Consideriamo un cerchio contenuto in un quadrato. Se distribuiamo a caso uniformemente N_{tot} punti all'interno del quadrato, ci aspettiamo che il rapporto N_{in}/N_{tot} tra il numero N_{in} di punti che cadono dentro il cerchio e N_{tot} sia una stima del rapporto tra l'area $S_\circ$ del cerchio e l'area $S_\diamond$ del quadrato. Ci aspettiamo anche che questa stima sia tanto più accurata quanto maggiore è N_{tot}.

Questo esempio del cerchio e del quadrato introduce in modo intuitivo il paradigma dell'integrazione Monte Carlo: definiamo $I = S_\circ/S_\diamond$ e riscriviamolo come

$$I = \frac{\int d\mathbf{r} f(\mathbf{r}) g(\mathbf{r})}{\int d\mathbf{r} g(\mathbf{r})}, \tag{1}$$

dove il dominio di integrazione è il quadrato, $g(\mathbf{r}) = 1$ e

$$f(\mathbf{r}) = \begin{cases} 1 & \text{se} \quad \mathbf{r} \quad \text{è dentro il cerchio} \\ 0 & \text{se} \quad \mathbf{r} \quad \text{è fuori;} \end{cases} \tag{2}$$

è immediato verificare che $N_{in}/N_{tot} = \frac{1}{N_{tot}} \sum_{i=1}^{N_{tot}} f(\mathbf{r_i})$, dove $\{\mathbf{r_i}\}$ sono i punti che abbiamo distribuito nel quadrato.

In generale, un integrale $I = \int g(x)dx$ si può calcolare con il metodo Monte Carlo fattorizzando l'integrando come $g(x) = f(x)\pi(x)$, dove $\pi(x)$ è una distribuzione di probabilità, cioè $\pi(x) \geq 0$ e $\int \pi(x)dx = 1$. Qui x rappresenta un punto del dominio di integrazione, definito in uno spazio con un numero arbitrario di dimensioni. La procedura consiste nel campionare un insieme di N punti $\{x_i\}$ dalla distribuzione $\pi(x)$, e nel valutare la media $1/N \sum_i f(x_i)$. Nell'esempio del cerchio e del quadrato, $\pi(x) = g(x)/\int g(x)dx$, e questa distribuzione si può campionare direttamente usando un generatore di numeri pseudocasuali uniformemente distribuiti tra 0 e 1, disponibile sul computer.

Una quantità fondamentale per l'integrazione stocastica è la varianza, o scarto quadratico medio, di f sulla distribuzione π,

$$\sigma^2 = \int (f(x) - I)^2 \pi(x) dx. \tag{3}$$

Si può mostrare (a lezione è stato fatto il caso $N = 2$) che $\sigma_N^2 = \sigma^2/N$, dove σ_N^2 è la varianza della variabile aleatoria $1/N \sum_i f(x_i)$. Poichè quest'ultima è la nostra stima dell'integrale I, questo significa che l'incertezza statistica (la standard deviation, $\sqrt{\sigma_N^2}$) sul calcolo di I diminuisce come $1/\sqrt{N}$, indipendentemente dalla dimensionalità di x. Per esempio, per ridurre l'errore statistico di un fattore 2 ci vuole un numero di punti N quattro volte maggiore. Questo va confrontato con i metodi di quadratura, per i quali per ridurre l'errore di discretizzazione di un fattore 2 ci vorrebbe un numero di valutazioni dell'integrando maggiore di un fattore α^d, dove d è il numero di dimensioni e α un numero che dipende dall'algoritmo di integrazione, ma è sempre maggiore di 1. Questo fa sì che per d abbastanza grande non ci siano alternative pratiche al metodo Monte Carlo.

Se la varianza è finita vale il teorema del limite centrale, secondo il quale la distribuzione di $1/N \sum_i f(x_i)$ tende a una gaussiana nel limite $N \to \infty$. Questo significa che, se N è abbastanza grande, la probabilità che la stima di I sia fuori dal valore esatto di una standard deviation σ_N è circa 1/3, di $2\sigma_N$ circa 1/20, di $3\sigma_N$ circa 3/1000, e così via.

Il numero di casi come quello del cerchio e del quadrato in cui la distribuzione di probabilità si può campionare direttamente è molto limitato. Fortunatamente esistono metodi per campionare distribuzioni arbitrariamente complicate, il più utilizzato dei quali è l'algoritmo di Metropolis.