

Optimización Monte Carlo

¿Qué es la Integración de Monte Carlo?

Imagina enfrentarte a una integral tan complicada que los métodos tradicionales, como la regla del trapecio o Simpson, se vuelven inútiles. Aquí entra en juego **Monte Carlo**, un enfoque probabilístico que utiliza números aleatorios para aproximar integrales.

La idea central es simple: si podemos muestrear puntos al azar dentro de un dominio, podemos estimar el valor promedio de una función en ese dominio y usarlo para aproximar la integral.

Formalmente, si queremos calcular:

$$I = \int_a^b f(x) dx,$$

podemos reescribirlo como:

$$I = (b - a) \cdot \mathbb{E}[f(X)],$$

donde X es una variable aleatoria uniformemente distribuida en $[a, b]$.

Integral de Gauss

Consideremos la integral:

$$I = \int_{-\infty}^{\infty} e^{-x^2} dx.$$

Esta integral no tiene solución analítica en términos de funciones elementales, pero sabemos que su valor exacto es $\sqrt{\pi}$.

Para aplicar Monte Carlo, primero truncamos el dominio a un intervalo finito, digamos $[-10, 10]$, ya que e^{-x^2} decae rápidamente.

Integral de Gauss

Luego:

1. Generamos N puntos aleatorios x_1, x_2, \dots, x_N distribuidos uniformemente en $[-10, 10]$.
2. Evaluamos $f(x) = e^{-x^2}$ en cada punto, obteniendo $f(x_1), f(x_2), \dots, f(x_N)$.
3. Calculamos el promedio de estos valores:

$$\bar{f} = \frac{1}{N} \sum_{i=1}^N f(x_i).$$

4. Multiplicamos este promedio por la longitud del intervalo $(b - a) = 20$:

$$I \approx 20 \cdot \bar{f} = 20 \cdot \frac{1}{N} \sum_{i=1}^N e^{-x_i^2}.$$

Integrales Múltiples

Ahora, enfrentémonos a un problema más complejo. Consideremos la integral doble:

$$I = \int_0^1 \int_0^1 \frac{\sin(x+y)}{1+x^2+y^2} dx dy.$$

Este tipo de integral es difícil de resolver analíticamente debido a la interacción entre las variables x y y .

Integrales múltiples

1. Generamos N pares (x_i, y_i) distribuidos uniformemente en $[0, 1] \times [0, 1]$.
2. Evaluamos la función:

$$f(x, y) = \frac{\sin(x + y)}{1 + x^2 + y^2},$$

obteniendo $f(x_1, y_1), f(x_2, y_2), \dots, f(x_N, y_N)$.

3. Calculamos el promedio de estos valores:

$$\bar{f} = \frac{1}{N} \sum_{i=1}^N f(x_i, y_i).$$

4. Multiplicamos este promedio por el área del dominio (en este caso, $1 \times 1 = 1$):

$$I \approx 1 \cdot \bar{f} = \frac{1}{N} \sum_{i=1}^N \frac{\sin(x_i + y_i)}{1 + x_i^2 + y_i^2}.$$

Definición

Estimar:

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

con $x_i \sim \mathcal{U}(a, b)$:

$$\hat{I}_n = (b - a) \cdot \frac{1}{n} \sum_{i=1}^n f(x_i)$$

$$\int_0^1 x^2 dx$$

```
set.seed(1)
n <- 10000
x <- runif(n)
fx <- x^2
mean(fx)
```


$$\int_0^1 e^{-x^2} dx$$

No tiene primitiva elemental.

```
x <- runif(n)
fx <- exp(-x^2)
mean(fx)
```

Integral doble

$$I = \int_0^1 \int_0^1 xy \, dx \, dy = \frac{1}{4}$$

```
x <- runif(n)
y <- runif(n)
fx <- x * y
mean(fx)
```

Optimización determinista vs estocástica

- ▶ Métodos deterministas: gradiente descendente, Newton-Raphson.
- ▶ Métodos estocásticos: simulación de Monte Carlo, algoritmos genéticos, recocido simulado.

Funciones de distribución y densidad

La función de distribución acumulada de una variable aleatoria real X se define como:

$$P(x) = \Pr\{X \leq x\}$$

La correspondiente función de densidad de probabilidad es:

$$p(x) = \frac{dP}{dx}$$

Esto implica:

$$\Pr\{X \in A\} = \int_A p(x) dx = P(A)$$

Distribución y densidad conjuntas

Para un vector aleatorio $X = (X_1, X_2, \dots, X_s)$:

- Función de distribución acumulada conjunta:

$$P(x_1, \dots, x_s) = \Pr\{X_i \leq x_i \text{ para todo } i = 1, \dots, s\}$$

- Función de densidad conjunta:

$$p(x_1, \dots, x_s)$$

- Para un subconjunto medible $D \subset \mathbb{R}^s$:

$$\Pr\{X \in D\} = \int_D p(x_1, \dots, x_s) dx_1 \cdots dx_s$$

Medida de probabilidad y generalización

Sea X una variable aleatoria con valores en un dominio Ω .
Su **medida de probabilidad** es una función medida P tal que,
para todo conjunto medible D :

$$P(D) = \Pr\{X \in D\}$$

Se debe cumplir:

$$P(\Omega) = 1$$

Densidad como derivada de Radon-Nikodym

La función de densidad p se define como la derivada de Radon-Nikodym:

$$p(x) = \frac{dP}{d\mu}(x)$$

Esto significa:

$$P(D) = \int_D p(x) d\mu(x)$$

Lo anterior generaliza los casos anteriores con integración respecto a la medida de Lebesgue.

Valor esperado

El valor esperado o esperanza de una variable aleatoria Y se define como:

$$\mathbb{E}[Y] = \int f(x) p(x) dx$$

Donde $f(x)$ es una función medible de interés y $p(x)$ es la densidad de probabilidad correspondiente.

Asumimos que $\mathbb{E}[Y]$ existe, es decir, que la integral es finita.

Varianza

La varianza de una variable aleatoria Y se define como:

$$\text{Var}[Y] = \mathbb{E} \left[(Y - \mathbb{E}[Y])^2 \right]$$

Si $\mathbb{E}[Y]$ y $\mathbb{E}[Y^2]$ son finitas, entonces la varianza también lo es.

Propiedades del valor esperado y la varianza

Para cualquier constante a , se cumplen las siguientes propiedades:

$$\mathbb{E}[aY] = a\mathbb{E}[Y]$$

$$\text{Var}[aY] = a^2 \text{Var}[Y]$$

También se cumple:

$$\mathbb{E}\left[\sum_{i=1}^N Y_i\right] = \sum_{i=1}^N \mathbb{E}[Y_i]$$

Y si Y_1, \dots, Y_N son independientes:

$$\text{Var}\left[\sum_{i=1}^N Y_i\right] = \sum_{i=1}^N \text{Var}[Y_i]$$

Forma alternativa de la varianza

A partir de las definiciones, se puede deducir una forma equivalente de la varianza:

$$\text{Var}[Y] = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2$$

Esta identidad es útil para simplificar cálculos analíticos y computacionales.

Desviación estándar

La desviación estándar de Y se define como:

$$\sigma_Y = \sqrt{\text{Var}[Y]}$$

Esta cantidad también se conoce como *error cuadrático medio* o *RMS error*.

Densidad conjunta

Sean X y Y variables aleatorias con densidad conjunta $p(x, y)$.
Para cualquier conjunto medible D :

$$P(D) = \int_D p(x, y) \, dx \, dy$$

Densidad marginal y condicional

Densidad marginal de X :

$$p(x) = \int p(x, y) dy$$

Densidad condicional de Y dado $X = x$:

$$p(y|x) = \frac{p(x, y)}{p(x)}$$

Identidad útil:

$$p(x, y) = p(y|x) p(x) = p(x|y) p(y)$$

Esperanza condicional

Sea $G = g(X, Y)$. El valor esperado condicional dado X es:

$$\mathbb{E}[G|X] = \int g(x, y) p(y|x) dy = \frac{\int g(x, y) p(x, y) dy}{\int p(x, y) dy}$$

También usamos la notación:

$$\mathbb{E}_Y[G]$$

para enfatizar integración respecto a la densidad de Y .

Descomposición de la varianza

Se cumple la identidad:

$$\text{Var}[G] = \mathbb{E}_X[\text{Var}_Y[G]] + \text{Var}_X[\mathbb{E}_Y[G]]$$

Esto expresa la varianza total como:

- ▶ La media de la varianza condicional
- ▶ Más la varianza de la media condicional

Demostración de la identidad de varianza

$$\text{Var}[F] = \mathbb{E}[F^2] - (\mathbb{E}[F])^2$$

Luego:

$$\mathbb{E}_X[\text{Var}_Y[G]] + \text{Var}_X[\mathbb{E}_Y[G]] = \mathbb{E}[G^2] - (\mathbb{E}[G])^2 = \text{Var}[G]$$

Esta identidad se usa en métodos de reducción de varianza, como el muestreo estratificado.

Integración Monte Carlo

Queremos estimar:

$$I = \int_{\mathcal{X}} f(x) dx$$

Tomamos N muestras independientes $X_1, \dots, X_N \sim p(x)$ y usamos el estimador:

$$F_N = \frac{1}{N} \sum_{i=1}^N \frac{f(X_i)}{p(X_i)}$$

Relación con Horvitz-Thompson

Este estimador es aleatorio. Su esperanza es:

$$\mathbb{E}[F_N] = \frac{1}{N} \sum_{i=1}^N \int \frac{f(x)}{p(x)} p(x) dx = \int f(x) dx = I$$

Esta forma se conoce también como el estimador de Horvitz-Thompson.

Se requiere que $f(x)/p(x)$ sea finito donde $f(x) \neq 0$.

Caso uniforme

Si X_i se distribuyen uniformemente sobre un dominio finito de volumen V , entonces:

$$F_N = \frac{1}{N} \sum_{i=1}^N f(X_i)$$

Este caso tiene la misma forma que una regla de cuadratura, pero con puntos aleatorios.

Convergencia independiente de la dimensión

- El error del método Monte Carlo disminuye a una tasa de $O(N^{-1/2})$, donde N es el número de muestras, sin depender de la dimensión del espacio ni de la suavidad de la función integrando. Esto contrasta con métodos deterministas, cuya eficiencia suele deteriorarse en espacios de alta dimensión. Gracias a esta propiedad, Monte Carlo es especialmente útil en contextos donde es común evaluar integrales de alta dimensión sobre funciones que pueden ser discontinuas o presentar fuertes variaciones locales.

Simplicidad

- ▶ Solo se requieren dos operaciones básicas: muestreo y evaluación puntual. Esto favorece el uso de interfaces modulares tipo caja negra, lo que proporciona gran flexibilidad en el diseño de software Monte Carlo.

Generalidad

- ▶ Como se basa en el muestreo aleatorio, puede aplicarse incluso en dominios que no tienen una correspondencia natural con \mathbb{R}^s y que por lo tanto no son adecuados para la cuadratura numérica. Por ejemplo, en simulaciones físicas, es común estimar integrales sobre espacios de estados con restricciones complejas y discontinuidades, donde los métodos deterministas tradicionales resultan ineficientes o inviables. Monte Carlo permite abordar directamente estas integrales mediante muestreo en espacios de alta dimensión.

Manejo de singularidades

- ▶ Los métodos Monte Carlo son más adecuados que las reglas de cuadratura para integrandos con singularidades. El muestreo por importancia¹ puede aplicarse para manejar tales integrandos de manera efectiva, incluso en situaciones donde no existe una transformación analítica que elimine la singularidad.

¹El muestreo por importancia es una técnica que mejora la eficiencia del método Monte Carlo al tomar muestras de una distribución distinta de la original, dando mayor peso a las regiones del dominio que más contribuyen al valor esperado de la función. Esto reduce la varianza del estimador sin introducir sesgo.

¿Qué es el muestreo por importancia?

El muestreo por importancia es una técnica utilizada para estimar integrales del tipo:

$$\mu = \int h(x)f(x) dx$$

cuando $f(x)$ es una densidad difícil de muestrear directamente o cuando $h(x)f(x)$ tiene alta varianza.

Se introduce una densidad auxiliar $g(x)$ (llamada densidad de importancia), tal que $f(x) > 0 \Rightarrow g(x) > 0$, y se reescribe la integral como:

$$\mu = \int h(x) \frac{f(x)}{g(x)} g(x) dx = \mathbb{E}_g \left[h(X) \frac{f(X)}{g(X)} \right]$$

Esta expectativa se puede estimar mediante simulaciones de $X \sim g$.

Algoritmo de muestreo por importancia

Algorithm 1 Muestreo por importancia

- 1: **Entrada:** número de muestras N , función $h(x)$, densidades $f(x)$ y $g(x)$
- 2: **for** $i = 1$ to N **do**
- 3: Generar $X_i \sim g(x)$
- 4: Calcular peso: $w_i = \frac{f(X_i)}{g(X_i)}$
- 5: Calcular valor ponderado: $Y_i = h(X_i) \cdot w_i$
- 6: **end for**
- 7: Estimador:

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N Y_i$$

¿Qué es el muestreo por rechazo?

El muestreo por rechazo es un método para generar observaciones de una distribución objetivo $f(x)$ cuando no se puede muestrear directamente de ella.

La idea es:

- ▶ Encontrar una densidad auxiliar $g(x)$ fácil de muestrear.
- ▶ Asegurar que existe una constante $M > 0$ tal que $f(x) \leq Mg(x)$ para todo x .
- ▶ Aceptar o rechazar muestras de g según una regla probabilística.

Algoritmo de muestreo por rechazo

Algorithm 2 Muestreo por rechazo

- 1: **Entrada:** densidad objetivo $f(x)$, densidad propuesta $g(x)$, constante $M > 0$
 - 2: **while** TRUE **do**
 - 3: Generar $X \sim g(x)$
 - 4: Generar $U \sim \text{Uniform}(0, 1)$
 - 5: **if** $U \leq \frac{f(X)}{Mg(X)}$ **then**
 - 6: **return** X ▷ Aceptar muestra
 - 7: **end if**
 - 8: **end while**
-

Estimación Monte Carlo

En Monte Carlo se estima un valor esperado

$$\mu = \mathbb{E}[h(X)]$$

usando N muestras independientes $X_1, \dots, X_N \sim f(x)$ y el estimador:

$$\hat{\mu}_N = \frac{1}{N} \sum_{i=1}^N h(X_i)$$

Este estimador es una variable aleatoria. Su comportamiento asintótico está respaldado por las leyes de los grandes números.

Ley débil de los grandes números (WLLN)

Supuesto: X_1, X_2, \dots son i.i.d. y $\mathbb{E}[|h(X)|] < \infty$.

Entonces:

$$\hat{\mu}_N = \frac{1}{N} \sum_{i=1}^N h(X_i) \xrightarrow{P} \mu = \mathbb{E}[h(X)]$$

Es decir, $\hat{\mu}_N$ converge en probabilidad a μ cuando $N \rightarrow \infty$.

Esto garantiza que estimadores Monte Carlo son consistentes en el sentido débil.

Ley fuerte de los grandes números (SLLN)

Supuesto: X_1, X_2, \dots son i.i.d. y $\mathbb{E}[|h(X)|] < \infty$.

Entonces:

$$\hat{\mu}_N = \frac{1}{N} \sum_{i=1}^N h(X_i) \xrightarrow{\text{c.s.}} \mu = \mathbb{E}[h(X)]$$

Esto significa que la convergencia ocurre casi seguramente:

$$\Pr \left(\lim_{N \rightarrow \infty} \hat{\mu}_N = \mu \right) = 1$$

La SLLN asegura que el error Monte Carlo se estabiliza con $N \rightarrow \infty$, con probabilidad 1.

Estimadores Monte Carlo

Dado un valor esperado

$$\mu = \mathbb{E}[h(X)]$$

el estimador de Monte Carlo basado en N muestras i.i.d.
 $X_1, \dots, X_N \sim f(x)$ es:

$$\hat{\mu}_N = \frac{1}{N} \sum_{i=1}^N h(X_i)$$

La Ley de los Grandes Números garantiza que $\hat{\mu}_N \rightarrow \mu$, pero no describe su distribución para N finito.

Teorema Central del Límite

Supuestos:

- ▶ X_1, X_2, \dots son i.i.d.
- ▶ $h(X)$ tiene varianza finita: $\mathbb{V}[h(X)] = \sigma^2 < \infty$

Entonces:

$$\sqrt{N}(\hat{\mu}_N - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

Equivalente:

$$\hat{\mu}_N \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{N}\right) \quad \text{para } N \text{ grande}$$

Consecuencias

- ▶ Permite construir intervalos de confianza para μ :

$$\hat{\mu}_N \pm z_{\alpha/2} \cdot \frac{\hat{\sigma}}{\sqrt{N}}$$

donde $\hat{\sigma}^2$ es una estimación empírica de la varianza.

- ▶ Justifica el uso de métodos de control del error y precisión deseada.
- ▶ Respalda el diseño adaptativo de simulaciones: incrementar N hasta lograr un ancho de intervalo determinado.

Estimación de la varianza

La varianza del estimador se estima como:

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^N (h(X_i) - \hat{\mu}_N)^2$$

Luego, el error estándar del estimador es:

$$\text{SE}(\hat{\mu}_N) = \frac{\hat{\sigma}}{\sqrt{N}}$$

Motivación

En Monte Carlo estándar, todas las muestras se obtienen de una única distribución sobre el dominio completo.

Sin embargo, si la función integrando $f(x)$ presenta variación significativa en distintas regiones, es posible reducir la varianza dividiendo el dominio en subregiones homogéneas, conocidas como **estratos**.

Definición formal

Sea \mathcal{X} el dominio de integración. Lo dividimos en L estratos disjuntos:

$$\mathcal{X} = \bigcup_{\ell=1}^L \mathcal{X}_{\ell}, \quad \mathcal{X}_i \cap \mathcal{X}_j = \emptyset \text{ si } i \neq j$$

Queremos estimar:

$$\mu = \int_{\mathcal{X}} f(x) dx = \sum_{\ell=1}^L \int_{\mathcal{X}_{\ell}} f(x) dx$$

Simulamos n_{ℓ} muestras dentro de cada estrato \mathcal{X}_{ℓ} y calculamos:

$$\hat{\mu}_{\ell} = \frac{1}{n_{\ell}} \sum_{i=1}^{n_{\ell}} f(x_{\ell,i})$$

$$\hat{\mu} = \sum_{\ell=1}^L w_{\ell} \hat{\mu}_{\ell}, \quad w_{\ell} = \Pr(X \in \mathcal{X}_{\ell})$$

Varianza del estimador estratificado

La varianza total es:

$$\text{Var}[\hat{\mu}] = \sum_{\ell=1}^L \frac{w_{\ell}^2}{n_{\ell}} \sigma_{\ell}^2$$

donde σ_{ℓ}^2 es la varianza de $f(x)$ dentro del estrato \mathcal{X}_{ℓ} .

Asignación óptima de muestras:

$$\frac{n_{\ell}}{n} = \frac{w_{\ell} \sigma_{\ell}}{\sum_{j=1}^L w_j \sigma_j}$$

Esto minimiza la varianza total bajo un presupuesto fijo n .

Visualización

División del dominio en dos estratos

