

Integración Monte Carlo

Estas notas están en plena construcción, por lo mismo, es posible que encuentre errores. Si tal es el caso, repoórtelos.

1. Una breve historia

Los métodos Monte Carlo se originaron en el Laboratorio Nacional de Los Álamos en los primeros años después de la Segunda Guerra Mundial. La primera computadora electrónica en los Estados Unidos acababa de ser completada (el ENIAC), y los científicos de Los Álamos estaban considerando cómo usarla para el diseño de armas termonucleares (la bomba H). A fines de 1946, Stanislaw Ulam sugirió el uso del muestreo aleatorio para simular las trayectorias de vuelo de los neutrones, y John von Neumann desarrolló una propuesta detallada a comienzos de 1947. Esto llevó a simulaciones a pequeña escala cuyos resultados fueron indispensables para completar el proyecto. Metropolis y Ulam publicaron un artículo en 1949 describiendo sus ideas, lo que dio lugar a una gran cantidad de investigaciones en la década de 1950. El nombre del método Monte Carlo proviene de una ciudad en Mónaco, famosa por sus casinos (como lo sugirió Nick Metropolis, otro pionero del método Monte Carlo).

En casos aislados, el muestreo aleatorio había sido utilizado mucho antes para resolver problemas numéricos. Por ejemplo, en 1777 el Conde de Buffon realizó un experimento en el que se dejaba caer una aguja muchas veces sobre una tabla marcada con líneas paralelas equidistantes. Siendo L la longitud de la aguja y $d > L$ la distancia entre las líneas, demostró que la probabilidad de que la aguja intercepte una línea es:

$$p = \frac{2L}{\pi d}$$

Muchos años después, Laplace señaló que esto podría usarse como un medio rudimentario para estimar el valor de π .

De manera similar, Lord Kelvin utilizó lo que hoy llamaríamos un método Monte Carlo para estudiar algunos aspectos de la teoría cinética de los gases. Su generador de números aleatorios consistía en sacar pedazos de papel de un frasco de vidrio. La posibilidad de sesgo era una preocupación importante; temía que los papeles no se mezclaran lo suficiente debido a la electricidad estática. Otro experimentador temprano del método Monte Carlo fue Student (un seudónimo de W. S. Gosset),

quien utilizó el muestreo aleatorio como una ayuda para adivinar la forma de su famosa distribución t.

Una excelente referencia sobre los orígenes de los métodos Monte Carlo es el número especial de *Los Alamos Science* publicado en memoria de Stanislaw Ulam (1987). Los libros de Kalos y Whitlock (1986) y de Hammersley y Handscomb (1964) también contienen breves historias, incluyendo información sobre los experimentos de muestreo aleatorio anteriores a la guerra descritos anteriormente.

2.2 Reglas de cuadratura para la integración numérica

En esta sección explicamos por qué las técnicas estándar de integración numérica no funcionan muy bien en dominios de alta dimensión, especialmente cuando el integrando no es suave.

Consideremos una integral de la forma:

$$I = \int_{\Omega} f(x) d\mu(x)$$

donde Ω es el dominio de integración, $f : \mathbb{R}^s \rightarrow \mathbb{R}$ es una función con valores reales, y μ es una función medida sobre Ω .¹ Por ahora, supongamos que el dominio es el hipercubo unidad de dimensión s y que la función medida es $d\mu(x) = dx$, donde x_j denota la j -ésima componente del punto $x = (x_1, \dots, x_s)$.

Integrales de este tipo suelen aproximarse usando una regla de cuadratura, que es simplemente una suma de la forma:

$$I \approx \sum_{i=1}^N w_i f(x_i)$$

donde los pesos w_i y los puntos de muestreo x_i se determinan de antemano. Ejemplos comunes de reglas de cuadratura unidimensionales incluyen las reglas de Newton-Cotes (es decir, la regla del punto medio, la del trapecio, la regla de Simpson, etc.) y las reglas de Gauss-Legendre (véase Davis y Rabinowitz, 1984, para más detalles). Las formas de n puntos de estas reglas típicamente logran una tasa de convergencia de $O(n^{-r})$ para algún entero $r \geq 1$, siempre que el integrando tenga un número suficientemente grande de derivadas continuas. Por ejemplo, el error al usar la regla de Simpson es $O(n^{-4})$, siempre que f tenga al menos cuatro derivadas continuas (Davis y Rabinowitz, 1984).

Aunque estas reglas de cuadratura funcionan muy bien en integrales unidimensionales, aparecen problemas al extenderlas a dimensiones superiores. Por ejemplo, un enfoque común es usar reglas producto tensorial de la forma:

$$I \approx \sum_{i_1=1}^n \sum_{i_2=1}^n \cdots \sum_{i_s=1}^n w_{i_1} w_{i_2} \cdots w_{i_s} f(x_{i_1}, x_{i_2}, \dots, x_{i_s})$$

¹Ejemplos familiares de medidas incluyen longitud, área superficial, volumen y ángulo sólido; véase Halmos (1950) para una introducción a la teoría de la medida.

donde s es la dimensión, y los w_i y x_i son los pesos y puntos de muestreo de una regla unidimensional dada. Este método tiene la misma tasa de convergencia que la regla unidimensional en la que se basa (digamos $O(n^{-r})$), pero utiliza un número mucho mayor de puntos de muestra (es decir, $N = n^s$). Por lo tanto, en términos del número total de muestras, la tasa de convergencia es solo $O(N^{-r/s})$. Esto implica que la eficiencia de las reglas producto tensorial disminuye rápidamente con la dimensión, un hecho que se conoce como la maldición de la dimensionalidad (Niederreiter, 1992, p. 2).

La tasa de convergencia puede incrementarse utilizando una regla unidimensional con un valor mayor de r , pero esto presenta dos problemas. Primero, el número total de muestras $N = n^s$ puede volverse impráctico en dimensiones altas, ya que n crece linealmente con r (específicamente, $n \geq r$). Por ejemplo, la cuadratura de Gauss de dos puntos requiere al menos 2^s muestras, mientras que la regla de Simpson requiere al menos 3^s muestras. Segundo, tasas de convergencia más rápidas requieren mayor suavidad del integrando. Por ejemplo, si la función f tiene una discontinuidad, entonces la tasa de convergencia de cualquier regla de cuadratura unidimensional es como mucho $O(n^{-1})$ (suponiendo que la ubicación de la discontinuidad no se conoce de antemano), por lo que la correspondiente regla producto tensorial converge a una tasa no mejor que $O(N^{-1/s})$.

Por supuesto, no todas las reglas de integración multidimensional toman la forma de productos tensoriales. Sin embargo, hay un resultado importante que limita la tasa de convergencia de cualquier regla de cuadratura determinista, llamado el teorema de Bakhvalov (Davis y Rabinowitz, 1984, p. 354). Esencialmente, dice que dada cualquier regla de cuadratura de dimensión s , existe una función f con r derivadas continuas y acotadas, para la cual el error es proporcional a $N^{-r/s}$. Específicamente, sea C_M^r el conjunto de funciones $f : \mathbb{R}^s \rightarrow \mathbb{R}$ tales que $\|\nabla^r f(x)\| \leq M$ para todo $x \in \Omega$, donde M es una constante. Consideremos ahora cualquier regla de cuadratura con N puntos:

$$I(f) \approx \sum_{i=1}^N w_i f(x_i)$$

donde cada x_i es un punto en $\Omega \subset \mathbb{R}^s$, y supongamos que deseamos aproximar alguna integral

$$I(f) = \int_{\Omega} f(x_1, \dots, x_s) dx_1 \cdots dx_s$$

Entonces, de acuerdo con el teorema de Bakhvalov, existe una función $f \in C_M^r$ tal que el error cumple:

$$|I(f) - \sum_{i=1}^N w_i f(x_i)| \geq k N^{-r/s}$$

donde la constante k depende solo de M y r . Así, incluso si f tiene una derivada primera acotada y continua, ninguna regla de cuadratura puede tener una cota de error mejor que $O(N^{-1/s})$.

2.3 Un poco de teoría de la probabilidad

Antes de describir la integración Monte Carlo, revisamos algunos conceptos de probabilidad y estadística. Véase Pitman (1993) para una introducción a la probabilidad y Halmos (1950) para una introducción a la teoría de la medida. Introducciones breves a la teoría de la probabilidad también se pueden encontrar en las referencias de Monte Carlo citadas anteriormente.

2.3.1 Funciones de distribución acumulada y funciones de densidad

Recordemos que la función de distribución acumulada de una variable aleatoria real X se define como:

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

y que la correspondiente función de densidad de probabilidad es:

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

(también conocida como la función de densidad o pdf). Esto lleva a la relación importante:

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

Las nociones correspondientes para un vector aleatorio multidimensional $X = (X_1, X_2, \dots, X_s)$ son la función de distribución acumulada conjunta:

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

y la función de densidad conjunta:

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

de modo que tenemos la relación:

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

para cualquier subconjunto medible de Lebesgue $D \subset \mathbb{R}^s$.

Más generalmente, para una variable aleatoria X con valores en un dominio arbitrario Ω , su medida de probabilidad (también conocida como distribución de probabilidad o simplemente distribución) es una función medida P tal que para cualquier conjunto medible D :

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

En particular, una medida de probabilidad debe satisfacer $P(\Omega) = 1$.

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

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

que es simplemente la función p que satisface:

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

Así, la probabilidad de que $X \in D$ se puede obtener integrando $p(x)$ sobre la región dada D . Esto debe compararse con las ecuaciones anteriores, que son casos especiales de esta relación general.

Obsérvese que la función de densidad p depende de la medida μ utilizada para definirla. Usaremos la notación $p = \frac{dP}{d\mu}$ para denotar la densidad con respecto a una medida en particular μ , correspondiente a la notación u_x o $u(x)$ que se utiliza a menudo en análisis. Esta notación será útil cuando haya varias funciones de medida relevantes definidas en el mismo dominio. Véase Halmos (1950) para mayor información sobre espacios de medida y derivadas de Radon-Nikodym.

2.3.2 Valor esperado y varianza

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

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

mientras que su varianza es:

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

Siempre asumiremos que el valor esperado y la varianza de toda variable aleatoria existen (es decir, que la integral correspondiente es finita).

A partir de estas definiciones, es fácil ver que para cualquier constante a se cumple:

$$\mathbb{E}[aY] = a \mathbb{E}[Y], \quad \text{Var}[aY] = a^2 \text{Var}[Y]$$

La siguiente identidad también es útil:

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

la cual vale para cualquier conjunto de variables aleatorias Y_1, \dots, Y_N . Por otro lado, la siguiente identidad solo se cumple si las variables Y_i son independientes:

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

Nótese que a partir de estas reglas, podemos derivar una expresión más simple para la varianza:

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

Otra cantidad útil es la desviación estándar de una variable aleatoria, que es simplemente la raíz cuadrada de su varianza:

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

Esto también se conoce como el error cuadrático medio (RMS error).

2.3.3 Densidades marginales y condicionales

Sea X y Y un par de variables aleatorias, de modo que el vector aleatorio (X, Y) esté definido en un dominio conjunto. Sea P la medida de probabilidad conjunta de (X, Y) , de modo que $P(D)$ representa la probabilidad de que $(X, Y) \in D$ para cualquier subconjunto medible D . Entonces, la función de densidad conjunta $p(x, y)$ satisface:

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

La función de densidad marginal de X se define como:

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

mientras que la función de densidad condicional $p(y|x)$ se define como:

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

De manera similar se definen la densidad marginal $p(y)$ y la densidad condicional $p(x|y)$, lo que lleva a la identidad útil:

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

Otro concepto importante es el valor esperado condicional de una variable aleatoria $G = g(X, Y)$, definido como:

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

También utilizaremos la notación $\mathbb{E}_Y[G]$ para el valor esperado condicional, que enfatiza que Y es la variable cuya densidad está siendo integrada.

Existe una expresión muy útil para la varianza de G en términos de su esperanza condicional y su varianza condicional, a saber:

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

En otras palabras, la varianza total de G es igual a la media de la varianza condicional más la varianza de la media condicional. Para demostrar esta identidad, recordemos que:

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

y observamos que:

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

Utilizaremos esta identidad más adelante para analizar ciertas técnicas de reducción de varianza, incluyendo el muestreo estratificado y el uso de valores esperados.

2.4 Integración Monte Carlo básica

La idea de la integración Monte Carlo es evaluar la integral:

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

utilizando muestreo aleatorio. En su forma básica, esto se realiza tomando N puntos X_1, X_2, \dots, X_N muestreados independientemente de acuerdo con alguna función de densidad conveniente $p(x)$, y luego computando el estimador:

$$F_N = \frac{1}{N} \sum_{i=1}^N \frac{f(X_i)}{p(X_i)} \quad (2.12)$$

Aquí usamos la notación F_N en lugar de I para enfatizar que el resultado es una variable aleatoria, y que sus propiedades dependen de cuántos puntos de muestra se hayan elegido. Nótese que este tipo de estimador fue utilizado por primera vez en la literatura de muestreo por encuestas (para dominios discretos en lugar de continuos), donde se le conoce como el estimador de Horvitz-Thompson (Horvitz y Thompson, 1952).

Por ejemplo, supongamos que el dominio es \mathbb{R}^s y que las muestras X_i son tomadas de manera independiente y uniforme. En ese caso, el estimador se reduce a:

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

lo cual tiene la misma forma que una regla de cuadratura, excepto que los puntos de muestra son aleatorios.

Es directo demostrar que el estimador F_N entrega el resultado correcto en promedio. Específicamente, se cumple:

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

siempre que $f(x)/p(x)$ sea finita siempre que $f(x) \neq 0$.

Ventajas de la integración Monte Carlo

La integración Monte Carlo tiene las siguientes ventajas principales:

- **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 gráficos por computador, 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. En el contexto de gráficos por computadora, por ejemplo, es sencillo incluir efectos como desenfoque por movimiento, profundidad de campo, medios participantes, superficies procedurales, entre otros.
- **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 de física de partículas, 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 importancia² puede aplicarse para manejar tales integrandos de manera efectiva, incluso en situaciones donde no existe una transformación analítica que elimine la singularidad (véase la discusión sobre el muestreo por rechazo y el método de Metropolis más adelante).

²El 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.

1.1. Tasas de convergencia

Para determinar la tasa de convergencia de la integración Monte Carlo, comenzamos calculando la varianza de F_N . Para simplificar la notación, definimos $Y_i = \frac{f(X_i)}{p(X_i)}$, de modo que:

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

Sea $Y = Y_i$. Entonces se cumple que:

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

Suponiendo que esta cantidad es finita, es fácil verificar que la varianza de F_N decrece linealmente con N :

$$\text{Var}[F_N] = \text{Var}\left(\frac{1}{N} \sum_{i=1}^N Y_i\right) = \frac{1}{N^2} \sum_{i=1}^N \text{Var}[Y_i] = \frac{1}{N} \text{Var}[Y] \quad (2.13)$$

donde se usó que $\text{Var}[aY] = a^2 \text{Var}[Y]$ y que las muestras Y_i son independientes.

Por lo tanto, la desviación estándar es:

$$\sigma_{F_N} = \sqrt{\text{Var}[F_N]} = \frac{\sigma_Y}{\sqrt{N}}$$

lo que muestra inmediatamente que el error RMS (root-mean-square) converge a una tasa de $O(N^{-1/2})$.

También es posible obtener cotas probabilísticas sobre el error absoluto, usando la desigualdad de Chebychev:

$$\Pr(|F - \mathbb{E}[F]| > \varepsilon) \leq \frac{\text{Var}[F]}{\varepsilon^2}$$

la cual se cumple para cualquier variable aleatoria F tal que $\text{Var}[F]$ sea finita. Aplicando esta desigualdad a la varianza, obtenemos:

$$\Pr(|F_N - I| > \varepsilon) \leq \frac{\text{Var}[Y]}{N\varepsilon^2}$$

Por lo tanto, para cualquier umbral fijo ε , el error absoluto decrece a una tasa de $O(N^{-1/2})$.

Se pueden obtener cotas más estrictas sobre el error absoluto utilizando el teorema central del límite, el cual afirma que F_N converge a una distribución normal en el límite cuando $N \rightarrow \infty$. Específicamente, establece que:

$$\lim_{N \rightarrow \infty} \Pr\left(\frac{\sum_{i=1}^N Y_i - \mathbb{E}[Y]}{\sqrt{N \text{Var}[Y]}} < t\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-x^2/2} dx$$

donde la expresión en el lado derecho es la función de distribución acumulada de la normal estándar. Esta ecuación puede reorganizarse como:

$$\Pr(|F_N - I| > t \cdot \sigma_{F_N}) \approx 2 \left(1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-x^2/2} dx \right)$$

La integral en el lado derecho decrece muy rápidamente con t ; por ejemplo, cuando $t = 3$, el valor es aproximadamente 0,003. Por lo tanto, hay solo una probabilidad de aproximadamente 0.3 % de que F_N difiera de su media en más de tres desviaciones estándar, siempre que N sea suficientemente grande para que el teorema central del límite se aplique.

Finalmente, nótese que la integración Monte Carlo puede converger incluso si la varianza $\text{Var}[Y]$ es infinita, siempre que la esperanza $\mathbb{E}[Y]$ exista (aunque la convergencia será más lenta). Esto está garantizado por la ley fuerte de los grandes números, la cual establece que:

$$\Pr \left(\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N Y_i = \mathbb{E}[Y] \right) = 1$$

1.2. Muestreo de variables aleatorias

Existen diversas técnicas para muestrear variables aleatorias, que revisamos brevemente aquí. Más detalles pueden encontrarse en las referencias citadas en la introducción.

Una técnica es el *método de transformación* o *método de inversión*. En una dimensión, supongamos que queremos muestrear desde una función de densidad p . Sea P la función de distribución acumulada correspondiente. El método de inversión consiste en definir $X = P^{-1}(U)$, donde U es una variable aleatoria uniforme en $[0, 1]$. Es fácil verificar que X tiene la densidad requerida p .

Esta técnica puede extenderse fácilmente a varias dimensiones, ya sea calculando las distribuciones marginales y condicionales e invirtiendo cada dimensión por separado, o más generalmente derivando una transformación $x = g(u)$ con un determinante Jacobiano apropiado (tal que $p(x) = |\det J_g(x)|$, donde J_g denota el Jacobiano de g).

La principal ventaja de la técnica de transformación es que permite estratificar fácilmente las muestras, estratificando el espacio de parámetros $[0, 1]^s$ y mapeando estas muestras a \mathcal{X} . Otra ventaja es que la técnica tiene un costo fijo por muestra, el cual puede estimarse fácilmente. La desventaja principal es que la densidad $p(x)$ debe integrarse analíticamente, lo que no siempre es posible. También es preferible que la función de distribución acumulada tenga una inversa analítica, ya que la inversión numérica es típicamente más lenta.

Una segunda técnica de muestreo es el *método de rechazo*, atribuida a von Neumann (véase Ulam, 1987). La idea es muestrear desde alguna densidad conveniente q tal que:

$$p(x) \leq Mq(x)$$

para alguna constante M . Generalmente, las muestras de q son generadas usando el método de transformación. Luego se aplica el siguiente procedimiento:

Algorithm 1: Muestreo por rechazo

```
1 Función RECHAZO_MUESTREO():  
2   repeat  
3     muestrear  $X$  según  $q$   
4     muestrear  $U \sim \text{Uniforme}[0, 1]$   
5     if  $U \leq \frac{p(X)}{M \cdot q(X)}$  then  
6       return  $X$   
7   until se acepta una muestra
```

Es fácil verificar que este procedimiento genera una muestra X cuya función de densidad es p .

La principal ventaja del muestreo por rechazo es que puede utilizarse con cualquier función de densidad, incluso aquellas que no pueden integrarse analíticamente. Sin embargo, aún necesitamos poder integrar alguna función Mq que actúe como cota superior de p . Además, esta cota debería ser razonablemente ajustada, ya que el número promedio de muestras necesarias antes de aceptar una es M . Por lo tanto, la eficiencia del muestreo por rechazo puede ser muy baja si se aplica de forma ingenua.

Otra desventaja es que es difícil aplicar esta técnica con estratificación: la aproximación más cercana es estratificar el dominio del vector aleatorio $X = U$, pero la estratificación resultante no es tan buena como la obtenida con el método de transformación.

2. Ejercicios

2.1.

Sea:

$$p(x) = 2x, \quad q(x) = 1, \quad x \in [0, 1]$$

El cociente es:

$$\frac{p(x)}{q(x)} = 2x$$

El máximo en $[0, 1]$ ocurre en $x = 1$, por lo tanto:

$$M = \max_{x \in [0, 1]} 2x = 2$$

2.2. Densidad polinómica en $[0, 1]$

Sea:

$$p(x) = 30x^2(1 - x)^2, \quad q(x) = 1^3, \quad x \in [0, 1]$$

³Cuando escribimos $q(x) = 1$ en el intervalo $[0, 1]$, nos referimos a la densidad uniforme sobre dicho intervalo. En términos generales, la densidad de una variable aleatoria $X \sim \text{Uniforme}(a, b)$ es $q(x) = \frac{1}{b-a}$ en $[a, b]$ y cero fuera. Por

El cociente es:

$$\frac{p(x)}{q(x)} = 30x^2(1-x)^2$$

Para encontrar el máximo derivamos:

$$f(x) = 30x^2(1-x)^2$$

$$f'(x) = 60x(1-x)^2 - 60x^2(1-x) = 60x(1-x)(1-2x)$$

Los puntos críticos son $x = 0, \frac{1}{2}, 1$. Evaluando:

$$f\left(\frac{1}{2}\right) = 30\left(\frac{1}{4} \cdot \frac{1}{4}\right) = \frac{30}{16} = 1,875$$

Por lo tanto:

$$M = 1,875$$

Normal truncada en [0,3]

Sea:

$$p(x) = \frac{1}{Z}e^{-x^2/2}, \quad q(x) = \frac{1}{3}, \quad x \in [0, 3]$$

con $Z = \Phi(3) - \Phi(0)$

El cociente es:

$$\frac{p(x)}{q(x)} = \frac{3}{Z}e^{-x^2/2}$$

El valor máximo de $e^{-x^2/2}$ en $[0, 3]$ ocurre en $x = 0$, así que:

$$M = \frac{3}{\Phi(3) - \Phi(0)} \approx 6,016$$

2.3. Estimadores y sus propiedades

El propósito de un estimador Monte Carlo es aproximar el valor de alguna cantidad de interés Q (también llamada el estimando). Normalmente definiremos Q como el valor de una integral dada, aunque también pueden considerarse situaciones más generales (por ejemplo, Q podría ser el cociente de dos integrales). Un estimador se define entonces como una función de la forma:

$$F_N = F_N(X_1, X_2, \dots, X_N) \tag{2.14}$$

donde los X_i son variables aleatorias. Un valor numérico particular de F_N se denomina una estimación. Nótese que los X_i no tienen por qué ser independientes, y pueden tener distribuciones distintas.

tanto, $q(x) = 1$ en $[0, 1]$ es una forma válida de expresar la densidad uniforme estándar.

Existe una diferencia en el uso de términos según el contexto. En algunos enfoques, a cada valor individual X_i se le llama observación, y al conjunto completo X_1, \dots, X_N se le denomina muestra, siendo N el tamaño muestral. Aquí, utilizaremos una convención distinta: cada X_i será referido como una muestra, y N como el número total de muestras generadas.

Definimos ahora varias propiedades útiles de los estimadores Monte Carlo. La cantidad $F_N - Q$ se llama el *error*, y su valor esperado se llama el *sesgo* (bias):

$$\text{Sesgo} = \mathbb{E}[F_N] - Q \quad (2.15)$$

Un estimador se dice no sesgado (unbiased) si:

$$\mathbb{E}[F_N] = Q \quad \text{para todo } N$$

Por ejemplo, la variable aleatoria

$$F_N = \frac{1}{N} \sum_{i=1}^N \frac{f(X_i)}{p(X_i)} \quad (2.16)$$

es un estimador no sesgado de la integral $I = \int f(x) dx$ (como vimos en la Sección 2.4).

Un estimador se dice *consistente* si el error $F_N - Q$ tiende a cero con probabilidad uno, es decir, si:

$$\Pr \left(\lim_{N \rightarrow \infty} F_N = Q \right) = 1 \quad (2.17)$$

Una condición suficiente para que un estimador sea consistente es que tanto su sesgo como su varianza tiendan a cero al aumentar N :

$$\lim_{N \rightarrow \infty} \mathbb{E}[F_N] = Q, \quad \lim_{N \rightarrow \infty} \text{Var}[F_N] = 0$$

En particular, un estimador no sesgado es consistente siempre que su varianza decrezca a cero cuando $N \rightarrow \infty$.

La razón principal para preferir estimadores no sesgados es que es más fácil estimar el error. Típicamente, nuestro objetivo es minimizar el *error cuadrático medio* (MSE), definido como:

$$\text{MSE}(F) = \mathbb{E} [(F - Q)^2] \quad (2.18)$$

En general, el MSE puede reescribirse como:

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

Por tanto, para estimar el error se requiere una cota superior del sesgo, lo cual generalmente requiere conocimiento adicional del estimando Q , y muchas veces es difícil encontrar una cota adecuada.

En cambio, para estimadores no sesgados se cumple que $\mathbb{E}[F] = Q$, por lo que el error cuadrático medio es igual a la varianza:

$$\text{MSE}(F) = \text{Var}[F]$$

Esto facilita mucho la estimación del error, ya que basta con tomar varias muestras independientes. Siendo Y_1, Y_2, \dots, Y_N muestras independientes de un estimador no sesgado Y , y definiendo:

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

entonces la siguiente expresión:

$$\hat{\text{Var}}[F_N] = \frac{1}{N(N-1)} \sum_{i=1}^N (Y_i - F_N)^2$$

es un estimador no sesgado de la varianza $\text{Var}[F_N]$ (véase Kalos y Whitlock, 1986). Por lo tanto, las estimaciones de error son fáciles de obtener en el caso de estimadores no sesgados.

Nótese que tomando muchas muestras independientes, el error de un estimador no sesgado puede hacerse tan pequeño como se desee, ya que:

$$\text{Var}[F_N] = \frac{\text{Var}[F]}{N}$$

Sin embargo, esto también incrementa el tiempo de ejecución en un factor de N . Idealmente, querríamos encontrar estimadores cuya varianza y tiempo de ejecución sean ambos pequeños. Este compromiso se resume en la *eficiencia* de un estimador Monte Carlo:

$$\varepsilon_F = \frac{1}{\text{Var}[F] \cdot T_F} \tag{2.19}$$

donde T_F es el tiempo requerido para evaluar F . Así, mientras mayor sea la eficiencia, menor será la varianza obtenida para un tiempo de ejecución fijo.