

Simulación

6. Técnicas de Reducción de Varianza

Jorge de la Vega Góngora

Departamento de Estadística,
Instituto Tecnológico Autónomo de México

Clase 11



Técnicas de Reducción de varianza

- El objeto de las técnicas de reducción de varianza en los métodos de Montecarlo es mejorar la velocidad y eficiencia estadística de un estudio de simulación.
- Un estudio de simulación que utiliza entradas aleatorias genera salidas aleatorias, y por lo tanto tienen variabilidad que es necesario comprender.
- Por otra parte un estudio puede consumir muchos recursos como grandes cantidades de números aleatorios, y múltiples replicaciones para obtener un número puede ser demasiado costoso, por lo que es importante tratar de minimizar su variabilidad para obtener mayor precisión.

- Usualmente en simulación estocástica se desea estimar $\theta = E(h(\mathbf{X}))$. Como hemos visto, el algoritmo de simulación estándar dice:
 - 1 Genera $\mathbf{X}_1, \dots, \mathbf{X}_n$
 - 2 Estima θ con $\hat{\theta}_n = \sum_{i=1}^n Y_i$ donde $Y_i = h(\mathbf{X}_i)$
- Un intervalo de confianza aproximado del $100(1 - \alpha) \%$ está dado por:

$$\left[\hat{\theta}_n - z_{1-\alpha/2} \frac{\hat{\sigma}_n}{\sqrt{n}}, \hat{\theta}_n + z_{1-\alpha/2} \frac{\hat{\sigma}_n}{\sqrt{n}} \right]$$

donde $\hat{\sigma}_n = \hat{\text{Var}}(\hat{\theta}_n) = \frac{\sum (Y_j - \bar{Y})^2}{n-1}$

- Una forma de medir la calidad de un estimador es con la *longitud media HW* del intervalo de confianza,

$$HW = z_{1-\alpha/2} \frac{\hat{\sigma}_n}{\sqrt{n}}$$

- Usualmente se desea que HW sea lo más pequeño posible, pero esto es a veces difícil de lograr.

- Una forma de reducir la longitud media puede ser a través de las **Técnicas de reducción de varianza**. Éstas incluyen:
 - Variadas antitéticas
 - Variadas de control
 - Condicionamiento
 - Números pseudo-aleatorios comunes
 - Importance sampling
 - Muestreo estratificado
- Hay muchas otras técnicas sofisticadas o generalizaciones de las que se mencionan aquí.
- En la mayoría de los casos, consideraremos la siguiente situación: X es una variable aleatoria *de salida*, es decir, es una variable relevante del estudio de simulación. En muchas situaciones lo que se desea es estimar $E(X) = \mu$, la media del proceso. Típicamente, entonces, este número es una integral.

Variadas antitéticas: Idea básica

- Este método trata de inducir correlación negativa entre series de números pseudo-aleatorios.
- La idea básica es generar pares de corridas de un modelo tal que una observación pequeña en la primera corrida tiende a compensarse por una observación grande en la otra corrida, para que se obtenga una correlación negativa.
- Por ejemplo, si $u_k \sim \mathcal{U}(0, 1)$ fue generado para obtener un parámetro particular de la primera corrida (por ejemplo, un tiempo de espera) entonces $1 - u_k$ se usa para obtener el mismo parámetro en la segunda corrida. En este caso se dice que u_k y $1 - u_k$ están *sincronizados* en el sentido de que fueron utilizados para el mismo propósito (en este caso generar el mismo parámetro del modelo).

- Supóngase que $(X_1^{(1)}, X_2^{(1)}, \dots, X_n^{(1)})$ y $(X_1^{(2)}, X_2^{(2)}, \dots, X_n^{(2)})$ son 2 corridas de una simulación, en donde $X_j^{(1)}$ se estimó con u y $X_j^{(2)}$ se estimó con $1 - u$. Noten que:
 - $E(X_j^{(1)}) = E(X_j^{(2)}) = \mu$
 - Se tienen pares independientes: $(X_{j_1}^{(1)}, X_{j_1}^{(2)}) \perp\!\!\!\perp (X_{j_2}^{(1)}, X_{j_2}^{(2)})$.
 - Definan $X_j = \frac{X_j^{(1)} + X_j^{(2)}}{2}$ y $\bar{X} = \sum_{j=1}^n X_j$.

Entonces:

- \bar{X} es un estimador insesgado de μ , y
- $\text{Var}(\bar{X}) = \text{Var}(X_1)/n = \frac{\text{Var}(X_j^{(1)}) + \text{Var}(X_j^{(2)}) + 2\text{Cov}(X_j^{(1)}, X_j^{(2)})}{4n}$
- De esta forma, si se induce correlación negativa entre $X_j^{(1)}$ y $X_j^{(2)}$, entonces se puede reducir la varianza del estimador \bar{X} .
- Sin embargo, *no siempre* se puede garantizar que se logre el objetivo, depende del modelo. A veces se puede elaborar un estudio piloto para medir la magnitud de la reducción.

Si queremos estimar $\theta = \int_a^b f(x)dx$ por el método de Montecarlo crudo, entonces

$$\hat{\theta} = \frac{(b-a)}{n} \sum_{i=1}^n f(x_i),$$

donde $x_i \sim \mathcal{U}(a, b)$. Si por cada x_i se usa su variable antitética $\tilde{x}_i = a + (b - x_i)$, entonces el estimador se convierte en

$$\hat{\theta} = \frac{(b-a)}{n} \sum_{i=1}^{n/2} (f(x_i) + f(\tilde{x}_i))$$

Como probamos antes, como la varianza de la suma es la suma de las covarianzas mas dos veces la covarianza y la covarianza es negativa para variables antitéticas, entonces se reduce la varianza de la suma.

Variadas de control: Idea básica I

- Supóngase que queremos estimar $\mu = E(X)$ donde X es una variable aleatoria de salida, como indicamos previamente.
- En lugar de estimar μ directamente, se considera la diferencia entre el problema de interés y un modelo analítico: Y es otra variable relacionada con μ , y que está correlacionada con X , pero además se conoce $\nu = E(Y)$. A Y se le llama **variada de control** para X .
- Sea $X_c = X - a(Y - \nu)$ una nueva variable. Entonces
 - $E(X_c) = E(X) = \mu$, por lo que X_c es un estimador insesgado de μ .
 - $\text{Var}(X_c) = \text{Var}(X - a(Y - \nu)) = \text{Var}(X) + a^2\text{Var}(Y) - 2a\text{Cov}(X, Y)$ Entonces:

$$\text{Var}(X_c) \leq \text{Var}(X) \text{ si } 2a\text{Cov}(X, Y) > a^2\text{Var}(Y).$$

La varianza mínima se alcanza si

$$a^* = \frac{\text{Cov}(X, Y)}{\text{Var}(Y)}$$

En este caso $\text{Var}(X_c) = (1 - \rho_{X,Y}^2)\text{Var}(Y)$.

- En la práctica, se puede conocer o no el valor de $\text{Var}(Y)$ y muy difícilmente conocemos $\text{Cov}(X, Y)$, por lo que es difícil conocer el valor de a .

- Una alternativa es estimar a a través de un estudio piloto (Lavenberg, Moeller y Welch, 1982):

$$\hat{a} = \frac{C_{\bar{X}, \bar{Y}}}{S_Y^2}$$

y entonces $\bar{X}_c^* = \bar{X} - \hat{a}(\bar{Y} - \nu)$. Noten que \bar{X}_c^* ya no es un estimador insesgado de μ . Para reducir el sesgo se puede utilizar, por ejemplo el *jackknife*.

- Se puede mostrar (ejercicio) que el método de variadas antitéticas es un caso particular del método de variadas de control.

Ejemplo (a)

Supongan que $X \sim \mathcal{N}(0, 1)$ y se requiere estimar $E(\frac{X^6}{1+X^2})$. Como

$$\frac{x^6}{1+x^2} = x^4 - x^2 + 1 - \frac{1}{1+x^2}$$

entonces podemos aproximar $\frac{x^6}{1+x^2}$ con $Y = g(x) = x^4 - x^2 + 1$. Para esta Y , $E(Y) = E(X^4) - E(X^2) + 1 = 3 - 1 + 1 = 3$, ya que en una normal estándar $E(X^4) = 3$ (curtosis). De este modo,

$$\theta = E(\frac{X^6}{1+X^2} - (X^4 - X^2 + 1)) + 3 = 3 - E(\frac{1}{1+X^2})$$

Así que podemos aplicar Montecarlo crudo sólo a la función $h(x) = \frac{1}{1+x^2}$ muestreando de una normal estándar.

Ejemplo (b)

Se desea estimar $\theta = E(e^{(U+W)^2})$ donde $U, W \sim \mathcal{U}(0, 1)$ y son independientes. Sea $X = e^{(U+W)^2}$. Elegimos una variable de control Y .

Una posible variable de control es $Y_1 = U + W$. Por la distribución de U y W , sabemos que $\nu_1 = E(Y_1) = 1$ y se puede ver que $\text{Cov}(X, Y_1) > 0$. Otra posibilidad es usar $Y_2 = (U + W)^2$, y $E(Y_2) = 7/6$.

El siguiente código muestra como se puede hacer un pequeño piloto en R.

Código R

```
#Primero hacemos un pequeño piloto
p <- 100;n <- 1000
u <- runif(p);w <- runif(p)
x <- exp((u+w)^2)
y <- (u+w)^2
covest <- cov(cbind(x,y))
a <- -covest[1,2]/covest[2,2]

# Ahora hacemos la simulación
u <- runif(n);w <- runif(n)
x <- exp((u+w)^2); y <- (u+w)^2
v <- x + a*(y-7/6)
estimadorusual <- c(mean(x),sd(x))
estimadorcont <- c(mean(v),sd(v))
CI <- c(mean(v)-1.96*sd(v)/sqrt(n),mean(v)+1.96*sd(v)/sqrt(n))
estimadorusual

[1] 4.745260 5.597591

estimadorcont

[1] 4.867967 2.949073

CI

[1] 4.685182 5.050753
```

Condicionamiento: idea básica

- En algunos modelos es posible reemplazar un estimado de alguna variable por su valor analítico exacto.
- Al remover esta fuente de variabilidad, se espera una salida más estable, aunque de nuevo, *no siempre hay garantía de que cumpla el objetivo*.
- Supongamos otra vez que X es una variable de salida de una simulación y $E(X) = \mu$ es lo que se quiere estimar. Si Z es otra variable aleatoria tal que se conoce analíticamente la esperanza condicional $E(X|Z = z)$, entonces podemos calcular indirectamente la esperanza de X a través de la esperanza condicional:

$$\mu = E(X) = E(E(X|Z)).$$

Además con las fórmulas iteradas, podemos establecer la siguiente relación:

$$\text{Var}(E(X|Z)) = \text{Var}(X) - E(\text{Var}(X|Z)) \leq \text{Var}(X).$$

- Entonces, conviene que Z tenga las siguientes propiedades:
 - Z puede ser generado de manera eficiente.
 - $E(X|Z = z)$ se puede calcular analíticamente
 - $E(\text{Var}(X|Z))$ tiene un valor grande.

Ejemplo

Supongan un modelo jerárquico de la siguiente forma:

$$\begin{aligned}X|W &\sim \mathcal{Be}(w, w^2 + 1) \\ W &\sim \mathcal{P}(10)\end{aligned}$$

El problema es encontrar $\theta = E(X)$.

Como sabemos que $E(X|W = w) = \frac{w}{w^2 + w + 1}$, entonces basta con muestras $W_1, W_2, \dots, W_n \sim \mathcal{P}(10)$ y construir

$$\tilde{\theta} = \frac{1}{n} \sum_{j=1}^n E(X|W = w_j) = \frac{1}{n} \sum_{j=1}^n \frac{w_j}{w_j^2 + w_j + 1}.$$

Noten que en este ejemplo, no tuvimos que generar ningún valor de X , sólo valores de una distribución Poisson conocida.

En este caso es muy ineficiente utilizar el método crudo de Montecarlo, porque obliga a muestras de una distribución difícil de calcular.

Ejemplo 2 I

- Un proyecto de construcción tiene una duración $X | (\mu, \sigma^2) \sim \mathcal{N}(\mu, \sigma^2)$ donde $\mu \sim \mathcal{N}(100, 16)$ y $\sigma^2 \sim \exp(1/4)$, con $\mu \perp \sigma^2$. La compañía que construye debe pagar 1000 USD por cada día (y prorratea por partes de día) que la duración del proyecto excede K días. ¿Cuál es el costo esperado del retraso?
- Si usamos simulación 'ingenua':

```
K <- 20 # Supongamos 20 días
n <- 1000 # valor arbitrario
sigma <- rexp(n, 1/4) # Genera exponencial
mu <- rnorm(n, mean=4, sd=1) # Genera normal
x <- rnorm(n, mean=mu, sd=sigma) # Genera normal con mu y sigma
C1 <- 1000*ifelse(x - K <= 0, 0, x - K)
c(mean(C1), sd(C1))

[1] 91.21771 1142.96786
```


Ejemplo 2 II

- Usando condicionamiento, no es necesario generar muestras de X : El estimador tendrá la forma: $C_2 = E_{\sigma,\mu}(E_{X|\sigma,\mu}[1000 \max\{X - K, 0\}])$

$$\begin{aligned} E_{X|\sigma,\mu}[1000 \max\{X - K, 0\}] &= 1000 \int_K^\infty (x - K) \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left[-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right] dx \\ &= 1000 \int_{(K-\mu)/\sigma}^\infty (\sigma\nu + \mu - K) \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\nu^2\right] d\nu \\ &= 1000 \left(-\frac{\sigma e^{-\nu^2/2}}{\sqrt{2\pi}} \right)_{(K-\mu)/\sigma}^\infty + 1000(\mu - K)\Phi\left(\frac{\mu - K}{\sigma}\right) \\ &= 1000 \left[\sigma\phi\left(\frac{K - \mu}{\sigma}\right) + (\mu - K)\Phi\left(\frac{\mu - K}{\sigma}\right) \right] \end{aligned}$$

Entonces la simulación queda como:

Ejemplo 2 III

```
C2 <- 1000*(sigma*dnorm((K-mu)/sigma) - (K-mu)*pnorm((mu-K)/sigma))
c(mean(C2),sd(C2))
[1] 107.4373 412.1331

#porcentaje de reducción de varianza:
100*(sd(C1)-sd(C2))/sd(C1)
[1] 63.94185
```

Muestreo de Importancia I

Mejorando la eficiencia de la estimación

Hay ciertos problemas en los que el método clásico MC no nos será de utilidad, por la precisión que se necesita, o bien porque es un evento raro; por ejemplo, si deseamos calcular una probabilidad en la cola de una distribución (usualmente se requiere calcular p -values).

El método de *muestreo de importancia* (importance sampling) sirve para ganar eficiencia y precisión en la estimación, haciendo un cambio en la medida de referencia de los datos para mejorar el estimador de la varianza. Esta es una de las técnicas de *reducción de varianza* que detallaremos más adelante.

Muestreo de Importancia II

Mejorando la eficiencia de la estimación

Importance sampling

$$\theta = E_f[h(X)] = \int_{\mathcal{X}} h(x)f(x) dx = \int h(x) \frac{f(x)}{g(x)} g(x) dx = E_g \left[\frac{h(X)f(X)}{g(X)} \right]$$

donde $\text{supp}(g) \supset \text{supp}(hf)$ (g es estrictamente positiva cuando hf es diferente de cero). El estimador IS de θ será

$$\hat{\theta}_{n,IS} = \frac{1}{n} \sum_{i=1}^n \frac{f(y_i)h(y_i)}{g(y_i)},$$

donde $y_1, \dots, y_n \sim g$

Entonces, el problema se cambia a estimar observaciones de una densidad g en lugar de la densidad f .

Problema : Estimar $\theta = P(X > 20)$ donde $X \sim \mathcal{N}(0, 1)$.

```
pnorm(20,lower.tail=F)
```

```
[1] 2.753624e-89
```

En este caso pueden comprobar que generando observaciones de la distribución normal estándar no funciona, ya que el evento es muy raro. Pero podemos escribir la integral trasladando el problema a la cola de la distribución de interés:

$$\begin{aligned}\theta = \mathbb{E}[I(X > 20)] &= \int_{-\infty}^{\infty} I(X > 20) \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\&= \int_{-\infty}^{\infty} I(X > 20) \frac{\frac{1}{\sqrt{2\pi}} e^{-x^2/2}}{\frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2}} \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2} dx \\&= \int_{-\infty}^{\infty} I(X > 20) e^{-\mu x + \mu^2/2} \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2} dx \\&= \mathbb{E}_{\mu} \left[I(X > 20) e^{-\mu x + \mu^2/2} \right]\end{aligned}$$

Ahora podemos muestrear de $\mathcal{N}(\mu, 1)$. Podemos escoger $\mu = 20$ para estar cerca del punto de interés.

Ejemplo IS III

```
#Usando importance sampling:
n <- 1e6
mu <- 20
y <- rnorm(n, mean = mu)
I <- rep(0, n)
I[which(y > 20)] <- 1
theta_is <- mean(I*exp(-mu*y+mu^2/2))
#intervalo de confianza: noten la precisión.
theta_is

[1] 2.77473e-89

theta_is + c(-1,1)*sd(I*exp(-mu*y+mu^2/2))/n

[1] 2.774717e-89 2.774744e-89
```

Varianza del estimador IS I

La varianza del estimador IS se puede obtener de la siguiente manera:

$$\begin{aligned}\text{Var}_g(\hat{\theta}_{n,IS}) &= \int (\hat{\theta}_{n,IS} - \mathbb{E}(\hat{\theta}_{n,IS}))^2 g(x) dx \\ &= \int \hat{\theta}_{n,IS}^2 g(x) dx - \theta^2 \\ &= \int \left(\frac{h(x)f(x)}{g(x)} \right)^2 g(x) dx - \theta^2 \\ &= \int \frac{h^2(x)f(x)}{g(x)} f(x) dx - \theta^2\end{aligned}$$

Recordando que la varianza del estimador usual de θ es:

$$\text{Var}_f(\hat{\theta}_n) = \int h^2(x)f(x) dx - \theta^2,$$

la reducción de varianza que se puede dar es:

$$\text{Var}_f(\hat{\theta}_n) - \text{Var}_g(\hat{\theta}_{n,IS}) = \int h^2(x) \left(1 - \frac{f(x)}{g(x)}\right) f(x) dx$$

Necesitamos que sea positiva. Esto se puede dar cuando:

- $\frac{f(x)}{g(x)} \geq 1$ cuando $h(x)f(x)$ es pequeño, o
- $\frac{f(x)}{g(x)} \leq 1$ cuando $h(x)f(x)$ es grande.

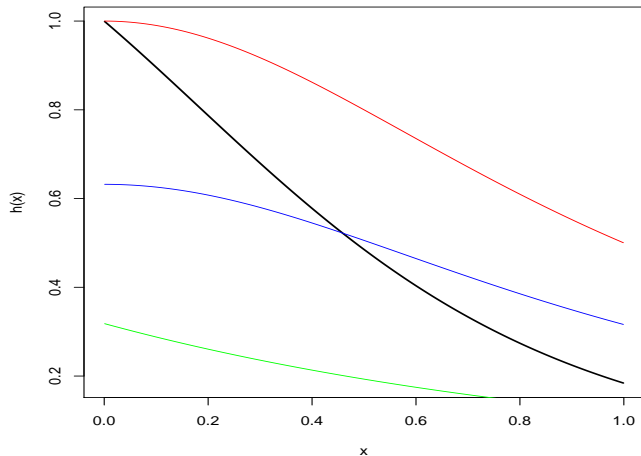
Más adelante regresaremos al problema de cómo se puede tratar de elegir g de manera adecuada.

- Consideremos $\int_0^1 \frac{e^{-x}}{1+x^2} dx$. Podemos usar como función de importancia varias opciones:
 - $g_0(x) = I(x)[0, 1]$
 - $g_1(x) = e^{-x}$
 - $g_2(x) = \frac{\pi}{1+x^2}$
 - $g_3(x) = e^{-x}/(1 + e^{-1})$

```
n <- 10000
theta.hat <- se <- numeric(4)
h <- function(x){exp(-x*log(1+x^2))*(x>0)*(x<1)}
g1 <- function(x)exp(-x)
g2 <- function(x)pi/(1+x^2)
g3 <- function(x)exp(-x)/(1-exp(-1))
```

```
curve(h(x),from=0.001, to=0.999, lwd=2)
curve(h(x)/g1(x),from=0.001,to=0.999,col="red",add=T)
curve(h(x)/g2(x),from=0.001,to=0.999,col="green",add=T)
curve(h(x)/g3(x),from=0.001,to=0.999,col="blue",add=T)
```

Ejemplo II



Ejemplo III

```
#Usando g0
x <- runif(n)
fg <- h(x)
theta.hat[1] <- mean(fg)
se[1] <- sd(fg)
```

```
#Usando g1
x <- rexp(n,1)
fg <- h(x)/exp(-x)
theta.hat[2] <- mean(fg)
se[2] <- sd(fg)
```

```
#Usando g2
x <- rcauchy(n)
i <- c(which(x>1),which(x<0))
x[i] <- 2 #para evitar los extremos fuera del intervalo
fg <- h(x)/dcauchy(x)
theta.hat[3] <- mean(fg)
se[3] <- sd(fg)
summary(fg)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.0000  0.0000  0.0000  0.5042  0.0000  3.1391
```

```
#usando g3
u <- runif(n)
x <- -log(1-u*(1-exp(-1)))
fg <- h(x)/(exp(-x)/(1-exp(-1)))
theta.hat[4] <- mean(fg)
se[4] <- sd(fg)
```

- Comparando todos los resultados:

```
W <- rbind(theta.hat,se)
colnames(W) <- paste0("g",0:3)
W
```

	g0	g1	g2	g3
theta.hat	0.5249991	0.5279436	0.5042298	0.52328794
se	0.2448457	0.4170262	0.9348325	0.09674531

Ejemplo 2 I

Supongan que se tiene una estadística de prueba $T(\mathbf{X})$ que rechaza una hipótesis H_0 si $T(\mathbf{X}) \geq c$. Lo que podría querer estudiarse en este contexto puede ser el nivel de confianza de la prueba

$$\theta = P_{H_0}(T(\mathbf{X}) \geq c) = \int I_D(\mathbf{x})f(\mathbf{x})d\mathbf{x}$$

donde $D = \{\mathbf{x} | T(\mathbf{x}) \geq c\}$.

- El método de Montecarlo dice: obtén $\mathbf{X}_1, \dots, \mathbf{X}_n$ una muestra aleatoria de la distribución conjunta F y calcula

$$\hat{\theta} = \frac{\#\{\mathbf{X}_i \in D\}}{n},$$

con $\text{Var}(\hat{\theta}) = \frac{\theta(1-\theta)}{n}$.

- El método de importance sampling dice: encuentra $g(\mathbf{x})$ una densidad tal que el cociente $\frac{f(\mathbf{x})}{g(\mathbf{x})}$ sea pequeño en D . Una vez encontrada esta g se puede mostrar de G :

$$\theta = \int I_D(\mathbf{x})f(\mathbf{x})d\mathbf{x} = \int I_D(\mathbf{x})f(\mathbf{x})/g(\mathbf{x})g(\mathbf{x})d\mathbf{x} = E\left(I_D(\mathbf{X})\frac{f(\mathbf{X})}{g(\mathbf{X})}\right)$$

- Si $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ es una muestra aleatoria de G ,

$$\tilde{\theta} = \frac{1}{n} \sum_{j=1}^n I_D(\mathbf{y}_j) \frac{f(\mathbf{y}_j)}{g(\mathbf{y}_j)}$$

tiene varianza $\text{Var}(\tilde{\theta}) = \frac{1}{n} \int_D \frac{f^2(\mathbf{y})}{g^2(\mathbf{y})} g(\mathbf{y}) d\mathbf{y} - \theta^2/n$ que será menor a $\text{Var}(\hat{\theta})$.

Métodos complementarios

- Este procedimiento se utiliza cuando se realiza un estudio y se desea comparar el desempeño de dos o más sistemas. Frecuentemente, el objetivo al comparar es estimar las diferencias en los parámetros de dos procesos estocásticos.
- Los dos parámetros pueden estar positivamente correlacionados, y en ese caso la varianza de las diferencias individuales posiblemente es menor que la varianza de la diferencia de los estimadores en conjunto. Si T y S son estimadores insesgados, queremos conocer la diferencia de las varianzas $\text{Var}(T) - \text{Var}(S)$, ya que el que tenga menor varianza será mejor.
- Supongan que cada estimador es una función de una muestra aleatoria x_1, \dots, x_n . Un estimador de montecarlo de la varianza de T para una muestra de tamaño n se obtiene generando m muestras de tamaño n de la distribución dada, se calcula T_i para la i -ésima muestra y luego se calcula:

$$\hat{\text{Var}}(T) = \frac{\sum_{i=1}^m (T_i - \bar{T})^2}{m - 1}$$

- Más que hacer esto separadamente para T y S , noten que como los estimadores son insesgados, $E(S) = E(T)$ y por lo tanto

$$\begin{aligned}\text{Var}(T) - \text{Var}(S) &= E(T^2) - E(T)^2 - (E(S^2) - E(S)^2) \\ &= E(T^2) - E(S^2) \\ &= E(T^2 - S^2)\end{aligned}$$

así que podemos obtener un estimado de $T^2 - S^2$, utilizando números aleatorios comunes.

- Consideren un sistema de linea de espera en donde los clientes llegan de acuerdo a un proceso Poisson $N(t)$. El operador requiere instalar un servidor para atender las llegadas que puede ser de dos tipos, M y N . Si elige M , entonces para el cliente i sea el tiempo de servicio S_i^m , X^m el tiempo de espera total en el sistema de todos los clientes que llegaron antes del tiempo T y W_i es el tiempo total en el sistema. Las variables se relacionan como:

$$X^m = \sum_{i=1}^{N(T)} W_i^m$$

- Noten que $W_i^m = S_i^m + Q_i^m$ donde Q_i^m es el tiempo de espera de i antes de ser atendido.
- Para el servido N , también se tienen las variables S_i^n , X^n , W_i^n y Q_i^n . El operador desea estimar $\theta = E(X^m) - E(X^n)$

Ejemplo

Los números aleatorios comunes se pueden usar para estimar la delta de una opción cuando no puede ser calculada de modo explícito, aunque hay mejores métodos para hacerlo. Aquí se menciona este como una forma de aplicación. Si $C(S)$ es el precio de una opción particular cuando el precio del bien subyacente es S , entonces se define a la delta como

$$\Delta = \frac{\partial C}{\partial S}$$

que mide la sensibilidad del precio de la opción a cambios en el precio S_t del bien subyacente. Entonces, si $\epsilon > 0$, se puede aproximar Δ con la razón de diferencia finita

$$\Delta_\epsilon = \frac{C(S + \epsilon) - C(S - \epsilon)}{2\epsilon}$$

En este caso, $C(S + \epsilon)$ y $C(S - \epsilon)$ no deben ser estimados de manera independiente, sino utilizando el mismo conjunto de números aleatorios comunes. Para una ϵ pequeña, la reducción de varianza puede ser dramática.

Muestreo estratificado I

- Este es un caso particular de *importance sampling*. Si suponemos que en el problema $\theta = \int_D f(x)dx$ $f(x)$ se puede descomponer como una mezcla de densidades, tal que $f(x) = \sum_{j=1}^k w_j f_j(x)$ con w_j pesos convexos, entonces si $\theta_j = \int_D f_j(x)dx$,

$$\theta = \sum_{j=1}^k w_j \theta_j.$$

- Se puede obtener una muestra de cada parte de manera separada, con n_j observaciones, lo que da $\hat{\theta}_j$ con $\text{Var}(\hat{\theta}_j) = \sigma_j^2/n_j$, donde σ_j es la varianza para el estimador $\hat{\theta}_j$. Combinando se obtiene

$$\text{Var}(\hat{\theta}) = \sum_{j=1}^k w_j^2 \frac{\sigma_j^2}{n_j}$$

- Se puede elegir n_j de manera óptima para $n = \sum_{j=1}^k n_j$, w_j y σ_j^2 fijas, obteniendo

$$n_j = \frac{w_j \sigma_j}{\sum_{j=1}^k w_j \sigma_j}$$

Ejemplo

Consideren obtener una muestra de una *normal contaminada*:

$$X \sim \begin{cases} \mathcal{N}(0, 1) & \text{con probabilidad } 1 - \alpha \\ \mathcal{N}(0, \sigma^2) & \text{con probabilidad } \alpha \end{cases}$$

Si α es pequeña entonces en una muestra de tamaño n se tendrán una o dos observaciones del componente contaminado. Si M es el número de casos contaminados en una muestra de tamaño n , entonces $M \sim \text{Bin}(n, \alpha)$.

Entonces podemos estratificar por la variable M en los casos $M = 0, 1, \dots, n$, con $w_j = \binom{n}{j} \alpha^j (1 - \alpha)^{n-j}$.

Observación: usualmente la varianza de un estimador se incrementa conforme la contaminación aumenta, así que la n_j óptima dependerá de j así como de w_j .

Sampling Importance Resampling (SIR) I

(Remuestreo de muestreo de importancia)

A partir de las ideas desarrolladas para el muestreo de importancia, se puede obtener un nuevo modelo para simular observaciones de f , como se indica a continuación.

- El método IS produce una muestra $X_1, \dots, X_n \sim g$ y un conjunto de pesos de 'importancia' $\omega_i = f(X_i)/g(X_i)$.
- Los pesos ω_i no suman 1, y algunos son mayores que uno. Pero se pueden normalizar:
$$w_i = \frac{\omega_i}{\sum_{i=1}^n \omega_i}$$
- Si se considera obtener muestras X^* con probabilidad w_i , extraídas **con reemplazo de** $\{X_1, \dots, X_n\}$ (esta fase se conoce como remuestreo), entonces la distribución de X_i^* es f , ya que:

$$\begin{aligned} P(X^* \in A) &= \sum_{i=1}^n P(X^* \in A \& X^* = X_i) \\ &= \int_A f(x)/g(x)g(x) dx = \int_A f(x) dx \end{aligned}$$

Sampling Importance Resampling (SIR) II

(Remuestreo de muestreo de importancia)

Estimador Sampling Importance Resampling

El estimador SIR está dado por:

$$\hat{\theta}_{n,SIR} = \sum_{i=1}^n w_i h(X_i)$$

donde $w_i = \frac{f(X_i)/g(X_i)}{\sum_{j=1}^n f(X_j)/g(X_j)}$

Ejemplo SIR I

Este método es muy general, así que podemos aplicarlo a cualquier función: Consideremos obtener la integral $\theta = E_f(X^2)$ donde $f(x) = e^{0.4(x-0.4)^2 - 0.08x^4}$. Notemos que f no es una densidad, por lo que necesitamos encontrar la constante para normalizar y poder comparar valores. Por otra parte, uso $g \sim \mathcal{U}(-4, 4)$ como cambio de medida.

```
f <- function(x, a=0.4,b=0.08){exp(a*(x-a)^2 -b*x^4)}
x <- seq(-4,4,0.1)
#Consideramos a g ~ U[-4,4]
n <- 10000
x <- runif(n,-4,4)
w <- f(x)/dunif(x,-4,4)
q <- w/sum(w)
xb <- sample(x,prob=q,replace = T) #remuestreo
thetahatSIR <- sum(q*xb^2)
cc <- integrate(function(x)f(x),-4,4)$val #constante de estandarización
integrate(function(x){(1/cc)*f(x)*x^2},-4,4)
```

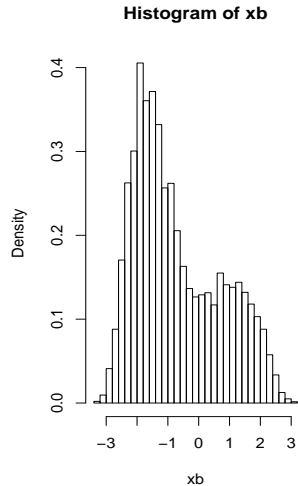
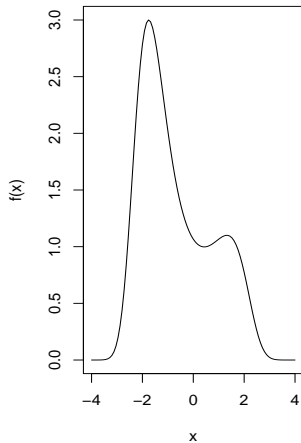
```
2.413271 with absolute error < 1.7e-09
```

```
thetahatSIR
```

```
[1] 2.406864
```

```
par(mfrow=c(1,2))
curve(f,type="l",from = -4,to=4)
hist(xb,breaks=30,prob=T)
```

Ejemplo SIR II



- $X \sim T(\nu, \theta, \sigma^2)$ con densidad

$$f(x) = \frac{\Gamma((\nu + 1)/2)}{\sigma \sqrt{\nu \pi} \Gamma(\nu/2)} \left(1 + \frac{(x - \theta)^2}{\nu \sigma^2} \right)^{-(\nu+1)/2}$$

- Tomar $\theta = 0$ y $\sigma = 1$
- Estimar

$$\int_{2.1}^{\infty} x^5 f(x) dx$$

- Con candidatos:
 - la misma f
 - Cauchy
 - Normal
 - Uniforme en $(0, 1/2.1)$