Numerical integration

1) MC integration, for high dim.

2) Deterministic algo., for low dim.

MC integration

- Consider a 1D integral

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

From calculus

- Definition of <u>average function value</u> on $x \in [a,b]$

$$\bar{f} = \frac{1}{b-a} \int_a^b f(x)\, dx \implies I = (b-a)\bar{f}$$

- Now consider <u>expectation value</u> for f given that $x \sim U(a,b)$
  that is $p(x) = \begin{cases} \frac{1}{b-a} & x \in [a,b] \\ 0 & \text{else} \end{cases}$

From prob. theory

$$E[f] = \langle f \rangle = \mu_f = \int_a^b f(x)\, p(x)\, dx = \frac{1}{b-a}\int_a^b f(x)\, dx = \bar{f}$$

→ Average f value = expected $f(x)$ when $x \sim$ uniform dist.

$$\implies \boxed{I = (b-a)\langle f \rangle}$$ ← Basis for MC int (1D)

- We can estimate $\langle f \rangle$ using <u>sample mean</u> $\bar{\mu}_f$
  - Draw N samples $x \sim U(a,b)$

$$\langle f \rangle \approx \bar{\mu}_f = \frac{1}{N}\sum_{\text{samples}} f(x_i)$$

- Estimate of integral $I \approx \hat{I} = (b-a)\bar{\mu}_f = (b-a)\frac{1}{N}\sum_{\text{samples}} f(x_i)$

- Error estimation: We want variance in $\hat{I}$: $\sigma_{\hat{I}}^2$

Remember that $\hat{I}$ is a rand. var., since $\bar{\mu}$ is.

- Start from sample variance for f_i samples

$$\sigma_f^2 = \frac{1}{N-1} \sum_{i=1}^{N} (f_i - \bar{\mu}_f)^2$$

samples

[Unbiased estimator for the true variance $E[\sigma_f^2] = \sigma^2$]

- Variance of the mean $\sigma_{\bar{\mu}_f}^2$ is then

$$\left[\sigma_{\bar{\mu}_f}^2 = \frac{\sigma_f^2}{N} \right]$$

[Result from the Central Limit Theorem]

- Since $\hat{I} = (b-a)\bar{\mu}_f$ $\Leftarrow$

we have

$$\sigma_{\hat{I}}^2 = (b-a)^2 \sigma_{\bar{\mu}_f}^2$$

$$\sigma_{\hat{I}}^2 = (b-a)^2 \frac{\sigma_f^2}{N}$$

$$\boxed{\sigma_{\hat{I}} = (b-a)\frac{\sigma_f}{\sqrt{N}}}$$

[Let $Y(x) = cx$, c const

$$\mathrm{Var}[Y] = E[Y^2] - E[Y]^2$$

$$= E[c^2x^2] - E[cx]^2$$

$$= c^2\left(E[x^2] - E[x]^2\right)$$

$$\mathrm{Var}(Y) = c^2\,\mathrm{Var}(x)$$
]

$$\boxed{I \approx \hat{I} \pm \sigma_{\hat{I}} = \frac{(b-a)}{N}\sum_{i=1}^{N} f_i \pm (b-a)\frac{\sigma_f}{\sqrt{N}}}$$

- Now go to d dimensions:

$$I = \int_{a_1}^{b_1} dx_1 \int_{a_2}^{b_2} dx_2 \dots \int_{a_d}^{b_d} dx_d \, f(x_1, x_2, \dots, x_d) = \int_V f(\bar{x}) d\bar{x}$$

- Nothing changes in derivation, except $(b-a) \rightarrow V$

- Basic MC integration:
  - Sample N points $\bar{x}_i$ uniformly on V
  - Evaluate function $f(\bar{x}_i) \equiv f_i$

$$\boxed{I \approx \frac{V}{N} \sum_{i=1}^{N} f_i}$$

  - with uncertainty estimate:

$$\sigma_f^2 = \frac{1}{N-1} \sum_{i=1}^{N} (f_i - \bar{\mu}_f)^2 \quad , \quad \sigma_f = \sqrt{\sigma_f^2}$$

$$\Rightarrow \boxed{I \approx \frac{V}{N} \sum_{i=1}^{N} f_i \pm \frac{V \sigma_f}{\sqrt{N}}}$$

- Observations:

  1) Error scales as $\mathcal{O}(\frac{1}{\sqrt{N}}) = \mathcal{O}(\frac{1}{N^{1/2}})$ with number of samples N, independent of number of dimensions d !!

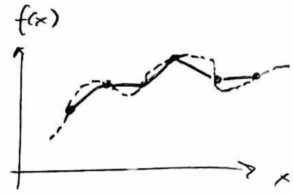
  2) If the variance in f-values is small (small σ_f), the error on I is small.
  [Makes sense: small σ_f $\hookrightarrow$ slowly varying function $f(x)$
  $\hookrightarrow$ need few samples to estimate $\langle f \rangle$ fairly well]

  Extreme limit: Flat function $\rightarrow$ Need 1 sample to integrate. $(\sigma_f = 0)$

In all dimensions, doubling number of samples will reduce error by factor $\frac{1}{\sqrt{2}}$

When to use MC integration?

- Example: compare to trapezoidal rule
  (linear interpolation between grid points)



  - Step size h in each dimension
  - Error: $\mathcal{O}(h^2)$ (Intuition: with linear interpolation, h^2 is the leading Taylor exp. term we ignore.)
  - Number of points $N = N_{x_1} N_{x_2} N_{x_3} \dots N_{x_d} \sim \left(\frac{1}{h}\right)^d = \frac{1}{h^d}$

$$\Rightarrow \quad h \sim \frac{1}{N^{1/d}}$$

  - Error scaling with number of points:

$$\mathcal{O}(h^2) \sim \mathcal{O}\left(\frac{1}{N^{2/d}}\right)$$

[we want the error to decrease quickly as func. of N, so want large denominator]

- MC error scaling: $\mathcal{O}\left(\frac{1}{N^{1/2}}\right)$

$\Rightarrow$ MC integration preferable over trapezoidal rule when

$$N^{1/2} > N^{2/d}$$

$$\frac{1}{2} > \frac{2}{d}$$

$$d > 2 \cdot 2$$

comes from $\mathcal{O}(h^2)$ error for trapezoidal rule

$$d > 4$$

- In general for $\mathcal{O}(h^k)$ method: MC int. preferred when

$$\boxed{d > 2k}$$

e.g. $d > 8$ for Simpson's rule ($k=4$)

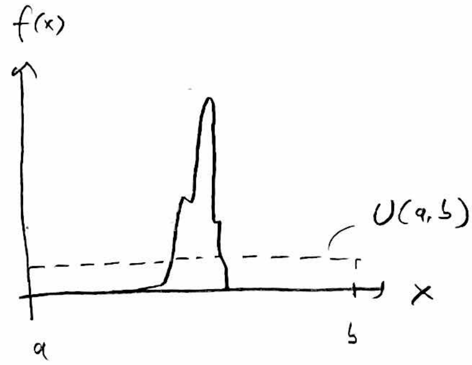
Importance sampling

start here

(Then esc @.)

↳ (o) Vanilla MC integration will be ineff. in this case

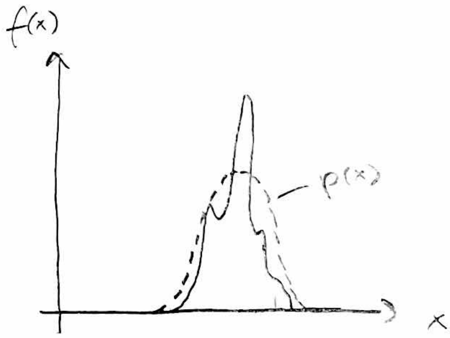
[Most samples will not contribute to I est.]



- Can we use a different pdf than the uniform?

$$I = \int_a^b f(x)dx = \int_a^b \frac{f(x)}{p(x)} \underset{\text{pdf}}{p(x)}\, dx \equiv \int_a^b q(x)\, p(x)\, dx = \langle q \rangle_{p(x)}$$

- So $I \approx \frac{(b-a)}{N} \sum_i q_i$ (if $x_i \sim p(x)$)



- Since $q(x) = \frac{f(x)}{p(x)}$ is a flatter function than $f(x)$,

  the sample variance is smaller for q_i than for f_i

So $$\boxed{I \approx \frac{V}{N} \sum_{i=1}^{N} q_i \pm \frac{V \sigma_q}{\sqrt{N}}}$$ (Took $(b-a) \rightarrow V$ for generality)

with $q(x) = \frac{f(x)}{p(x)}$ will be a better estimate of I if the sampling distr. $p(x)$ resembles $f(x)$, i.e. such that $q(x)$ is as flat as possible.

- Can be difficult to find suitable pdf. $p(x)$!

End here