

Simulación Estocástica

Integración de Monte Carlo y técnicas de reducción de varianza

Vanda Inácio de Carvalho

Primer Semestre 2015

Integración de Monte Carlo

- Sea $g(x)$ una función y suponga que queremos calcular $\int_a^b g(x)dx$ (asumiendo que la integral existe).
- Recuerde que si X es una variable aleatoria con función de densidad $f(x)$ y soporte en el intervalo $[a, b]$, entonces la esperanza de la variable aleatoria $Y = g(X)$ es

$$\mathbb{E}(g(X)) = \int_a^b g(x)f(x)dx.$$

- Empezemos con el caso más sencillo, el estimador de Monte Carlo simple.
- La idea del método es escribir las integrales como esperanzas.
- Considere el problema de estimar $I = \int_0^1 g(x)dx = \mathbb{E}(g(X))$, $X \sim U(0, 1)$.
- Si $X_1, \dots, X_m \stackrel{\text{iid}}{\sim} U(0, 1)$, entonces

$$\hat{I} = \frac{1}{m} \sum_{i=1}^m g(X_i)$$

converge para $\mathbb{E}(g(X)) = I$ con probabilidad 1, cuando $m \rightarrow \infty$ por la ley fuerte de los grandes números.

Integración de Monte Carlo

Ejemplo

- Calcular una estimación de Monte Carlo de $I = \int_0^1 e^{-x} dx$ y comparar con el valor exacto.

- El valor exacto es

$$I = [-e^{-x}]_0^1 = 1 - e^{-1} \approx 0.6321.$$

- Sea $X_i \stackrel{iid}{\sim} U(0, 1)$, $i = 1, \dots, m$. Entonces,

$$I \approx \hat{I} = \frac{1}{m} \sum_{i=1}^m e^{-X_i}.$$

```
set.seed(123)
m=10000; x=runif(m)
I.hat=mean(exp(-x))
0.6333057
```

Integración de Monte Carlo

- Para calcular $\int_a^b g(x)dx$, una posibilidad es hacer una transformación de variable tal que los límites de integración sean de 0 a 1.
- La transformación lineal es

$$y = \frac{x - a}{b - a} \quad y \quad \frac{dx}{dy} = b - a.$$

- Reemplazando, tenemos

$$\int_a^b g(x)dx = \int_0^1 g((b - a)y + a)dy.$$

- Alternativamente, se puede reemplazar la distribución $U(0, 1)$ por la distribución $U(a, b)$.
- Así,

$$I = \int_a^b g(x)dx = (b - a) \int_a^b \frac{1}{b - a} g(x)dx = (b - a)\mathbb{E}(g(X)), \quad X \sim U(a, b).$$

Integración de Monte Carlo

• Entonces,

$$\hat{\lambda} = (b - a) \frac{1}{m} \sum_{i=1}^m g(X_i), \quad X_i \stackrel{iid}{\sim} U(a, b), \quad i = 1, \dots, m.$$

Algoritmo

Algoritmo para calcular $\int_a^b g(x)dx$.

- 1 Generar $X_1, \dots, X_m \stackrel{iid}{\sim} U(a, b)$.
- 2 Calcular $g(X_1), \dots, g(X_m)$.
- 3 Calcular $\bar{g}(X) = \frac{1}{m} \sum_{i=1}^m g(X_i)$.
- 4 Hacer $\hat{\lambda} = (b - a)\bar{g}(X)$.

Integración de Monte Carlo

Ejemplo

- *Calcular una estimación de Monte Carlo de*

$$I = \int_0^{2\pi} e^{\cos(x)} dx.$$

- *Solución*

$$I \approx \hat{I} = 2\pi \frac{1}{m} \sum_{i=1}^m e^{\cos(X_i)}, \quad X_i \stackrel{iid}{\sim} U(0, 2\pi), \quad i = 1, \dots, m.$$

```
f=function(x){exp(cos(x))}  
I=integrate(f,lower=0,upper=2*pi)$value  
  
set.seed(123); m=10000; x=runif(m,0,2*pi)  
I.hat=2*pi*mean(exp(cos(x)))  
I.hat
```

Integración de Monte Carlo

Ejemplo

- Sea $X \sim \text{Cauchy}(0, 1)$ y suponga que queremos calcular $\Pr(X > 2)$.
- Consideremos $A = \{x : x > 2\}$ y notemos que

$$\Pr(X > 2) = \int_{-\infty}^{\infty} I_A(x) f(x) dx = \int_2^{\infty} f(x) dx = \int_2^{\infty} \frac{1}{\pi(1+x^2)} dx.$$

- El valor exacto da probabilidad referida es $1 - p_{\text{cauchy}}(2, 0, 1) \approx 0.1476$
- El algoritmo presentado anteriormente no se puede aplicar directamente porque la integral no está definida en un intervalo limitado.
- Sin embargo, haciendo la transformación $y = 1/x$, la integral pasa a estar definida en un intervalo limitado.
- Veamos
 - $x = 2 \Rightarrow y = 1/2$.
 - $x = \infty \Rightarrow y = 0$.

Integración de Monte Carlo

Ejemplo

- Además, tenemos

$$y = \frac{1}{x} \Rightarrow \frac{dx}{dy} = -\frac{1}{y^2}.$$

- Así,

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

y entonces la probabilidad referida puede ser aproximada por

$$\frac{1}{2} \frac{1}{m} \sum_{i=1}^m \frac{Y_i^{-2}}{\pi(1+Y_i^{-2})}, \quad Y_i \stackrel{iid}{\sim} U(0, 1/2).$$

```
set.seed(123); m=10000; y=runif(m,0,1/2)
I.hat=(1/2)*mean((y**-2)/(pi*(1+y**-2)))
I.hat
```


Integración de Monte Carlo

- La generalización para el caso en que la integral es la esperanza de una función $g(x)$, donde X tiene función densidad $f(x)$ (que no necesariamente es la uniforme) es trivial

$$I = \int g(x)f(x)dx = \mathbb{E}(g(x)).$$

- Así, si $X_i \stackrel{\text{iid}}{\sim} X$, entonces

$$\hat{I} = \frac{1}{m} \sum_{i=1}^m g(X_i).$$

Integración de Monte Carlo

Ejemplo

- *Vamos a volver al ejemplo anterior, donde $X \sim \text{Cauchy}(0, 1)$ y estamos interesados en la probabilidad $\Pr(X > 2)$.*
- *Ya sabemos que*

$$\Pr(X > 2) = \int_2^{\infty} \frac{1}{\pi(1+x^2)} dx = \int_{-\infty}^{\infty} I_A(x) \frac{1}{\pi(1+x^2)} dx, \quad \text{con } A = \{x : x > 2\}.$$

- *En este caso se puede considerar*

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad g(x) = I_A(x).$$

- *Así,*

$$I = \Pr(X > 2) = \mathbb{E}(g(X)) = \mathbb{E}(I_A(X)), \quad X \sim \text{Cauchy}(0, 1).33$$

Integración de Monte Carlo

Ejemplo

- *Tenemos entonces que*

$$\hat{\lambda} = \hat{\Pr}(X > 2) = \frac{1}{m} \sum_{i=1}^m I_A(X_i), \quad X_i \stackrel{iid}{\sim} \text{Cauchy}(0, 1), \quad i = 1, \dots, m.$$

```
m=10000; x=rcauchy(m,0,1)
I.hat=mean(ifelse(x>2,1,0))
I.hat
```

Integración de Monte Carlo

Ejemplo

- Suponga que queremos calcular $\Pr(X_1 < 1, X_2 < 1)$, donde el vector aleatorio (X_1, X_2) tiene distribución normal bivariada estándar con correlación 0.5, es decir

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2 \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \right).$$

- Note que

$$\Pr(X_1 < 1, X_2 < 1) = \int I_A(x_1, x_2) f(x_1, x_2) dx_1 dx_2,$$

donde f es la función de densidad conjunta de (X_1, X_2) (o sea, es la densidad de la distribución normal estándar bivariada) y $A = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 < 1 \& x_2 < 1\}$.

- Así, se simulamos pares (X_{1i}, X_{2i}) desde una distribución normal estándar bivariada, la probabilidad referida puede ser estimada como

$$\hat{\Pr}(X_1 < 1, X_2 < 1) = \frac{1}{m} \sum_{i=1}^m I_A(x_{1i}, x_{2i}),$$

o sea, tenemos simplemente que contar el número de puntos simulados que caen en A .

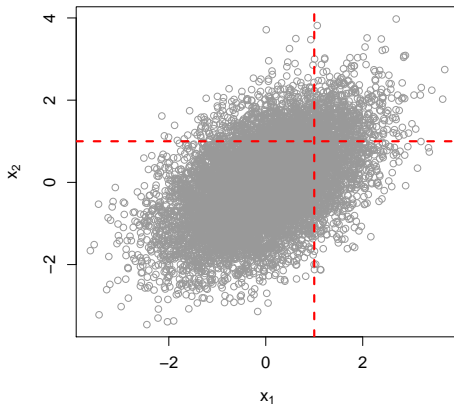
Integración de Monte Carlo

Ejemplo

```
require(pbivnorm) # para calcular el verdadero valor de la probabilidad
pbivnorm(1,1,0.5)
0.7452036

require(MASS)
set.seed(123); m=10000
sigma=matrix(c(1,0.5,0.5,1),byrow=T,nrow=2)
mu=c(0,0)
x=mvrnorm(m,mu,sigma)
phat=(1/m)*sum(x[,1]<1 & x[,2]<1)
0.7442
```

Integración de Monte Carlo



Integración de Monte Carlo

Ejemplo

- Use simulación para aproximar $\text{Cov}(U, e^U)$, donde $U \sim U(0, 1)$. Compare su aproximación con la respuesta exacta.
- Recuerde que

$$\begin{aligned}\text{Cov}(X, Y) &= \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y),\end{aligned}$$

con

$$\mu_X = \mathbb{E}(X), \quad \mu_Y = \mathbb{E}(Y).$$

- Así,

$$\text{Cov}(U, e^U) = \mathbb{E}(Ue^U) - \mathbb{E}(U)\mathbb{E}(e^U)$$

- Tenemos que

$$\mathbb{E}(Ue^U) = \int_0^1 ue^u f(u) du = \int_0^1 ue^u du = [ue^u]_0^1 - \int_0^1 e^u du = 1$$

Integración de Monte Carlo

Ejemplo

- Por otra parte, tenemos que $\mathbb{E}(U) = 1/2$, mientras que

$$\mathbb{E}(e^U) = \int_0^1 e^u du = e - 1.$$

- Así,

$$\text{Cov}(U, e^U) = 1 - \frac{1}{2}(e - 1) = \frac{1}{2}(3 - e) \approx 0.1409.$$

- Usando simulación, simplemente hacemos

$$\text{Cov}(U, e^U) \approx \frac{1}{m} \sum_{i=1}^m U_i e^{U_i} - \left(\frac{1}{m} \sum_{i=1}^m U_i \right) \left(\sum_{i=1}^m e^{U_i} \right), \quad U_i \stackrel{iid}{\sim} U(0, 1).$$

```
set.seed(123); m=10000; u=runif(m)
covap=mean(u*exp(u))-mean(u)*mean(exp(u))
covap
0.1386525
```


Integración de Monte Carlo

Ejemplo

- *Emplear simulación para aproximar la siguiente integral*

$$I = \int_0^1 \int_0^1 e^{x^2+y^2} dx dy.$$

- *Para aproximar la integral, simplemente tenemos que simular $U_1(i)U_2 \sim U(0,1)$ y hacer*

$$I \approx \hat{I} = \frac{1}{m} \sum_{i=1}^m e^{U_{1i}^2 + U_{2i}^2}$$

```
set.seed(123);m=100000; u1=runif(m); u2=runif(m)
```

```
Ihat=mean(exp((u1**2)+(u2**2)))
```

```
Ihat
```

```
2.136237
```

```
#comparar con el valor obtenido usando integración numerica  
(métodos de quadratura)
```

```
require(pracma)
```

```
f=function(x,y){exp((x**2)+(y**2))}
```

```
quad2d(f,0,1,0,1) 2.13935
```

Integración de Monte Carlo

- Consideremos

$$I = \int g(x)f(x)dx = \mathbb{E}(g(X)),$$

donde X es una variable aleatoria con función de densidad de probabilidad f .

- Hemos ya visto que si generamos $X_1, \dots, X_m \stackrel{\text{iid}}{\sim} X$, entonces el estimador de Monte Carlo de I está dado por

$$\hat{I}_m = \frac{1}{m} \sum_{i=1}^m g(X_i)$$

que converge para I cuando $m \rightarrow \infty$, por la ley fuerte de los grandes números.

- Hemos afirmado anteriormente que \hat{I}_m es un estimador no sesgado de I , es decir, que $\mathbb{E}(\hat{I}_m) = I$. Verifiquemos

$$\mathbb{E}(\hat{I}_m) = \mathbb{E}\left(\frac{1}{m} \sum_{i=1}^m g(X_i)\right) = \frac{1}{m} \sum_{i=1}^m \mathbb{E}(g(X_i)) = \mathbb{E}(g(X)) = I.$$

Integración de Monte Carlo

- Además, veamos cual es la varianza asociada al estimador de Monte Carlo

$$\text{Var}(\hat{I}_m) = \text{Var}\left(\frac{1}{m} \sum_{i=1}^m g(X_i)\right) = \frac{1}{m^2} \text{Var}\left(\sum_{i=1}^m g(X_i)\right) = \frac{1}{m} \text{Var}(g(X)).$$

- Se observa que la varianza del estimador disminuye cuando $m \rightarrow \infty$, por lo que aumentando el tamaño de la simulación irá mejorar la precisión del estimador.
- La varianza del estimador puede ser estimada reemplazando $\text{Var}(g(X))$ por la varianza muestral de g , que está dada por

$$\widehat{\text{Var}}(g(X)) = \frac{1}{m-1} \sum_{i=1}^m \{g(X_i) - \bar{g}(X)\}^2,$$

con

$$\bar{g}(X) = \frac{1}{m} \sum_{i=1}^m g(X_i).$$

Integración de Monte Carlo

- Así, la varianza de \hat{l}_m puede ser estimada por

$$\widehat{\text{Var}}(\hat{l}_m) = \frac{1}{m(m-1)} \sum_{i=1}^m \{g(X_i) - \bar{g}(X)\}^2.$$

- El correspondiente error estándar está dado por

$$\widehat{\text{se}}(\hat{l}_m) = \sqrt{\frac{1}{m(m-1)} \sum_{i=1}^m \{g(X_i) - \bar{g}(X)\}^2}$$

Integración de Monte Carlo

Ejemplo

- Usando solo variables aleatorias $U(0, 1)$, use un algoritmo de Monte Carlo para aproximar el valor de la función gamma

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$$

e indicar una estimación de la varianza del estimador.

- Podemos considerar $g(x) = x^{\alpha-1}$ y $f(x) = e^{-x}$. Así, $X \sim \text{Exp}(1)$.
- Entonces

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx \approx \frac{1}{m} \sum_{i=1}^m (-\log U_i)^{\alpha-1} = \hat{\Gamma}(\alpha), \quad U_i \stackrel{iid}{\sim} U(0, 1).$$

Integración de Monte Carlo

```
set.seed(123); m=10000; u=runif(m); alpha=0.75
g=(-log(u))**(alpha-1)
Ihat=mean(g)
1.21915
gamma(0.75)
1.225417

varIhat=var(g)/m
2.614636e-05
seIhat=sqrt(varIhat)
0.005113351
```

Técnicas de reducción de varianza

- Vamos a considerar varios enfoques para reducir la varianza del estimador \hat{I} de $I = \mathbb{E}(g(X))$.
- Si \hat{I}_1 y \hat{I}_2 son estimadores de I y $\text{Var}(\hat{I}_2) < \text{Var}(\hat{I}_1)$, entonces el porcentaje de reducción en la varianza logrado mediante el uso de \hat{I}_2 en lugar de \hat{I}_1 es

$$100 \times \left(\frac{\text{Var}(\hat{I}_1) - \text{Var}(\hat{I}_2)}{\text{Var}(\hat{I}_1)} \right).$$

Técnicas de reducción de varianza

Variables antitéticas

- Suponga como usualmente que queremos estimar $I = \mathbb{E}(g(X)) = \mathbb{E}(Y)$.
- Además suponga que hemos generado Y_1 y Y_2 idénticamente distribuidas con media I .
- Luego, el estimador

$$\hat{I}_A = \frac{Y_1 + Y_2}{2},$$

es un estimador no sesgado de I . Además

$$\text{Var}(\hat{I}_A) = \text{Var}\left(\frac{Y_1 + Y_2}{2}\right) = \frac{\text{Var}(Y_1) + \text{Var}(Y_2) + 2\text{Cov}(Y_1, Y_2)}{4}.$$

- Por lo tanto, sería ventajoso (en el sentido que la varianza se reduciría) si Y_1 y Y_2 en lugar de ser independientes, se correlacionaron negativamente, i.e., $\text{Cov}(Y_1, Y_2) < 0$.

Técnicas de reducción de varianza

Variables antitéticas

- Empezemos con el caso en que Y es una función de variables aleatorias uniformes y así $I = \mathbb{E}(g(U))$.
- El estimador de Monte Carlo simple en este caso es

$$\hat{I} = \frac{1}{m} \sum_{i=1}^m g(U_i) = \frac{1}{m} \sum_{i=1}^m Y_i, \quad U_i \stackrel{\text{iid}}{\sim} U(0, 1).$$

- En este estimador, podríamos también haber utilizado las variables $1 - U_i$, ya que si $U \sim U(0, 1)$, entonces también $1 - U \sim U(0, 1)$.
- Adicionalmente, U y $1 - U$ están perfectamente correlacionadas negativamente. Decimos que U y $1 - U$ son variables antitéticas.

Técnicas de reducción de varianza

Variables antitéticas

- Este argumento esta en la base de la construcción del estimador basado en variables antitéticas.
- Consideremos entonces $Y_i = g(U_i)$ como antes. Consideremos ahora también $\tilde{Y}_i = g(1 - U_i)$.
- Note que $\mathbb{E}(Y_i) = \mathbb{E}(\tilde{Y}_i) = I$, de manera que, en particular, si

$$Z_i = \frac{Y_i + \tilde{Y}_i}{2},$$

entonces $\mathbb{E}(Z_i) = I$.

Técnicas de reducción de varianza

Variables antitéticas

- El estimador de Monte Carlo basado en variables antitéticas es bastante simple de ser aplicado.
- Si queremos generar m replicas de Monte Carlo, entonces generamos $m/2$ replicas de $Y_i = g(U_i)$ y para las restantes $m/2$ replicas hacemos $\tilde{Y}_i = g(1 - U_i)$.
- **Atención:** Y_i y \tilde{Y}_i se basan en las mismas variables uniformes.
- Luego,

$$\begin{aligned}\text{Var}(\hat{l}_A) &= \frac{1}{m} \left\{ Y_1 + \tilde{Y}_1 + Y_2 + \tilde{Y}_2 + \dots + Y_{m/2} + \tilde{Y}_{m/2} \right\} \\ &= \frac{1}{m/2} \sum_{i=1}^{m/2} \left(\frac{Y_i + \tilde{Y}_i}{2} \right) \\ &= \frac{1}{m/2} \sum_{i=1}^{m/2} Z_i\end{aligned}$$

Técnicas de reducción de varianza

Variables antitéticas

- \hat{I}_A es un estimador no sesgado de I y por la ley fuerte de los grandes números sabemos que $\hat{I}_A \rightarrow I$ con probabilidad 1 cuando $m \rightarrow \infty$.
- Note que ambos los estimadores requieren aproximadamente el mismo esfuerzo computacional, así que la comparación de los dos estimadores será hecha en términos de cual estimador tiene menor varianza.
- Hemos ya visto que

$$\text{Var}(\hat{I}) = \text{Var}\left(\frac{1}{m} \sum_{i=1}^m Y_i\right) = \frac{\text{Var}(Y)}{m}.$$

Técnicas de reducción de varianza

Variables antitéticas

- Por otra parte, tenemos que

$$\begin{aligned}\text{Var}(\hat{I}_A) &= \text{Var}\left(\frac{\sum_{i=1}^{m/2} Z_i}{m/2}\right) \\&= \frac{\text{Var}(Z)}{m/2} \\&= \frac{\text{Var}(Y + \tilde{Y})}{2m} \\&= \frac{\text{Var}(Y) + \text{Var}(\tilde{Y}) + 2\text{Cov}(Y, \tilde{Y})}{2m} \\&= \frac{\text{Var}(Y)}{m} + \frac{\text{Cov}(Y, \tilde{Y})}{m} \\&= \text{Var}(\hat{I}) + \frac{\text{Cov}(Y, \tilde{Y})}{m}\end{aligned}$$

Técnicas de reducción de varianza

Variables antitéticas

- Por lo tanto,

$$\text{Var}(\hat{I}_A) < \text{Var}(\hat{I}),$$

si y solo si

$$\text{Cov}(Y, \tilde{Y}) < 0.$$

- Recordando que $Y = g(U)$ y $\tilde{Y} = g(1 - U)$, entonces

$$\text{Var}(\hat{I}_A) < \text{Var}(\hat{I}) \Leftrightarrow \text{Cov}(g(U), g(1 - U)) < 0.$$

- Averiguemos bajo que condiciones se tiene $\text{Cov}(g(U), g(1 - U)) < 0$.
- Supongamos que g es una función no decreciente en u en el intervalo $[0, 1]$. En este caso, si U es grande, $g(U)$ también tiende a ser grande; mientras que $1 - U$ y $g(1 - U)$ tienden a ser pequeños. Así, $\text{Cov}(g(U), g(1 - U)) < 0$.
- Podemos concluir de manera similar que si g es una función no creciente en u , entonces nuevamente $\text{Cov}(g(U), g(1 - U)) < 0$.
- Así, una condición suficiente para garantizar una reducción de varianza es que la función g sea monotona en el intervalo $[0, 1]$.

Técnicas de reducción de varianza

Variables antitéticas

Ejemplo

- Considere el problema de estimar

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

Compare el estimador de Monte Carlo usual con el estimador de Monte Carlo basado en variables antitéticas.

- Como $g = e^{u^2}$, con $U \sim U(0, 1)$, es una función monótona en u en el intervalo $[0, 1]$, entonces el estimador basado en las variables antitéticas tendrá menor varianza que el estimador de Monte Carlo usual.

Técnicas de reducción de varianza

Variables antitéticas

```
f=function(x)exp(x**2)
I=integrate(f,0,1)$value

# Estimador Monte Carlo simple
set.seed(123); m=10000; u=runif(m)
g=exp(u**2); Ihat=mean(g); varIhat=var(g)/m
Ihat
1.456122

# Estimador Monte Carlo variables antiteticas
set.seed(123); u1=runif(m/2); v1=1-u1
z=(exp(u1**2)+exp(v1**2))/2
IhatA=mean(z); varIhatA=var(z)/(m/2)
IhatA
1.462117

# porcentaje reducción varianza
100*((varIhat-varIhatA)/varIhat)
74.54129
```


Técnicas de reducción de varianza

Variables antitéticas

- Hasta ahora hemos considerado el caso donde $I = \mathbb{E}(g(U))$, con $U \sim U(0, 1)$.
- Por supuesto, en la práctica suele ser el caso que $I = \mathbb{E}(Y)$, con $Y = g(X)$.
- Todavía podemos usar el método de variables antitéticas para este tipo de problema si podemos usar el método de la transformación inversa para generar X .
- Sea F la función de distribución acumulada de X . Si $U \sim U(0, 1)$, por el método de la transformación inversa sabemos que

$$X \sim F^{-1}(U).$$

- Así,

$$Y_i = g(F^{-1}(U_i)), \quad \tilde{Y}_i = g(F^{-1}(1 - U_i)), \quad Z_i = \frac{Y_i + \tilde{Y}_i}{2}.$$

- Como F es monótona, F^{-1} también lo es, y entonces Y_i y \tilde{Y}_i se correlacionan negativamente si g es monótona.

Técnicas de reducción de varianza

Variables control

- Otra alternativa para reducir la varianza de un estimador de Monte Carlo de $I = \mathbb{E}(g(X))$ es el uso de variables control.
- Suponga que existe una función h tal que $\mu = \mathbb{E}(h(X))$ es conocido y que $h(X)$ está correlacionada con $g(X)$.
- Así, para cualquier constante $c \in \mathbb{R}$

$$\lambda_c = g(X) + c(h(X) - \mu),$$

es un estimador no sesgado de I .

Técnicas de reducción de varianza

Variables control

- La varianza

$$\begin{aligned}\text{Var}(\hat{l}_C) &= \text{Var}(g(X) + c(h(X) - \mu)) \\ &= \text{Var}(g(X) + ch(X) - c\mu) \\ &= \text{Var}(g(X) + ch(X)) \\ &= \text{Var}(g(X)) + c^2\text{Var}(h(X)) + 2c\text{Cov}(g(X), h(X)),\end{aligned}$$

es una función cuadrática en c y es minimizada para $c = c^*$, con

$$c^* = -\frac{\text{Cov}(g(X), h(X))}{\text{Var}(h(X))},$$

y para este valor la varianza del estimador es

$$\text{Var}(\hat{l}_A) = \text{Var}(g(X)) - \frac{\text{Cov}^2(g(X), h(X))}{\text{Var}(h(X))}.$$

Técnicas de reducción de varianza

Variables control

- Luego, si comprueba que desde que $\text{Cov}(g(X), h(X)) \neq 0$ se logra una reducción de varianza.
- La variable aleatoria $h(X)$ se llama variable de control para $g(X)$.
- Para calcular la constante c^* necesitamos de calcular $\text{Cov}(g(X), h(X))$ y $\text{Var}(h(X))$, lo que dependiendo del contexto puede no ser sencillo.
- Sin embargo, podemos estimar estas cantidades haciendo una simulación de Monte Carlo piloto/previa.
- Tenemos

$$\widehat{\text{Cov}}(g(X), h(X)) = \frac{\sum_{i=1}^p (g(X_i) - \bar{g}(X))(h(X_i) - \bar{h}(X))}{p - 1},$$
$$\widehat{\text{Var}}(h(X)) = \frac{\sum_{i=1}^p (h(X_i) - \bar{h}(X))^2}{p - 1},$$

donde p es el número de simulaciones hechas en estudio de Monte Carlo piloto.

Técnicas de reducción de varianza

Variables controlo

- Entonces,

$$\hat{c}^* = -\frac{\widehat{\text{Cov}}(g(X), h(X))}{\widehat{\text{Var}}(h(X))}.$$

Técnicas de reducción de varianza

Variables control

Ejemplo

- Usar el método de variables controlo para calcular

$$I = \mathbb{E}(U) = \int_0^1 e^u du, \quad U \sim U(0, 1).$$

- Obviamente, por integración, $I = e - 1 = 1.718282$.
- La varianza del estimador de Monte Carlo simple está dada por

$$\text{Var}(\hat{I}) = \text{Var}\left(\frac{1}{m} \sum_{i=1}^m e^{U_i}\right) = \frac{1}{m} \text{Var}(e^U),$$

con

$$\text{Var}(e^U) = \int_0^1 e^{2u} du - \left(\int_0^1 e^u du\right)^2 = \frac{e^2 - 1}{2} - (e - 1)^2 \approx 0.242036$$

Técnicas de reducción de varianza

Variables control

Ejemplo

- En este ejemplo, una elección natural para la variable control es $U \sim U(0, 1)$, para la cual se tiene

$$\mathbb{E}(U) = 1/2;$$

$$\text{Var}(U) = 1/12;$$

$$\text{Cov}(e^U, U) = \frac{1}{2}(3 - e) \approx 0.1408.$$

- Así,

$$c^* = -\frac{\text{Cov}(e^U, U)}{\text{Var}(U)} = -6(3 - e) \approx -1.6903.$$

- El estimador basado en la variable control está entonces dado por

$$\hat{I}_{c^*} = e^U - 1.6903(U - 0.5),$$

y, para m replicas, su varianza es $\text{Var}(\hat{I}_{A, c^*})/m$.

Técnicas de reducción de varianza

Variables control

Ejemplo

- Sabemos que

$$\text{Var}(\hat{I}_{A,c^*}) = \text{Var}(e^U) - \frac{\text{Cov}^2(e^U, U)}{\text{Var}(U)} \approx 0.003940$$

- El porcentaje de reducción de varianza es

$$100 \times \left(\frac{0.242036 - 0.003940}{0.242036} \right) \approx 98.4\%.$$

Técnicas de reducción de varianza

Variables control

```
# codigo R para repetir lo que hemos hecho anaíticamente

set.seed(123); m=10000; u=runif(m)
g=exp(u); Ihat=mean(g); varIhat=var(g)/m

gc=exp(u)-1.6903*(u-0.5); Ihatc=mean(gc); varIhatc=var(gc)/m

100*(varIhat-varIhatc)/varIhat
98.36

# Si no fuera posible calcular el valor optimo de c analíticamente
tendríamos que hacer un estudio de Monte Carlo piloto.

p=1000; u1=runif(u1); cast=-(cov(u1,exp(u1)))/var(u)
cast
-1.707979
```

Técnicas de reducción de varianza

Muestreo por importancia

- Consideremos, una vez más, que estamos interesados en la estimación de la cantidad

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

donde f es una función de densidad de probabilidad.

- El método de Monte Carlo usual para estimar I consiste en simular X_1, \dots, X_m desde la densidad f y hacer

$$\hat{I} = \frac{1}{m} \sum_{i=1}^m g(X_i).$$

- En este proceso, la distribución desde la cual estamos a simular, f , no tiene necesariamente que estar relacionada con g .
- Es posible que muchas de las variables simuladas sean de áreas donde $g(x)$ es cerca a cero o mismo cero.

Técnicas de reducción de varianza

Muestreo por importancia

- La idea del método de muestreo por importancia (*importance sampling*) es que sería preferible simular más variables en las áreas que son más importantes para la estimación de I .
- Simularíamos por tanto de una otra densidad, $\psi(x)$ que concentra más masa en áreas más importantes.
- Note que podemos re-escribir I de la siguiente manera

$$\begin{aligned} I &= \int g(x)f(x)dx \\ &= \int g(x)\frac{f(x)}{\psi(x)}\psi(x)dx \\ &= \int g(x)\omega(x)\psi(x)dx \\ &= \mathbb{E}(g(X)\omega(X)), \quad X \sim \psi. \end{aligned}$$

Técnicas de reducción de varianza

Muestreo por importancia

- La función de importancia $\psi(x)$ no tiene que ser positiva en todas las partes, es suficiente tener $\psi(x) > 0$, siempre que $g(x)f(x) \neq 0$.
- La función $\omega(X)$ es llamada función peso importancia (*importance weighting function*).
- El estimador de Monte Carlo basado en el método de muestreo por importancia está así dado por

$$\hat{l}_{IS} = \frac{1}{m} \sum_{i=1}^m g(X_i) \omega(X_i), \quad X_1, \dots, X_m \stackrel{\text{iid}}{\sim} \psi.$$

- Se tiene que $\mathbb{E}(\hat{l}_{IS})$, lo que implica que \hat{l}_{IS} es un estimador no sesgado de I .
- La varianza del estimador es

$$\begin{aligned} \text{Var}(\hat{l}_{IS}) &= \text{Var} \left(\frac{1}{m} \sum_{i=1}^m g(X_i) \omega(X_i) \right) \\ &= \frac{1}{m} \text{Var}(g(X) \omega(X)) \\ &= \frac{1}{m} \left\{ \int g^2(x) \frac{f^2(x)}{\psi(x)} dx - I^2 \right\}. \end{aligned}$$

Técnicas de reducción de varianza

Muestreo por importancia

- La elección de la función de importancia $\psi(x)$ se basa esencialmente en los siguientes criterios
 - Debe ser facil simular desde ψ .
 - $\text{Var}(\hat{I}_{IS}) < \text{Var}(\hat{I})$. Idealmente, nos gustaría que la varianza del estimador fuera la menor posible.
- Se prueba que la densidad $\psi(x)$ que minimiza $\text{Var}(\hat{I}_{IS})$ es proporcional a $|g(x)|f(x)$.
- Desafortunadamente, la función de importancia optima raramente está disponible.
- Sin embargo, debemos buscar densidades ψ cuya la forma sea cercana a la forma de $|g(x)|f(x)$ y que tenga el mismo soporte de f .

Técnicas de reducción de varianza

Muestreo por importancia

Ejemplo

- *Suponga que deseamos estimar*

$$I = \int_0^1 \frac{e^{-x}}{1+x^2} dx.,$$

usando el método de muestreo por importancia.

- *Las funciones candidatas a funciones de importancia son*

$$\psi_1(x) = 1, \quad 0 < x < 1,$$

$$\psi_2(x) = e^{-x}, \quad 0 < x < \infty,$$

$$\psi_3(x) = (1+x^2)^{-1}/\pi, \quad -\infty < x < \infty,$$

$$\psi_4(x) = e^{-x}/(1-e^{-1}), \quad 0 < x < 1,$$

$$\psi_5(x) = 4(1+x^2)^{-1}/\pi, \quad 0 < x < 1.$$

Técnicas de reducción de varianza

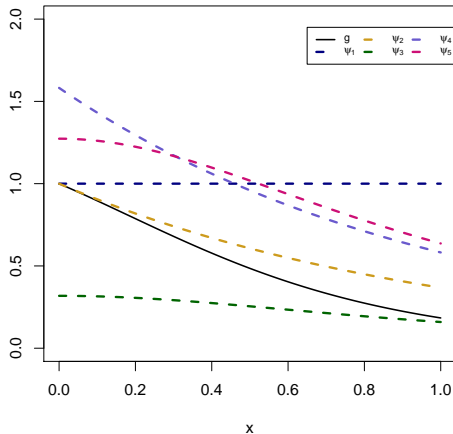
Muestreo por importancia

Ejemplo

- *Todas las 5 funciones son positivas en el intervalo $(0, 1)$.*
- *Sin embargo, ψ_2 y ψ_3 tienen soporte $(0, \infty)$ y $(-\infty, \infty)$ respectivamente, por lo que muchos de los valores simulados van a contribuir con ceros para la estimación, lo que claramente es ineficiente.*
- *Es fácil de muestrear desde todas estas distribuciones; en particular, ψ_3 es la función Cauchy estándar. Para simular desde ψ_4 y ψ_5 se puede usar el método de la transformación inversa.*

Técnicas de reducción de varianza

Muestreo por importancia



Técnicas de reducción de varianza

Muestreo por importancia

Ejemplo

- Las funciones ψ_4 y ψ_5 en la gráfica anterior tienen una forma similar a g y su soporte es el intervalo $(0, 1)$, por lo que esperamos que estas funciones de importancia logren una menor varianza para el estimador.
- Aún que la función ψ_2 tenga una forma muy similar a g , como su soporte es el intervalo $(0, \infty)$ esperamos, a priori, que la varianza asociada al estimador con esta función de importancia sea superior a la varianza usando ψ_4 o ψ_5 como funciones de importancia.
- El mismo es válido para la función ψ_3 , se bien que en este caso su forma no es tan cercana quanto ψ_2 y su soporte es $(-\infty, \infty)$, por lo que aún esperamos que la varianza en este caso sea mayor.

Técnicas de reducción de varianza

Muestreo por importancia

```
g=function(x) (exp(-x)/(1+x**2))
I=integrate(g, lower=0, upper=1)$value
0.5247971

set.seed(123); m=10000; u=runif(m)
g=function(x) { (exp(-x)/(1+x**2)) * (x>0) * (x<1) }
f=function(x) { x=ifelse(x>0 & x<1, 1, 0) }
phi1=function(x) { x=ifelse(x>0 & x<1, 1, 0) }
int1=g(u)*f(u)/phi1(u); Iis1=mean(int1); seIis1=sqrt(var(int1)/m)

phi2=function(x) exp(-x)
set.seed(123); x=rexp(m, 1)
int2=f(x)*g(x)/phi2(x); Iis2=mean(int2); seIis2=sqrt(var(int2)/m)
```

Técnicas de reducción de varianza

Muestreo por importancia

```
phi3=function(x) ((1+x**2)**(-1))/pi
set.seed(123); x=rcauchy(m)
#para evitar problemas numericos
ind=c(which(x>1), which(x<0)); x[ind]=2
int3=f(x)*g(x)/phi3(x); Iis3=mean(int3); seIis3=sqrt(var(int3)/m)

phi4=function(x) {(exp(-x)/(1-exp(-1)))*(x>0)*(x<1)}
set.seed(123); u=runif(m); x=-log(1-u*(1-exp(-1))) # transformacion
inversa
int4=f(x)*g(x)/phi4(x); Iis4=mean(int4); seIis4=sqrt(var(int4)/m)

phi5=function(x) {(4*((1+x**2)**(-1))/pi)*(x>0)*(x<1)}
set.seed(123); u=runif(m); x=tan(pi*u/4) # transformacion inversa
int5=f(x)*g(x)/phi5(x); Iis5=mean(int5); seIis5=sqrt(var(int5)/m)
```

Técnicas de reducción de varianza

Muestreo por importancia

Ejemplo

	\hat{I}_{IS, ψ_1}	\hat{I}_{IS, ψ_2}	\hat{I}_{IS, ψ_3}	\hat{I}_{IS, ψ_4}	\hat{I}_{IS, ψ_5}
<i>estimación</i>	0.5263	0.5236	0.5214	0.5259	0.5259
<i>se</i>	0.00244	0.0042	0.0095	0.00096	0.0014

- Como se observa, las funciones de importancia que logran una menor varianza son ψ_4 y ψ_5 .