Tarea 8

Marco Antonio Andrade Barrera

5 de abril de 2018

Ejemplo. Sea $X \sim Ca(x|0,1)$ y se quiere evaluear $P[X \geq 2]$.

Analíticamente se tiene

$$\theta = P[X \geq 2] = 1 - \frac{1}{\pi}arctan2 - 1/2 \approx 0.1475$$

Para cada uno de los estimadores (con 5,000 repeticiones),

- Generar una muestra de tamaño 1,000 y verficar la relación entre varianzas.
- Comprobar que si se generan n_i (n equivalente) entonces se obtiene aproximadamente la misma varianza que $\hat{\theta}_1$

1. Montecarlo simple

Usando $X_1, \ldots, X_{m_1} \sim Ca(x|0,1)$.

$$\hat{\theta}_1 = \frac{1}{m_1} \sum_{j=1}^{m_1} I_{[2,\infty)}(X_j)$$

$$V[\hat{\theta}_1] = \frac{\theta(1-\theta)}{m_1} \approx \frac{0.125}{m_1}$$

```
set.seed(32)
r <- 5000 #repeticiones
m1 <- 1000
thetas1 <- replicate(r,expr = {
  X <- rcauchy(m1)#generar m1 aleatorios Cauchy(x|0,1)
  theta1 <- (1/m1) * sum(2<=X)
  theta1
  })
```

```
#E[theta1]
mean(thetas1)
```

```
## [1] 0.1476234
```

```
#Varianza empírica de theta1
var(thetas1)
```

```
## [1] 0.0001220428
```

```
#Varianza real (aprox.)
0.125/m1
```

```
## [1] 0.000125
```

2. Montecarlo simple + propiedades de la Cauchy

$\theta = 0.5 \cdot \int\limits_{(-2,2)^c} \frac{dx}{\pi(1+x^2)}$ Entonces usando $X_1, \ldots, X_{m_2} \sim Ca(x|0,1)$,

$$\hat{\theta}_2 = \frac{1}{2m_2} \sum_{1}^{m_2} I_{(-2,2)^c}(X_j)$$

$V[\hat{\theta}_2] = 0.052/m_2$ y para obtener aproximadamente la misma varianza que $\hat{\theta}_1$ bastará $m_2 = 0.4134 m_1$.

```
set.seed(60)
m2 <- 1000
thetas2 <- replicate(r,expr = {
  X <- rcauchy(m2)#generar m2 aleatorios Cauchy(x|0,1)
  theta2 <- (1/2/m2) * sum(X<=-2 | 2<=X)
  theta2
  })
```

```
#E[theta2]
mean(thetas2)
```

```
## [1] 0.147626
```

```
#Varianza empírica de theta2
var(thetas2)
```

```
## [1] 5.264485e-05
```

```
#Varianza real (aprox.)
0.052/m2
```

```
## [1] 5.2e-05
```

Veamos que usando $m_2 = 0.4134 m_1 \approx 413$ se obtiene aproximadamente la misma varianza de $\hat{\theta}_1$.

```
set.seed(82)
m2eq <- 413
thetas2eq <- replicate(r,expr = {
  X <- rcauchy(m2eq)#generar m2eq aleatorios Cauchy(x|0,1)
  theta2eq <- (1/2/m2eq) * sum(X<=-2 | 2<=X)
  theta2eq
  })
```

```
#E[theta2eq]
mean(thetas2eq)
```

```
## [1] 0.1473424
```

```
#Varianza empírica de theta2eq
var(thetas2eq)
```

```
## [1] 0.0001210779
```

```
#Varianza real de theta1
0.125/m1
```

```
## [1] 0.000125
```

Note que efectivamente se obtiene una varianza muy cercana (apenas 4 millonésimas de diferencia) usando sólo 413 en $\hat{\theta}_2$ comparado con 1000 que se usan en $\hat{\theta}_1$.

3. Montecarlo simple + propiedades de la Cauchy

$\theta = 1/2 - 2/\pi \int\limits_0^2 \frac{dx}{2(1+x^2)} = 1/2 - 2/\pi E[\frac{1}{1+x_i^2}|X \sim U(u|0,2)]$. Entonces, usando $X_1, \ldots, X_{m_3} \sim U(x|0,2)$ tenemos que $\hat{\theta}_3 = 1/2 - \frac{2}{m_3\pi} \sum \frac{1}{1+x_j^2}$, $V[\hat{\theta}_3] = 0.0285/m_3$

y se obtiene una varianza equivalente a $\hat{\theta}_1$ con $m_3 = 0.22m_1$.

```
set.seed(107)
m3 <- 1000
thetas3 <- replicate(r,expr = {
  X <- runif(m3,0,2)#generar m3 aleatorios U(x|0,2)
  theta3 <- 1/2 - (2/pi/m3) * sum(1/(1+X^2))
  theta3
  })
```

```
#E[theta3]
mean(thetas3)
```

```
## [1] 0.1476244
```

```
#Varianza empírica de theta3
var(thetas3)
```

```
## [1] 2.818836e-05
```

```
#Varianza real (aprox.)
0.0285/m3
```

```
## [1] 2.85e-05
```

Note que el valor esperado de $\hat{\theta}_3$ es cercano al valor teórico, apenas una diezmilémisa de diferencia. También la varianza empírica dierefere en 1 millonésima de la real.

Veamos ahora que efectivamente haciendo $m_3 = 0.22m_1 = 220$ se obtiene prácticamente la misma varianza que en $\hat{\theta}_1$ usando 1000 observaciones.

```
set.seed(130)
m3eq <- 220
thetas3eq <- replicate(r,expr = {
  X <- runif(m3eq,0,2)#generar m3eq aleatorios U(x|0,2)
  theta3eq <- 1/2 - (2/pi/m3eq) * sum(1/(1+X^2))
  theta3eq
  })
```

```
#E[theta3eq]
mean(thetas3eq)
```

```
## [1] 0.1476511
```

```
#Varianza empírica de theta3eq
var(thetas3eq)
```

```
## [1] 0.0001315099
```

```
#Varianza real de theta1
0.125/m1
```

```
## [1] 0.000125
```

Efectivamente se obtiene una varianza semejante, con una diferencia de una cienmilésima.

4. Usando cambio de variable

$\theta = \frac{1}{2\pi} \int_0^{0.5} \frac{2dy}{1+y^2}$. Entonces usando $X_j \sim U(u|0, 1/2)$ tenemos

$$\hat{\theta}_4 = \frac{1}{2\pi m_4} \sum \frac{1}{1 + X_j^2}$$

y $V[\hat{\theta}_4] \approx 9.5 \times 10^{-5}/m_4$ y se obtiene una varianza equivalente a la de θ_1 usando $m_4 = 0.00076 \cdot m_1$.

```
set.seed(155)
m4 <- 1000
thetas4 <- replicate(r,expr = {
  X <- runif(m4,0,1/2)#generar m4 aleatorios U(x|0,1/2)
  theta4 <- 1/2/pi/m4 * sum(1/(1+X^2))
  theta4
  })
```

```
#E[theta4]
mean(thetas4)
```

```
## [1] 0.1475864
```

```
#Varianza empírica de theta4
var(thetas4)
```

```
## [1] 9.357839e-08
```

```
#Varianza real (aprox.)
9.5*10^(-5)/m4
```

```
## [1] 9.5e-08
```

Nuevamente obtenemos un valor muy cercano al valor teórico de θ y la varianza empírica y teórica prácticamente coinciden con una diferencia del orden de 10^{-9}. Comprobemos ahora que se obtiene una varianza equivalente a la de $\hat{\theta}_1$ usando sólo $m_4 = 0.00076 \cdot m_1 \approx 1$ observación.

```
set.seed(175)
m4eq <- 1
thetas4eq <- replicate(r,expr = {
  X <- runif(m4eq,0,1/2)#generar m4eq aleatorios U(x|0,1/2)
  theta4eq <- 1/2/pi/m4eq * sum(1/(1+X^2))
  theta4eq
  })
```

```
#E[theta4eq]
mean(thetas4eq)
```

```
## [1] 0.1477541
```

```
#Varianza empírica de theta4eq
var(thetas4eq)
```

```
## [1] 9.367995e-05
```

```
#Varianza real de theta1
0.125/m1
```

```
## [1] 0.000125
```

Note que en este caso, la diferencia entre las varianzas es 3 diezmilésimas (siendo la de $\hat{\theta}_4$ menor). Sorprendentemente este estimador sólo requiere una observación comparado con los 1000 que requiere $\hat{\theta}_1$.

5. Muestreo por importancia

Usando $X_j = \frac{2}{u} \sim g$, donde $u \sim U(u|0,1)$ tenemos el estimador

$$\hat{\theta}_5 = \frac{1}{2\pi m_5} \sum \frac{x_i^2}{1+x_i^2}$$

y $V[\hat{\theta}_5] = 0.000095/m_5$ y $m_5 = 0.00075 \cdot m_1$ para obtener una varianza equivalente.

```
set.seed(199)
m5 <- 1000
thetas5 <- replicate(r,expr = {
  X <- 2/runif(m5,0,1)
  theta5 <- 1/2/pi/m5 * sum(X^2/(1+X^2))
  theta5
  })
```

```
#E[theta5]
mean(thetas5)
```

```
## [1] 0.1475856
```

```
#Varianza empírica de theta5
var(thetas5)
```

```
## [1] 9.517094e-08
```

```
#Varianza real (aprox.)
0.000095/m5
```

```
## [1] 9.5e-08
```

Tanto el valor esperado de $\hat{\theta}_5$ como la varianza coinciden (aprox.) con los valores teóricos. Acontinuación se comprueba que para obtener una varianza equivalente a la de $\hat{\theta}_1$ es suficiente tomar $m_5 = 0.00075 \cdot m_1 \approx 1$.

```
set.seed(220)
m5eq <- 1
thetas5eq <- replicate(r,expr = {
  X <- 2/runif(m5eq,0,1)
  theta5eq <- 1/2/pi/m5eq * sum(X^2/(1+X^2))
  theta5eq
  })
```

```
#E[theta5eq]
mean(thetas5eq)
```

```
## [1] 0.1475498
```

```
#Varianza empírica de theta5eq
var(thetas5eq)
```

```
## [1] 9.463178e-05
```

```
#Varianza real de theta1
0.125/m1
```

```
## [1] 0.000125
```

Note que con una sóla observación se obtiene un valor muy cercano al valor real de θ. Además, efectivamente las varianza son prácticamente equivalente (difieren en 3 diezmilésimas).

6. Variables correlacionadas negativamente

Usando $u_1, \ldots, u_n \sim U(u|0,2)$ tenemos el estimador

$$\hat{\theta}_6 = 1/2 - \frac{1}{\pi m_6} \sum \left(\frac{1}{1+x_j^2} + \frac{1}{1+(2-x_j)^2}\right)$$

y $V[\hat{\theta}_6] = 5.8 \times 10^{-4}/m_6$ y una varianza equivalente a la de $\hat{\theta}_1$ se obtiene con $m_6 = 0.004 \cdot m_1$.

```r
set.seed(245)
m6 <- 1000
thetas6 <- replicate(r,expr = {
  X <- runif(m6,0,2)
  theta6 <- 1/2 - 1/pi/m6 * sum( 1/(1+X^2) + 1/(1+(2-X)^2) )
  theta6
  })
```

```r
#E[theta6]
mean(thetas6)
```

```
## [1] 0.1475793
```

```r
#Varianza empírica de theta6
var(thetas6)
```

```
## [1] 5.754771e-07
```

```r
#Varianza real (aprox.)
5.8 * 10^(-4) / m6
```

```
## [1] 5.8e-07
```

Práctivamente el valor esperado de $\hat{\theta}_6$ es el valor teórico de θ. También, las varianzas empírica y real difieren en el orden de 10^{-8}, es decir, se podrían considerar iguales. Comprobemos ahora que basta tomar $m_6 = 0.004 \cdot m_1 = 4$ para obtener una varianza equivalente.

```r
set.seed(266)
m6eq <- 4
thetas6eq <- replicate(r,expr = {
  X <- runif(m6eq,0,2)
  theta6eq <- 1/2 - 1/pi/m6eq * sum( 1/(1+X^2) + 1/(1+(2-X)^2) )
  theta6eq
  })
```

```r
#E[theta6eq]
mean(thetas6eq)
```

```
## [1] 0.1476424
```

```r
#Varianza empírica de theta6eq
var(thetas6eq)
```

```
## [1] 0.0001432075
```

```r
#Varianza real de theta1
0.125/m1
```

```
## [1] 0.000125
```

Observe que con 4 observaciones se obtiene también una estimación muy cercana a $\theta = 0.1475$. También, la varianza es equivalente pues difieren apenas en 2 diezmilésimas.

7. Rao-Blackwell

Tomando $y_1, \ldots, y_{m_7} \sim Ga(y|1/2, 1/2)$ tenemos el estimador

$$\hat{\theta}_7 = 1 - 1/m_7 \sum \Phi(2y_j^{0.5})$$

y $V[\hat{\theta}_7] \approx 0.0241/m_7$ y la varianza equivalente se obtiene con $m_7 = 0.0241/0.125 \cdot m_1 = 0.1928 \cdot m_1$.

```r
set.seed(290)
m7 <- 1000
thetas7 <- replicate(r,expr = {
  X <- rgamma(m7,shape = 1/2,rate = 1/2)
  theta7 <- 1 - 1/m7 * sum(pnorm(2*X^0.5))
  theta7
  })
```

```r
#E[theta7]
mean(thetas7)
```

```
## [1] 0.1475377
```

```r
#Varianza empírica de theta7
var(thetas7)
```

```
## [1] 2.314573e-05
```

```r
#Varianza real (aprox.)
0.0241 / m7
```

```
## [1] 2.41e-05
```

Observe como se obtiene un estimador muy cercano al valor real de θ y la varianza empírica difiere en una millonésima de la teórica. Verifiquemos ahora que $m_7 = 0.1928 \cdot m_1 \approx 193$ genera una varianza equivalente a la de $\hat{\theta}_1$.

```r
set.seed(311)
m7eq <- 193
thetas7eq <- replicate(r,expr = {
  X <- rgamma(m7eq,shape = 1/2,rate = 1/2)
  theta7eq <- 1 - 1/m7eq * sum(pnorm(2*X^0.5))
  theta7eq
  })
```

```r
#E[theta7eq]
mean(thetas7eq)
```

```
## [1] 0.1475946
```

```r
#Varianza empírica de theta7eq
var(thetas7eq)
```

```
## [1] 0.0001174021
```

```r
#Varianza real de theta1
0.125/m1
```

```
## [1] 0.000125
```

Efectivamente, las varianzas son práctivamente las mismas (equivalentes), con una diferencia de una diezmilésima.