text{Var}(I_2)$. Entonces

$$\begin{aligned} \text{Var}(\hat{I}) &= \frac{1}{2}\text{Var}(I_1) + \frac{1}{2}\text{Cov}(I_1, I_2) \\ &= \frac{1}{2}\text{Var}(I_1) \left\{ 1 + \frac{\text{Cov}(I_1, I_2)}{\sqrt{\text{Var}(I_1)\text{Var}(I_2)}} \right\} \\ &= \frac{1}{2}\text{Var}(I_1)\{1 + \rho(I_1, I_2)\}, \end{aligned}$$

donde $\rho(X, Y)$ denota al coeficiente de correlación entre las variables aleatorias X e Y . Entonces, la varianza de \hat{I} será mucho menor que la de I_1 si la correlación es negativa y cercana a -1 .

En este punto, la pregunta natural es, ¿Es posible generar variable correlacionadas negativamente sin aumentar demasiado el tamaño de la muestra original? La respuesta es afirmativa y además, asombrosamente, no se generan observaciones adicionales, al menos teóricamente. Veamos un ejemplo de esto.

Ejemplo 6.2.6. *Supongamos que queremos calcular*

$$I = \int_0^1 \sqrt{1-x^2} dx.$$

En este caso, esta integral tiene solución analítica. Si hacemos $x = \sin \theta$ y por lo tanto, $dx = \cos \theta d\theta$, tenemos que

$$\begin{aligned} I &= \int_0^{\pi/2} \cos^2 \theta d\theta = \int_0^{\pi/2} \left(\frac{1}{2} + \frac{1}{2} \cos 2\theta \right) d\theta \\ &= \left[\frac{\theta}{2} + \frac{1}{4} \sin 2\theta \right]_0^{\pi/2} = \frac{\pi}{4}. \end{aligned}$$

El estimador por Monte Carlo simple es

$$I_1 = \frac{1}{n} \sum_{i=1}^n \sqrt{1-U_i^2}$$

donde $\{U_1, \dots, U_n\}$ es una muestra de $U(0, 1)$ y cuya varianza resulta es

$$\text{Var}(I_1) = \frac{1}{n} \left[\int_0^1 (1-u^2) du - I^2 \right] \approx \frac{0.0498}{n}$$

Ahora, consideremos el estimador (de variables antitéticas) dado por

$$\hat{I} = \frac{1}{n} \sum_{i=1}^n \frac{1}{2} \left[\sqrt{1-U_i^2} + \sqrt{1-(1-U_i)^2} \right]$$

donde $\{U_1, \dots, U_n\}$ es una muestra proveniente de la distribución $U(0, 1)$. Entonces ocurre que

$$\text{Var}(\hat{I}) = \frac{1}{4n} \int_0^1 \left(\sqrt{1-u^2} + \sqrt{1-(1-u)^2} - 2I \right)^2 du \approx \frac{0.0068}{n}.$$

Hubo una reducción aproximada de la varianza original del 73 %.

6.3. Ejercicios

1. A partir de la expresión

$$\pi = 4 \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1},$$

estimar el valor de π por el método de Monte Carlo.

2. Implementar el Ejemplo 7.1.2 y estimar el valor k .

3. Aproximar la integral de la función de densidad de una v.a. Normal(0,1) con límite inferior igual a -3 y límite superior igual a 5 y calcular la desviación estándar del estimador.
4. Aproximar la integral del kernel de una v.a. gamma y calcula la desviación estándar del estimador.
5. Calcular los 4 estimadores para la probabilidad de que $X > 2$, donde $X \sim Cauchy(0,1)$ y sus varianzas
 - a) Usando Cauchy y la indicadora de 2 a infinito.
 - b) Pegándole una uniforme, con una transformación del estimador aprovechando la simetría.
 - c) Usando la Cauchy, con una indicadora de $(-2,2)$ complemento.
 - d) Haciendo el cambio de variable $x = 1/y$.
6. Para esa misma Cauchy pero ahora con muestreo por importancia y su varianza
7. Monte carlo condicional
 - a) X distribuye $\exp(10)$.
 - b) M sea va Poisson (2).
8. Ejemplo 2 de reducción de varianza , implementar en R (reducción de varianza con dos muestras correlacionadas negativamente, para la integral de 0 a 1 de la función $\sqrt{(1+x^2)}$.
9. Probar que solución al problema de minimización de varianza en el método de muestro por importancia es el de la expresión 6.3.
10. Estimar el parámetro del Ejemplo 6.2.2.
11. Del ejercicio anterior, analizar el caso $V(I_1) < V(I_2)$ en R.
12. En el ejemplo de las diferentes varianzas, analizar el caso en que $\mathbb{V}ar(I_1) > \mathbb{V}ar(I_2)$.