Integración de Montecarlo

Dr. Misael Erikson Maguiña Palma

Teorema del Estadístico Inconsciente

Consideremos una variable aleatoria X y $g:\mathbb{R}\to\mathbb{R}$ una función tal que $g(X)$ sea una variable aleatoria con esperanza finita. Entonces:

$$\mathrm{E}[g(X)] = \int_{\mathbb{R}} g(x)\, dF_X(x)$$

Al estar integrando con respecto a la medida inducida por la función de distribución de X, es decir F_X, nos permite generalizar el resultado para variables aleatorias discretas, continuas o mixtas.

Caso particular 1: Cuando X es v.a. *absolutamente discreta* entonces el resultado anterior puede verse como

$$\mathrm{E}[g(X)] = \sum_{\forall x} g(x) f_X(x)$$

Caso particular 2: Cuando X es v.a. *absolutamente continua* entonces el resultado anterior puede verse como

$$\mathrm{E}[g(X)] = \int_{\mathbb{R}} g(x) f_X(x)\, dx$$

Ejemplos aplicados hay muchísimos, como el cálculo de momentos de variables aleatorias.

Ley de los Grandes Números

Consideremos una muestra $X, X_1, X_2, \dots$ de variables aleatorias independientes e idénticamente distribuidas con media $\mathrm{E}[X] = \mu$. Entonces:

$$\overline{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i \to \mu$$

Existen dos versiones principales de la **Ley de los Grandes Números**, la débil y la fuerte. Su principal diferencia está en su tipo de convergencia, en **probabilidad** y casi segura respectivamente.

Ejemplo: Si tomamos una muestra $X, X_1, X_2, \dots, X_n$ de v.a.i.i.d. con $X \sim Bernoulli(p) \Rightarrow \mu = p$, donde pensamos a este parámetro como la **probabilidad de éxito**, que es precisamente cuando la variable aleatoria toma el valor 1. Recordemos que cada observación de este tipo toma valores únicamente en $\{0,1\}$. Entonces, invocando la Ley de los Grandes Números tendremos que:

$$\overline{X}_n = \frac{\#\acute{E}xitos}{\#Ensayos} \to p$$

La cual es una propiedad sumamente utilizada en estadística básica.

Derivado de este resultado, si contamos con una muestra homogénea materializada en $x_1, x_2, \dots, x_n$ entonces el promedio de estos datos convergerá a la media teórica de los mismos.

Sean X una variable aleatoria tal que $X \sim Unif(a,b)$ y g una función integrable en el intervalo (a,b) tal que g(sea variable aleatoria. Invocando el Teorema del Estadístico Inconsciente podemos obtener el siguiente resultado:

$$\mathrm{E}[\mathrm{g}(X)] = \int_{\mathbb{R}} \mathrm{g}(x) f_X(x) dx = \int_a^b \frac{\mathrm{g}(x)}{b-a} dx = \frac{1}{b-a} \int_a^b \mathrm{g}(x) dx$$

De donde se deduce la siguiente igualdad:

$$(b-a)\mathrm{E}[\mathrm{g}(X)] = \int_a^b \mathrm{g}(x) dx$$

Luego, si tomamos una muestra aleatoria $X, X_1, X_2, \dots, X_n$ de v.a.i.i.d. podemos ahora transformar cada una de estas con la función g, de tal manera que tendremos una muestra $\mathrm{g}(X), \mathrm{g}(X_1)$, $\mathrm{g}(X_2)$, …, $\mathrm{g}(X_n)$ y así podremos invocar la Ley de los Grandes Números sobre la transformación $\mathrm{g}(X)$ obteniendo el siguiente resultado:

$$\overline{\mathrm{g}(X)}_n \coloneqq \frac{1}{n} \sum_{i=1}^{n} \mathrm{g}(X_i) \to \mathrm{E}[\mathrm{g}(X)]$$

Y por lo tanto:

$$(b-a)\, \overline{\mathrm{g}(X)}_n \to \int_a^b \mathrm{g}(x) dx$$

Hemos encontrado una forma de aproximarnos una integral a partir de simulaciones de $X_i \sim Unif(a,b)$.

En resumen, si nosotros deseáramos integrar por ejemplo, una función g continua en el intervalo (a, b) podríamos seguir los siguientes pasos:

1. Simulamos una muestra de tamaño n de $X \sim Unif(a, b)$.
2. Transformamos cada observación de esta muestra como $g(X)$.
3. Calculamos la estadística $(b - a)\,\overline{g(X)}_n$.

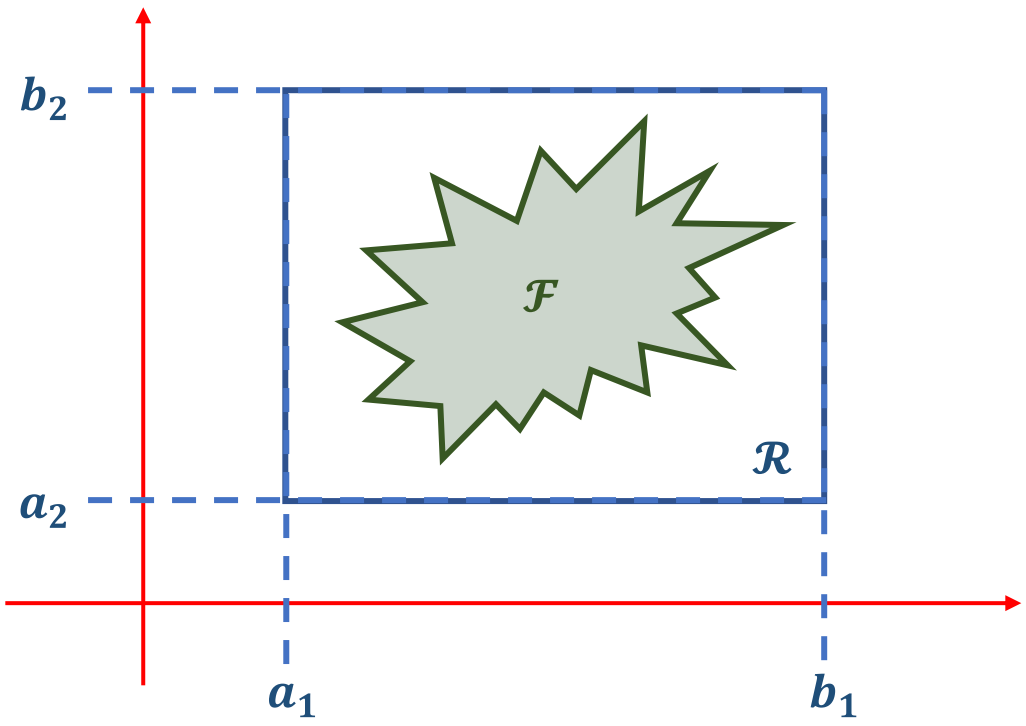
Y de esta manera tendemos una aproximación a la integral que nosotros buscamos:

$$\int_a^b g(x)dx$$

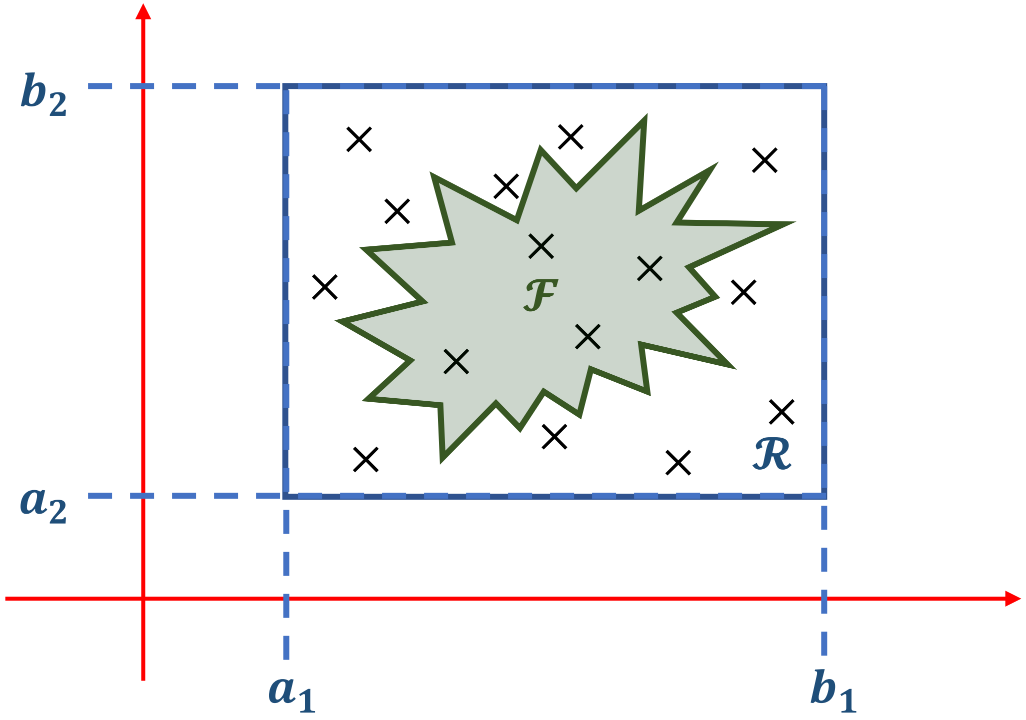
Todo esto suena bastante bien y de hecho, calcular el área por ejemplo de un rectángulo es bastante simple. embargo, la verdadera pregunta es **¿cómo podemos calcular $\boldsymbol{P}[\boldsymbol{V} \in \boldsymbol{\mathcal{F}}]$?**.

La respuesta a esto es que NO vamos a calcular $\boldsymbol{P}[\boldsymbol{V} \in \boldsymbol{\mathcal{F}}]$, la vamos a estimar.
¿Entonces cómo haremos esto?

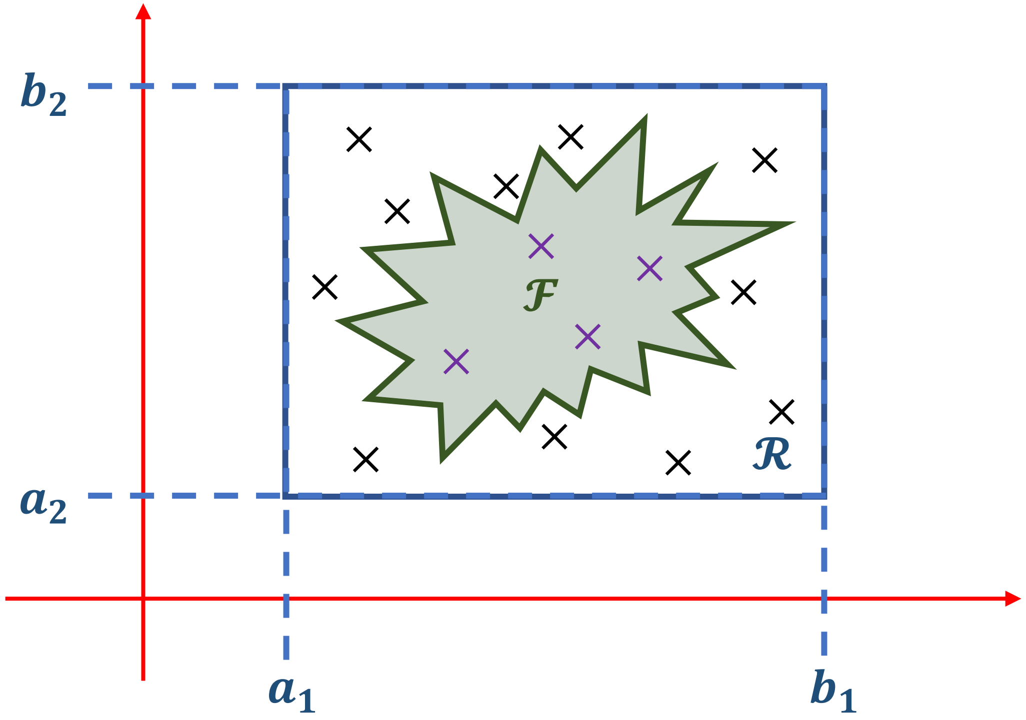
Primero, vamos a imaginar una *figura* $\boldsymbol{\mathcal{F}}$ arbitraria en el plano que esté *"encerrada"* en un *rectángulo* $\boldsymbol{\mathcal{R}}$:



Ahora, similar a cuando arrojamos *dardos* lanzaremos en todo el *rectángulo* $\mathcal{R}$ observaciones de la variable aleatoria $\boldsymbol{V}$. Estas deberían verse de la siguiente manera:



Luego, observemos del total de *dardos (ensayos)* cuántos caen dentro de la *figura* $\mathcal{F}$ a esta cantidad la podemos pensar como el número de éxitos.



De tal manera que podremos estimar la probabilidad que necesitamos como: $\boldsymbol{P}[\boldsymbol{V} \in \boldsymbol{\mathcal{F}}] \approx \frac{\#\text{éxitos}}{\#ensayos}$.

Nota: A mayor cantidad de ensayos obtendremos una mejor aproximación a la probabilidad real.

Por lo tanto, podemos aproximar el área de $\boldsymbol{\mathcal{F}}$ simplemente calculando:

$$\acute{A}rea(\boldsymbol{\mathcal{R}})\left(\frac{\#\text{éxitos}}{\#ensayos}\right) \approx \acute{A}rea(\boldsymbol{\mathcal{F}})$$

En resumen, para hacer esta metodología debemos:

1. Simular una muestra de $V \sim Unif(\boldsymbol{\mathcal{R}})$.
2. Contar el número de éxitos, es decir el total de observaciones $v_i \in \boldsymbol{\mathcal{F}}$.
3. Calcular $\acute{A}rea(\boldsymbol{\mathcal{R}})\left(\frac{\#\text{éxitos}}{\#ensayos}\right)$.

Y con esto tendremos una aproximación a $\acute{A}rea(\boldsymbol{\mathcal{F}})$ esta es una forma de realizar **integración Monte Carlo**.

Hagamos cálculos con…

Realice la siguiente integral mediante el método montecarlo

Si deseamos computar

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

$$\int_a^b g(x)\,dx$$

Sustituyendo $u = \frac{x-a}{b-a}$ cuando $x \longrightarrow a$ entonces $u \longrightarrow 0$; cuando $x \longrightarrow b$ entonces $u \longrightarrow 1$.

Derivando tenemos que $du = \frac{dx}{b-a}$, por lo cual $x = (b-a)\,u + a$, y $dx = (b-a)\,du$.

Reemplazando tenemos que

$$\int_0^1 g((b-a)u + a)\,(b-a)\,du = \int_0^1 h(u)\,du$$

donde $h(u) = (b-a)\,g(a + (b-a)u)$

Integración multivariada por aproximación de Monte Carlo

Supogase que g es una función con argumento n-dimensionaly estamos interesados en el computo en el computo de

$$\theta = \int_0^1 \int_0^1 \cdots \int_0^1 g(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \cdots dx_n$$

La clave de la aproximación de Monte Carlo para estimar θ puede expresarse como la siguiente esperanza

$$\theta = E[g(U_1, U_2, \ldots, U_n)]$$

donde $U_1, U_2, \ldots, U_n$ son v.a. uniformes independientes en $(0, 1)$. Por lo tanto, generando k conjuntos independientes, cada uno de n v.a. independientes uniformes $(0, 1)$

$$\begin{array}{cccc} U_{11} & U_{12} & \ldots & U_{1n} \\ U_{21} & U_{22} & \ldots & U_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ U_{k1} & U_{k2} & \ldots & U_{kn} \end{array}$$

Realice la siguiente integral médiante el método montecarlo

$$\int_0^1 \int_0^1 (x_1^2 + x_2^2)\, dx_2\, dx_1$$

Supongamos que $X = [x_1, x_2]$, por lo cual la función a tener en cuenta es la siguiente

Resolviendo la integral de forma analítica se tiene que

$$
\begin{aligned}
\theta &= \int_0^1 \int_0^1 (x_1^2 + x_2^2)\, dx_2\, dx_1 \\
&= \int_0^1 x_1^2 x_2 \Big|_{x_2=0}^{x_2=1} + \frac{1}{3} x_2^3 \Big|_{x_2=0}^{x_2=1} dx_1 \\
&= \int_0^1 x_1^2 + \frac{1}{3}\, dx_1 \\
&= \frac{1}{3} x_1^3 \Big|_{x_1=0}^{x_1=1} + \frac{1}{3} x_1 \Big|_{x_1=0}^{x_1=1} \\
&= \frac{1}{3} + \frac{1}{3} = \frac{2}{3} = 0.666667
\end{aligned}
$$

Emplee la simulación para aproximar las siguientes integrales. Compare su estimación con la respuesta exacta, si ésta se conoce

1. $\int_0^1 \exp(\exp(x))\,dx.$
2. $\int_0^1 (1-x^2)^{3/2}\,dx.$
3. $\int_{-2}^2 \exp\left(x+x^2\right)dx.$
4. $\int_0^\infty x(1+x^2)^{-2}\,dx.$
5. $\int_{-\infty}^\infty \exp(-x^2)\,dx.$
6. $\int_0^1\int_0^1 \exp(-(x^2+y^2))\,dy\,dx.$
7. $\int_0^\infty\int_0^x \exp(-(x+y))\,dy\,dx.$