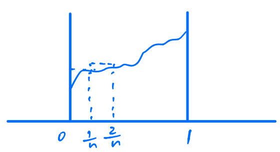
Monte Carlo Method

- visualization
- code reusability
- generality
- "dimensionally insensitive"



$\alpha = \int_0^1 h(x)\,dx$

$\alpha_n = \sum_{i=1}^{n} \frac{1}{n} h(\frac{i-1}{n}) \to \alpha$ as $n \to \infty$

"rectangular integration rule"

$$\alpha_n - \alpha = \sum_{i=1}^{n} \left(\frac{1}{n} h(\frac{i-1}{n}) - \int_{(i-1)/n}^{i/n} h(x)\,dx \right)$$

$$= \sum_{i=1}^{n} \int_{\frac{i-1}{n}}^{\frac{i}{n}} \underbrace{[h(\tfrac{i-1}{n}) - h(x)]}_{h'(\xi_i)(\frac{i-1}{n} - x)} dx$$

$\leftarrow$ Tylor

$\underbrace{h'(\xi_i)}_{\le m \text{ (const)}} \underbrace{(\tfrac{i-1}{n} - x)}_{\le \frac{1}{n}}$

$$= O(\tfrac{1}{n})$$

Suppose h is polynomial of degree k over $[0,1]$,

constant : evaluate h at 1 point
deg = 1 : . . . at 2 points
deg = k . . . at k+1 points

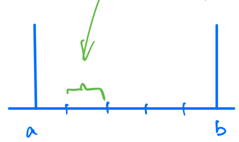
$$\sum_{j=0} w_j\, h(\tfrac{j}{k}) = \int_0^1 h(x)\,dx$$

$$h(x) = x^l, \quad 0 \le l \le k$$

$$\sum_{0}^{k} w_j (\tfrac{j}{k})^l = \int_0^1 x^l\,dx, \quad 0 \le l \le k$$

uniquely determined from the linear system

apply integration rule to each interval



$\Rightarrow$ integration rule that converges faster

More improvements:

decision variables

$$\sum_{j=0}^{k} w_j \, h(z_j) = \int_0^1 h(x)\,dx$$

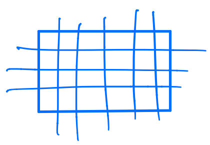
Guassian Quadrature Rule

z_j: orthogonal polynomials

poly of degree k $\Rightarrow$ degree $2k+1$

These high accuracy methods require lots of smoothness in h

$d = 2$



$$\alpha = \int_{[0,1]^2} h(x,y)\,dxy$$

In d-dimensions:

$$\alpha = \int_{[0,1]^d} h(\vec{x})\,d\vec{x}$$

$$\alpha_n = \sum_{i=1}^{n} \frac{1}{n} h(\vec{x})$$

$$\alpha_n - \alpha = \sum_{i=1}^{n} \int_{H_j} \underbrace{[h(\vec{x}_j) - h(\vec{x})]}_{\nabla h(\xi_j)(\vec{x}_j - \vec{x})} \, d\vec{x}$$

$\nabla h \le m$

$$\mathrm{Vol}(H_j) = \frac{1}{n}$$

$$\mathrm{side}(H_j) = s = n^{-1/d}$$

$$\Rightarrow \quad |\alpha_n - \alpha| = O(n^{-1/d})$$

e.g. $\varepsilon = 10^{-2}$

$n^{-1/d} = \varepsilon$

$n \approx \varepsilon^{-d} = \left(\frac{1}{\varepsilon}\right)^d = 10^{2d}$ $\leftarrow$ too big!

"curse of dimensionality"

- $h \in C^k$ over $[0,1]^d$; $\| D^j h \| \le c$ (derivatives)

- $\sum_{j=1}^{n} w_j h(z_j) = \alpha_n$

$$\Rightarrow \quad |\alpha_n - \alpha| = O(n^{-k/d})$$

Monte Carlo

- Convergence rate dimensionality insensitive
- Smoothness not needed

$$\int_{[0,1]^d} h(\vec{x})\, d\vec{x} = \int_0^1 \cdots \int_0^1 h(x_1, \cdots, x_d)\, dx_1 \cdots dx_d$$

$$= E\, h(u_1, \cdots u_d)$$

$$E\, h(\vec{x}) = \int_{[0,1]^d} h(\vec{x})\, f(\vec{x})\, d\vec{x}$$

$\alpha = EX$

Method :

$X_1, X_2, \cdots, X_n$ i.i.d. draws from the dist of X

$\alpha_n = \bar{X}_n$

- $\bar{X}_n \xrightarrow{P} \alpha$
- $\bar{X}_n \xrightarrow{a.s.} \alpha$
- If $\operatorname{var} X < \infty$, CLT holds:

$$n^{\frac{1}{2}} (\alpha_n - \alpha) \Rightarrow \sigma N(0,1)$$

std of X

$$\alpha_n \overset{D}{\approx} \alpha + \frac{\sigma}{\sqrt{n}} N(0,1)$$

① rate of convergence is $n^{-\frac{1}{2}}$

independent of d

② rate of convergence depends on X

through only 1 parameter : σ

③ error looks approx. normal

---

① $\alpha = P(A)$, $A \in \Omega$ ← very high dim

$I_1, I_2, \cdots, I_n$

$\alpha_n = \bar{I}_n$

$$P(|\alpha_n - \alpha| > \varepsilon) \le \frac{\mathrm{Var}\,\alpha_n}{\varepsilon^2}$$

$$= \frac{P(A)(1-P(A))}{n\varepsilon^2}$$

$$\le \frac{1}{4n\varepsilon^2}$$ insensitive to Ω

② $\alpha = P(A)$

$I_1, I_2, \cdots, I_n$ under P

Change of measure :

$$P(S_n > na) = E_{\theta(a)} I(S_n > na) \exp(\quad)$$

$$P(A) = \int_\Omega I(A,w) L(w) \, Q(dw)$$

$\{w\}$ "likelihood ratio"

$$= E_\theta I(A) L$$

$$P(dw) = L(w)\, Q(dw)$$

$$L(w) = \frac{P(dw)}{Q(dw)}$$ "likelihood ratio"

$$Z = I(A)\, L$$

$Z_1, Z_2, \cdots, Z_n$ under Q

$$\bar{Z}_n \to P(A)$$

Should I use $\bar{Z}_n$ or $\bar{I}_n$?

$$\bar{Z}_n \overset{D}{\approx} \alpha + \frac{\sigma_Z}{\sqrt{n}} N(0,1)$$

$$\bar{I}_n \overset{D}{\approx} \alpha + \sqrt{\frac{P(A)(1-P(A))}{n}}\, N(0,1)$$

Use Z if

$$\sigma_Z^2 \ll P(A)(1-P(A))$$

- Design of algorithms
- "variance reduction"
- "importance sampling"

  e.g. $P(S > na)$

  sample where it matters