MAT-468: Sesión 10, Integración Monte Carlo

Felipe Osorio

http://fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Integración Monte Carlo

Problemas habituales en Estadística son el cálculo de:

$$
\begin{aligned}
f(\boldsymbol{x}) &= \int f(\boldsymbol{x}, \boldsymbol{y}) \, \mathrm{d}\boldsymbol{y}, \\
\mathsf{E}\{\boldsymbol{g}(\boldsymbol{x})\} &= \int \boldsymbol{g}(\boldsymbol{x}) f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \\
f(\boldsymbol{x}|\boldsymbol{y}) &= \frac{f(\boldsymbol{x}, \boldsymbol{y})}{f(\boldsymbol{y})} = \frac{f(\boldsymbol{y}|\boldsymbol{x}) f(\boldsymbol{x})}{\int f(\boldsymbol{y}|\boldsymbol{x}) f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}},
\end{aligned}
$$

todos estos pueden ser vistos como el problema de evaluar la integral

$$
\mathsf{E}_f\{\boldsymbol{h}(\boldsymbol{x})\} = \int \boldsymbol{h}(\boldsymbol{x}) f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}. \tag{1}
$$

Observación:

Ecuaciones como en (1) usualmente no pueden ser resueltas analíticamente y por tanto debemos recurrir a métodos de integración numérica.[1]

[1] Aunque estos métodos no son eficientes para dimensiones altas.

Integración Monte Carlo

Esta técnica está basada en la ley de los grandes números: Si $\{\boldsymbol{X}_j\}_{j=1}^{M}$ es una secuencia de vectores aleatorios iid con la misma distribución $\boldsymbol{X}$, entonces

$$\mathsf{E}\{\boldsymbol{h}(\boldsymbol{X})\} = \lim_{M\to\infty} \frac{1}{M} \sum_{j=1}^{M} \boldsymbol{h}(\boldsymbol{X}_j), \tag{2}$$

con probabilidad 1.

Note que la igualdad dada en (2) sólo se satisface en el límite $M \to \infty$, así que podemos usar una aproximación para M fijo.

Método Monte Carlo

Note que la técnica Monte Carlo, propone el estimador

$$\widehat{\boldsymbol{\theta}}_M = \frac{1}{M}\sum_{j=1}^{M} \boldsymbol{h}(\boldsymbol{x}_j), \qquad (3)$$

para el parámetro poblacional (desconocido)

$$\boldsymbol{\theta} = \mathsf{E}_f\{\boldsymbol{h}(\boldsymbol{X})\}$$

usando una muestra aleatoria, $\boldsymbol{x}_1, \dots, \boldsymbol{x}_M$ desde f.

De este modo, tenemos que $\widehat{\boldsymbol{\theta}}_M \xrightarrow{\text{a.s}} \boldsymbol{\theta}$.[2]

[2] Es decir, $\mathrm{P}(\lim_{M\to\infty} \widehat{\boldsymbol{\theta}}_M = \boldsymbol{\theta}) = 1$.

Ejemplo 1

Considere $h(x) = x$, de este modo la estimación Monte Carlo (3) reduce a

$$\mathsf{E}\{X\} \approx \frac{1}{M}\sum_{j=1}^{M} x_j = \overline{x}.$$

Ejemplo 2

Suponga que deseamos calcular $\mathsf{E}\{\sin(X)^2\}$, donde $X \sim \mathsf{N}(\mu, \sigma^2)$. [3] De este modo, considere $x_1, \dots, x_M$ una m.a.(M) desde $\mathsf{N}(\mu, \sigma^2)$, y por la ley de los grandes números, tenemos

$$\mathsf{E}\{\sin(X)^2\} \approx \frac{1}{M}\sum_{j=1}^{M} \sin(x_j)^2.$$

[3] Un valor exacto es muy difícil de obtener analíticamente.

Note además que $\mathsf{P}(\boldsymbol{X} \in A) = \mathsf{E}\{I_A(\boldsymbol{X})\}$, pues en efecto,

$$\mathsf{P}(\boldsymbol{X} \in A) = \int_A f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \int f(\boldsymbol{x}) I_A(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}.$$

De este modo, podemos usar

$$\mathsf{P}(\boldsymbol{X} \in A) = \mathsf{E}\{I_A(\boldsymbol{X})\} \approx \frac{1}{M}\sum_{j=1}^{M} I_A(\boldsymbol{x}_j),$$

para M suficientemente grande.

Ejemplo 3

Sea $X \sim \mathsf{N}(0,1)$ y $a \in \mathbb{R}$. Entonces $p = \mathsf{P}(X \leq a)$ $(= \Phi(a))$ no tiene forma explícita, y

$$\widehat{p}_M = \frac{1}{M}\sum_{j=1}^{M} I_{(-\infty,a]}(x_j),$$

puede ser usado como una aproximación.

Integración Monte Carlo: Ejemplos

Usualmente, también es de interés aproximar integrales de la forma,

$$\theta = \int_a^b h(x)\,\mathrm{d}x.$$

De este modo, considere $x_1, \dots, x_M$ una muestra aleatoria desde $\mathsf{U}(a,b)$, cada uno con densidad

$$f(x) = \frac{1}{b-a} I_{(a,b)}(x).$$

Podemos escribir

$$\begin{aligned}
\theta &= \int_a^b \frac{h(x)}{f(x)}\, f(x)\,\mathrm{d}x = \mathsf{E}_f\left\{\frac{h(x)}{f(x)}\right\} \\
&= (b-a)\,\mathsf{E}_f\{h(X)\} \approx \frac{(b-a)}{M}\sum_{j=1}^{M} h(x_j),
\end{aligned}$$

para M suficientemente grande.

Ejemplo 4

Considere la integral $\int_0^{2\pi} e^{\alpha\cos(x)}\,\mathrm{d}x$. Podemos generar datos $X_j \sim \mathsf{U}(0, 2\pi)$ y usar la aproximación

$$\int_0^{2\pi} \exp(\alpha\cos(x))\,\mathrm{d}x \approx \frac{1}{2\pi}\sum_{j=1}^{M} e^{\alpha\cos(x_j)}.$$

Ejercicio

Suponga que deseamos calcular

$$\int_0^{\infty} \log(x) x^{a-1} e^{-x}\,\mathrm{d}x$$

¿Qué $f(x)$ utilizaría usted?

Algoritmo: Integración Monte Carlo

Entradas: una función $\boldsymbol{h}$, $M \in \mathbb{N}$ y copias iid $\{\boldsymbol{x}_j\}_{j\in\mathbb{N}}$ de $\boldsymbol{X}$.

1. $s \leftarrow 0$
2. `para` $j = 1, 2, \dots, M$ $\{$

    generar $\boldsymbol{x}_j$ con la misma distribución que $\boldsymbol{X}$.

    $s \leftarrow s + \boldsymbol{h}(\boldsymbol{x}_j)$

   $\}$

4. retornar s/M.

Ignorando el costo de la asignación inicial y la división por M, tenemos que el tiempo de ejecución del algoritmo es proporcional a M.[4]

[4] Es decir, la complejidad es de orden $O(M)$

Proposición 1

La estimación Monte Carlo $\widehat{\theta}_M$ para $\mathsf{E}_f(h(\boldsymbol{X}))$, satisface:

$$\begin{aligned}\mathsf{bias}(\widehat{\theta}_M) &= 0,\\ \mathsf{MSE}(\widehat{\theta}_M) &= \mathsf{var}(\widehat{\theta}_M) = \frac{1}{M}\,\mathsf{var}(h(\boldsymbol{X})).\end{aligned}$$

En efecto, basta notar que

$$\mathsf{E}(\widehat{\theta}_M) = \mathsf{E}\left\{\frac{1}{M}\sum_{j=1}^M h(\boldsymbol{X}_j)\right\} = \frac{1}{M}\sum_{j=1}^M \mathsf{E}\{h(\boldsymbol{X}_j)\} = \mathsf{E}\{h(\boldsymbol{X})\},$$

y

$$\mathsf{var}(\widehat{\theta}_M) = \mathsf{var}\left\{\frac{1}{M}\sum_{j=1}^M h(\boldsymbol{X}_j)\right\} = \frac{1}{M^2}\sum_{j=1}^M \mathsf{var}\{h(\boldsymbol{X}_j)\} = \frac{1}{M}\,\mathsf{var}\{h(\boldsymbol{X})\}.$$

Error Monte Carlo

Para hacer comparaciones más simples, se define la raíz del error cuadrático medio como:

$$\mathsf{RMSE}(\widehat{\theta}_M) = \sqrt{\mathsf{MSE}(\widehat{\theta}_M)} = \frac{\mathsf{stdev}(h(\boldsymbol{X}))}{\sqrt{M}}$$

Definición 1

Sean dos funciones $g : \mathbb{N} \to \mathbb{R}$ y $h : \mathbb{N} \to \mathbb{R}$, se dice que g es de orden $O(h)$, o bien $g(N) = O(h(N))$, para $N \to \infty$ si existen constantes $N_0 \in \mathbb{N}$ y $c > 0$ tal que

$$|f(N)| \leq c|g(N)|, \qquad \text{para todo } N \geq N_0.$$

De este modo, para el caso de integración Monte Carlo tenemos

$$\mathsf{MSE}(\widehat{\theta}_M) = O\big(\tfrac{1}{M}\big), \qquad \mathsf{RMSE}(\widehat{\theta}_M) = O\big(\tfrac{1}{\sqrt{M}}\big)$$

Error Monte Carlo

Note además que

$$\mathsf{var}(\widehat{\theta}_M) = \frac{1}{M} \int (\widehat{\theta}_M - \mathsf{E}_f\{h(\boldsymbol{X})\})^2 f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x},$$

que también puede ser estimado desde la muestra $\boldsymbol{x}_1, \dots, \boldsymbol{x}_M$ como

$$v_M = \frac{1}{M^2} \sum_{j=1}^{M} \{h(\boldsymbol{x}_j) - \widehat{\theta}_M\}^2.$$

Bandas de confianza y construcción de test de convergencia pueden ser obtenidos notando que para M grande,

$$\frac{\widehat{\theta}_M - \mathsf{E}_f\{h(\boldsymbol{X})\}}{\sqrt{v_M}} \stackrel{\cdot}{\sim} \mathsf{N}(0, 1).$$

EX UMBRA IN SOLEM

Observaciones:

- Una de las ventajas de Monte Carlo es que podemos aproximar integrales altamente dimensionales de manera bastante sencilla.
- Es posible mejorar la aproximación de la integral mediante simular un número mayor de observaciones.

Área de un cuarto del círculo unitario

Sea $h(x) = \sqrt{1 - x^2}$ y sea $f(x)$ la densidad uniforme entre $(0, 1)$. Se desea calcular $\frac{1}{4}$ del área del circulo unitario ($\pi/4 \approx 0.7854$). Usando 5000 dígitos generados desde U$(0, 1)$ tenemos

$$\widehat{\theta}_M = \frac{1}{5000} \sum_{j=1}^{5000} \sqrt{1 - x_j^2} = 0.7868223,$$

con error estándar $\sqrt{v_M} = 0.0031373$.

CDF Normal

Considere generar una muestra de tamaño M, digamos $x_1, \dots, x_M$. La aproximación de

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \, \mathrm{d}y,$$

mediante Monte Carlo es dada por:

$$\widehat{\Phi}(z) = \frac{1}{M} \sum_{j=1}^{M} I_{\{x_j \leq z\}}$$

con varianza exacta $\Phi(z)(1 - \Phi(z))/M$.[5]

Considere $z = 1.65$, entonces

$$\widehat{\Phi}(z) = 0.9514$$

con varianza

$$\Phi(1.65)(1 - \Phi(1.65))/10\,000 = 4.7024 \cdot 10^{-6}$$

[5] pues las v.a $I_{(-\infty,z]}(x_j)$ son Bernoulli con prob. de éxito $\Phi(z)$

Integración Monte Carlo: Ejemplos

Código en R:

```
# aproxima pnorm(1.65) usando 10000 observaciones
m <- 10^4
x <- rnorm(m)
p <- x < 1.65
mean(p)
[1] 0.9511
```

Mientras que el valor exacto es $\Phi(1.65) = 0.9505285$ (usando `pnorm(1.65)`).

Observación:

Existe varias técnicas para mejorar la precisión de la estimación Monte Carlo. (técnicas de reducción de varianza)

Una alternativa al muestreo directo desde f para evaluar $\mathrm{E}_f\{\boldsymbol{h}(\boldsymbol{X})\}$ es usar importance sampling (o ponderado).

Importance Sampling

Considere $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M$ muestra aleatoria simulada desde la densidad instrumental g. Importance Sampling permite aproximar $\mathsf{E}_f\{\boldsymbol{h}(\boldsymbol{X})\}$ como

$$\widetilde{\boldsymbol{\theta}}_M = \frac{1}{M} \sum_{j=1}^{M} \frac{f(\boldsymbol{x}_j)}{g(\boldsymbol{x}_j)} \boldsymbol{h}(\boldsymbol{x}_j).$$

Este método está basado en la representación alternativa

$$\mathsf{E}_f\{\boldsymbol{h}(\boldsymbol{X})\} = \int \boldsymbol{h}(\boldsymbol{x}) \frac{f(\boldsymbol{x})}{g(\boldsymbol{x})} g(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} =: \mathsf{E}_g \left\{ \boldsymbol{h}(\boldsymbol{x}) \frac{f(\boldsymbol{x})}{g(\boldsymbol{x})} \right\}$$

Observaciones:

- Note que el término $\omega_j = f(\boldsymbol{x}_j)/g(\boldsymbol{x}_j)$ luce como un peso y $\widetilde{\boldsymbol{\theta}}_M$ puede ser escrito como
$$\widetilde{\boldsymbol{\theta}}_M = \frac{1}{M}\sum_{j=1}^{M}\omega_j \boldsymbol{h}(\boldsymbol{x}_j).$$
- $\widetilde{\boldsymbol{\theta}}_M \xrightarrow{\text{a.s}} \boldsymbol{\theta}$ $(= \mathsf{E}_f\{\boldsymbol{h}(\boldsymbol{X}\})$. Sin embargo, es requerido que
  - $\text{supp}(g) \supseteq \text{supp}(f)$.[6]
  - g debe ser simple de simular.
- La misma muestra puede ser re-utilizada para diferentes $\boldsymbol{h}$ y f.
- Se siguiere que g no tenga colas más livianas que f.

[6] Si la razón f/g es no acotada los pesos ω pueden fluctuar bruscamente.

Importance Sampling: Ejemplos

Probabilidades en la cola de la normal

Suponga $Z \sim \mathsf{N}(0,1)$ y considere que se está interesado en calcular $P(Z > 4.5)$ $(= 3.3977 \cdot 10^{-6})$. Podemos simular $z_1, \dots, z_M$ desde $\mathsf{N}(0,1)$ y calcular

$$P(Z > 4.5) \approx \frac{1}{M} \sum_{j=1}^{M} I_{(4.5,\infty)}(z_j),$$

usando $M = 10\,000$ usualmente produce todos ceros de la función indicadora.

Considere $Y \sim \mathsf{TExp}(4.5, 1)$ una distribución exponencial truncada en 4.5 con escala 1, cuya densidad es dada por

$$g(y) = \frac{e^{-(y-4.5)}}{\int_{4.5}^{\infty} e^{-y}\, \mathrm{d}y}$$

Simulando desde g y usando Importance Sampling, obtenemos

$$P(Z > 4.5) \approx \frac{1}{M} \sum_{j=1}^{M} \frac{\phi(y_j)}{g(y_j)} I_{(4.5,\infty)}(y_j) = 0.000003377$$

Sugonga $-h(\boldsymbol{z})$ función unimodal, suave y acotada, con máximo en $\widehat{\boldsymbol{z}}$. El método de Laplace aproxima la integral

$$\theta = \int_{\mathbb{R}^q} f(\boldsymbol{z}) \exp\{-nh(\boldsymbol{z})\} \, \mathrm{d}\boldsymbol{z},$$

considerando una aproximación de segundo orden de $h(\boldsymbol{z})$ en torno de $\widehat{\boldsymbol{z}}$, como

$$h(\boldsymbol{z}) \approx h(\widehat{\boldsymbol{z}}) + (\boldsymbol{z} - \widehat{\boldsymbol{z}})^\top \boldsymbol{g}(\widehat{\boldsymbol{z}}) + \frac{1}{2}(\boldsymbol{z} - \widehat{\boldsymbol{z}})^\top \boldsymbol{H}(\widehat{\boldsymbol{z}})(\boldsymbol{z} - \widehat{\boldsymbol{z}})$$

donde

$$\boldsymbol{g}(\boldsymbol{z}) = \frac{\partial h(\boldsymbol{z})}{\partial \boldsymbol{z}}, \qquad \boldsymbol{H}(\boldsymbol{z}) = \frac{\partial^2 h(\boldsymbol{z})}{\partial \boldsymbol{z} \partial \boldsymbol{z}^\top}.$$

Como

$$\boldsymbol{g}(\widehat{\boldsymbol{z}}) = \frac{\partial h(\boldsymbol{z})}{\partial \boldsymbol{z}}\Big|_{z=\widehat{z}} = \mathbf{0}.$$

Tenemos

$$\begin{aligned}\theta &\approx \int_{\mathbb{R}^q} f(\widehat{\boldsymbol{z}}) \exp\Big\{ -n\Big[h(\widehat{\boldsymbol{z}}) + \frac{1}{2}(\boldsymbol{z}-\widehat{\boldsymbol{z}})^\top \boldsymbol{H}(\widehat{\boldsymbol{z}})(\boldsymbol{z}-\widehat{\boldsymbol{z}})\Big]\Big\}\, \mathrm{d}\boldsymbol{z} \\ &= f(\widehat{\boldsymbol{z}}) \exp\{-nh(\widehat{\boldsymbol{z}})\} \int_{\mathbb{R}^q} \exp\Big\{ -\frac{1}{2}(\boldsymbol{z}-\widehat{\boldsymbol{z}})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{z}-\widehat{\boldsymbol{z}})\Big]\Big\}\, \mathrm{d}\boldsymbol{z}\end{aligned}$$

con $\boldsymbol{\Sigma}^{-1} = n\boldsymbol{H}(\widehat{\boldsymbol{z}})$.

De este modo, la aproximación de Laplace de primer orden, está dada por

$$\widehat{\theta} = f(\widehat{\boldsymbol{z}})\sqrt{\frac{2\pi}{n}}\,|\boldsymbol{H}(\widehat{\boldsymbol{z}})|^{-1/2}\exp\{-nh(\widehat{\boldsymbol{z}})\}.$$

Se puede mostrar que para $q < \infty$ fijado,

$$I = \widehat{\theta}\{1 + O(n^{-1})\}.$$

Observaciones:

- Existen extensiones para aproximaciones de orden mayor.
- El método de Laplace convierte un problema de integración en uno de optimización.