Bonus Lecture: Numerical Integration

C.Conlon

Fall 2020

Numerical Integration

- We are interested in lots of problems that require computing difficult integrals (e.g.: evaluating expectations).
- Often the problem looks like this:

$$I = \int_a^b h(x) d\,x$$

$$E_{f|\theta}[g(x)] = \int_a^b g(x) f(x|\theta) d\,x$$

- Either just integrating $h(x)$ from $[a, b]$ or computing the expectation of $g(x)$ over some density $f(x)$.

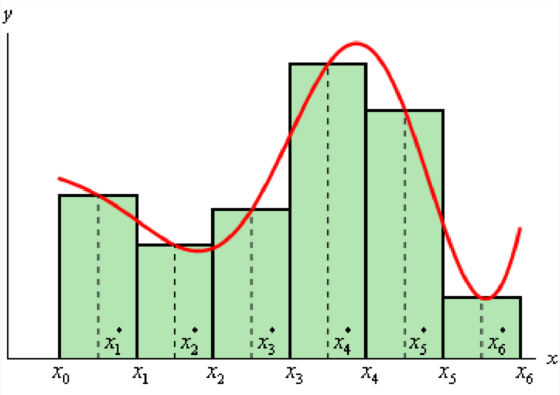
Integration Rules

- Our goal is to construct an integration rule to approximate the function:

$$\int_a^b f(x)dx \approx \sum_{s=1}^{S} w_s \cdot f(x_s)$$

- Rules consist of nodes x_s and weights w_s
- This is probably how you learned integration in high school.

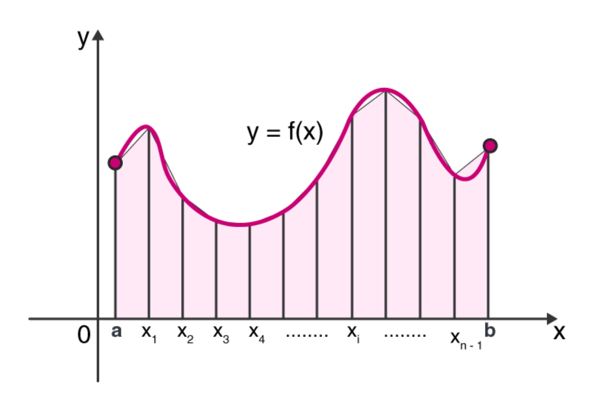
The Midpoint Rule



$$I = \frac{b-a}{S} \cdot \sum_{s=1}^{S} f(x_s^*) \quad \text{with } x_s^* = \frac{x_{s+1} + x_s}{2}$$

We are fitting piecewise constants and adding up rectangles

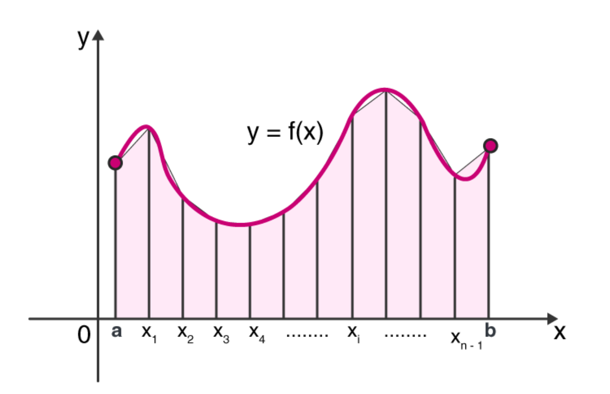
The Trapezoid Rule



$$I = \frac{b-a}{2S} \cdot [f(x_0) + 2 \cdot (f(x_1)+, \ldots, f(x_{S-1})) + f(x_S)]$$

(1) Fit a (piecewise) line from $f(x_s), f(x_{s+1})$; (2) Evaluate trapezoid area analytically; (3) Add up trapezoids.

Simpson's Rule



$$I = \frac{b-a}{3S} \cdot [f(x_0) + 4(f(x_1) + f(x_3) + \ldots) + 2(f(x_2) + f(x_4) + \ldots) + f(x_S)]$$

(1) Fit a (piecewise) quadratic through $f(x_s), f(x_{s+1}), f(x_{s+2})$; (2) Analytically integrate the quadratic; (3) Add up areas.

Choosing Integration Rules

Not a lot to choose here

- Choose the number of points S or interval width $h = \frac{b-a}{S}$.
- Points are generally equally spaced

How accurate is it?

- Bounds analysis is possible (based on series approximations)
- For Simpson's Rule:

$$\frac{h^4}{180}(b-a)\max_{\xi\in[a,b]}\left|f^{(4)}(\xi)\right|$$

- Quadratic approximation does poorly where $f^{(4)}$ is large (and gets up to order 3 polynomials exact).

Newton-Cotes Formulas

Summary of formulas for a single interval:

Degree n	Step size h	Common name	Formula	Error term
1	$b-a$	Trapezoid rule	$\frac{h}{2}(f_0+f_1)$	$-\frac{1}{12}h^3f^{(2)}(\xi)$
2	$\frac{b-a}{2}$	Simpson's rule	$\frac{h}{3}(f_0+4f_1+f_2)$	$-\frac{1}{90}h^5f^{(4)}(\xi)$
3	$\frac{b-a}{3}$	Simpson's 3/8 rule	$\frac{3h}{8}(f_0+3f_1+3f_2+f_3)$	$-\frac{3}{80}h^5f^{(4)}(\xi)$
4	$\frac{b-a}{4}$	Boole's rule	$\frac{2h}{45}(7f_0+32f_1+12f_2+32f_3+7f_4)$	$-\frac{8}{945}h^7f^{(6)}(\xi)$

`https://en.wikipedia.org/wiki/Newton-Cotes_formulas`

Can we do better?

Can use adaptive rules (built-in many software routines)

- Divide up $[a, b]$ evenly into sub-intervals
- Some are relatively flat or well-approximated, some are not.
- Use more points in the sub-intervals that are the most difficult to approximate
- Don't waste points in intervals that are easy to approximate.

Drawbacks

- Points used in approximation are $f(\cdot)$ dependent
- Points used in approximation x_s may not be the same for $f(x|\theta)$ at all values of parameters θ.

Gaussian Quadrature: Same Basic Idea

- Choose a degree of polynomial approximation for $f(x)$ on $[a, b]$.
- "Fit" the polynomial by evaluating $f(x_s)$ at various values of x_s
- Integrate the polynomial analytically by adjusting coefficients on $f(x_s)$.

In practice, much easier: get a pre-specified set of weights and nodes (w_s, x_s).

Gaussian Quadrature: But a little different

- $[1, x, x^2, x^3, \ldots]$ is not the only polynomial basis.
- For example Chebyshev Basis (of first kind) is orthogonal $[1, x, 2x^2 - 1, 4x^3 - 3x, \ldots]$
- Concept remains the same:
  1. Approximate $f(x)$ with a polynomial basis
  2. Integrate the polynomial exactly
  3. In practice we just get a pre-determined set of points and weights (x_s, w_s) for a given polynomial order.
- Rules have some additional properties
  - What interval to integrate over $[-1, 1]$ or $[0, \infty)$ or $(-\infty, +\infty)$.
  - Can we exploit properties of $f(x)$: is it proportional to $\frac{1}{\sqrt{1-x^2}}$ or e^{-x} or e^{-x^2}?
- May need to do change of variables on $f(x) \rightarrow f(v)$ to better satisfy conditions above.

Gaussian Quadrature

Formulas look like:

$$\int_a^b f(x)dx \approx \sum_{s=1}^{S} w_s f(x_s)$$

for some quadrature nodes $x_s \in [a, b]$ and weights w_s.

- Let $\mathcal{F}_k$ be the space of degree k polynomials
- Quadrature formulas are exact of degree k if it correctly integrates each function in $\mathcal{F}_k$
- Gaussian quadrature formulas use S points and are exact of degree $2s - 1$.

Approximation Error

$$\int_a^b w(x)f(x)dx - \sum_{s=1}^{S} w_s f(x_s) = \frac{f^{(2S)}(\xi)}{(2S)!}(p_S, p_S)$$

Gaussian Quadrature

Legendre Domain: $[-1, 1]$, $w(x) = 1$

Chebyshev Domain: $[-1, 1]$, $w(x) = \frac{1}{\sqrt{1-x^2}}$

Laguerre Domain: $[0, \infty)$, $w(x) = \exp[-x]$ (useful for present value)

Hermite Domain: $(-\infty, \infty)$, $w(x) = \exp[-x^2]$ (useful for normal)

Helpful if function is C^∞ or real-analytic (comports with series expansion).

Alternative: Monte Carlo Integration

$$
\begin{aligned}
E_{f|\theta}[g(x)] &= \int_a^b g(x) f(x|\theta) d\,x \\
&\approx \frac{1}{S} \sum_{s=1}^{S} g(x_s)
\end{aligned}
$$

- Sample $[x_1, \ldots, x_S]$ by drawing from $f(x|\theta)$.
- Weight everything equally $\frac{1}{S}$
- Can't bound errors, but can discuss rate of convergence
- Convergence is slow $\frac{1}{\sqrt{S}}$ but mostly unrelated to curvature of $f(\cdot)$.

Monte Carlo Integration: Change of Variables

Still often want to change variables. Consider $f(x|\theta) \sim N(\mu, \sigma^2)$

- Should we sample from $x_s \sim N(\mu, \sigma)$?
- Might be better to sample $z_s \sim N(0,1)$ and then $x_s = z_s \cdot \sigma + \mu$
- For complicated distributions $v_s \sim U[0,1]$ and then $x_s = \Phi^{-1}(v_s) \cdot \sigma + \mu$

Discuss approximating π.

Example: Gauss Herrmite

Let $Y \sim N(\mu, \sigma^2)$ and apply COV $x = (y - \mu)/\sqrt{2}\sigma$

$$E[f(Y)] = (2\pi\sigma^2)^{-\frac{1}{2}} \int_{-\infty}^{\infty} f(y) \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right] dy$$

$$\int_{-\infty}^{\infty} f(y) \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right] dy = \int_{-\infty}^{\infty} f(\sqrt{2}\sigma x + \mu) e^{-x^2} \sqrt{2}\sigma dx$$

Gives the quadrature formula using Gauss Hermite (x_s, w_s).

$$E[f(Y)] = \frac{1}{\sqrt{\pi}} \sum_{s=1}^{S} w_s f(\sqrt{2}\sigma x_s + \mu)$$

notice that we don't have the e^{-x^2} anymore.

Quasi Monte Carlo

Lots of ways to improve on purely pseudorandom sampling

- Antithetic draws
- Low-discrepancy sequence (Halton, Sobol, Low-discrepancy sequence etc.)
- Think about these as ensuring more even coverage (stratified).
- Koksma-Hlawka inequality gives bounds but is hard to characterize the Hardy-Kruse variation of $f(\cdot)$.

Many of these are built-in routines in Python, R, Matlab, etc.

Higher Dimensional Integration

- In higher dimension we can use product rules of 1-D integrals. $\mathbf{x_s^{(1)}} \times \mathbf{x_s^{(2)}}$ and $\mathbf{w_s^{(1)}} \times \mathbf{w_s^{(2)}}$
- This grows exponentially in dimension D (Curse of Dimensionality)
  - 10 points in dimension one, means 10,000 points in dimension 4.
- Monte Carlo is not cused but slow to converge $\frac{1}{\sqrt{S}}$ vs $\frac{1}{2S!} f^{(2S)}$
- Some monomial rules (Judd), (Skrainka and Judd) aren't cursed
- Sparse Grids show how to combine 1-D rules more efficiently (`www.sparse-grids.de`)