Math 459: Lecture 13

Todd Kuffner

Approximate Bayesian Inference

Exact analytic calculation of posterior quantities often not practical.

Alternatives:

1. asymptotic (large-sample) approximations (last time)
2. analytic integral approximations (coming later)
3. numerical integration (starting this week–MCMC)

Today: when do we need to approximate an integral?

When Must we Approximate Integrals?

Question: Is there a simple, fool-proof way to determine if the integral of a function can be computed in closed form?

One approach: let's consider elementary functions.

Definition

A function built using a finite combination of constant functions, algebraic operations (addition, multiplication, division, raising to integer power, root extractions–fractional power), logarithmic, exponential and algebraic functions and their inverses *under repeated compositions* is called an **elementary function**.

Types of elementary functions:

1. algebraic functions (can be expressed as solution of a polynomial equation): polynomials, rational functions, root extraction
2. *(non-algebraic)* transcendental functions: exponentials, logarithms, power functions, periodic functions (e.g. trigonometric: sine, cosine, etc.)

Example

$$\frac{\sin^{-1}(x^4 - 3)}{\sqrt{\log(6x) + \cos(x^{-2} + 9)}}$$

More about Elementary Functions

The set of elementary functions is **closed** under *arithmetic operations* (addition, subtraction, multiplication, division) and *differentiation*.

However, it is not closed under integration (**Liouville's theorem**, 1830s)

Implication of Liouville's Theorem

The integrals of certain elementary functions cannot themselves be expressed as elementary functions.

References: (i) Brian Conrad's article 'Impossibility theorems for elementary integration' (ii) M. Rosenlicht (1972), 'Integration in Finite Terms', *American Mathematical Monthly* 79(9), 963-972.

More advanced (Galois theory): Kontsevich & Zagier's article 'Periods'

A Non-Elementary Example

The CDF of the standard normal distribution is given by

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt.$$

- you were taught $f(x) = (2\pi)^{-1/2} \exp(-x^2/2)$ is a density, so we must have that $I = \int_{-\infty}^{\infty} f(x)dx = 1$
- <u>but</u> $\int e^{-x^2} dx$ is not an elementary function; there is no closed-form expression
- we *cannot* show that $I = 1$ by computing $\Phi(x)$ as an explicit function of x and then finding $\lim_{x\to\infty} \Phi(x)$

The Gaussian integral $\int e^{-x^2} dx$ is not an elementary function.

Another Example

The step function $\pi(x) = \#\{1 \leq n \leq x | n \text{ is prime}\}$ of a real variable x counts the number of primes up to x.

- The Prime Number Theorem states that (asymptotically) we can approximate $\pi(x)$ by $x/\log(x)$ since

$$\lim_{x\to\infty} \frac{\pi(x)}{x/\log(x)} = 1.$$

Such reasoning doesn't necessarily yield a good approximation.

- e.g. consider x^2 as an approximation to $x^2 + 3x$
- then $\lim_{x\to\infty} \frac{x^2}{x^2+3x} = 1$
- absolute error is $\varepsilon = |(x^2 + 3x) - x^2| = |3x|$ blows up as $x \to \infty$
- only the relative error, $|3x|/|x^2 + 3x|$, tends to zero

Let's find a better asymptotic approximation.

When Gauss was 15, he conjectured that $\lim_{x\to\infty} \frac{\pi(x)}{\mathrm{Li}(x)} = 1$ where $\mathrm{Li}(x)$ is the **logarithmic integral**

$$\mathrm{Li}(x) = \int_2^x \frac{dt}{\log(t)}, \quad x > 2.$$

- the logarithmic integral is not an elementary function
- with the change of variable $x = \log t$, we have

$$\int \frac{dt}{\log t} = \int \frac{e^x}{x} dx$$

$\int (e^x/x)dx$ is not an elementary function.

How to prove such statements?

Liouville's Approach

What Liouville showed: *if* a (meromorphic) function can be integrated in elementary terms, then such an elementary integral *must have* a very special form

Special case: For functions of the form fe^g with f and g rational functions, there is an elementary integrability condition in terms of the solution of a first-order differential equation with a rational function.

- e^{-x^2} has $f = 1$ and $g = -x^2$
- e^{-x}/x has $f = 1/x$ and $g = x$

Definition

A **rational function** is any function that can be written in the form

$$f(x) = \frac{P(x)}{Q(x)}, \quad x \in \{x : Q(x) \neq 0\}$$

where P and Q are both polynomials in x, and Q is not the zero polynomial.

Complex-valued functions

Let $\mathbb{C}$ be the complex numbers. Advantage to using $\mathbb{C}$-valued functions of a real variable x: $f(x) = u(x) + iv(x)$

- all trigonometric and inverse-trigonometric functions can be expressed in terms of exponentials and logarithms
- allows for more general notion of elementary functions
- makes the current problem simpler

Example

The relationship $e^{ix} = \cos(x) + i\sin(x)$, or the formulas

$$\sin(x) = \frac{e^{ix} - e^{-ix}}{2i}, \quad \cos(x) = \frac{e^{ix} + e^{-ix}}{2}$$

Properties of $\mathbb{C}$-valued functions

Let $f(x) = u(x) + iv(x)$.

Continuity f is continuous if u and v are continuous

Differentiability f is differentiable if u and v are differentiable

Analytic f is (complex) analytic if the real and imaginary parts, $u(x)$ and $v(x)$ are *locally* expressible as a convergent Taylor series

Most functions we can easily write down are analytic, including all elementary functions.

Properties of (Complex) Analytic Functions

- f is complex analytic on some region R if it is complex differentiable at every point z_0 in R

$$f'(z_0) = \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

- f is analytic $\Rightarrow$ infinitely differentiable on R
- *(complex) analytic function* a.k.a. holomorphic function

This property is preserved under the usual operations (sums, products, quotients, composition, exponentiation, differentiation, integration, inversion with non-vanishing derivative).

Connection with logs and diff. eq.

Recall: a differential equation specifies the relationship between a function and its derivatives.

If $f(x)$ is analytic and non-vanishing, then f'/f is also analytic.

- choose a point x_0; then the integral

$$(\log f)(x) = \int_{x_0}^{x} \frac{f'(t)}{f(t)} dt$$

  is an analytic function which we call the **logarithm of** f
- this function depends on the choice of x_0 up to an additive constant, but we ignore this
- equivalently we can consider the logarithm of f to be a *solution*, $y = \log f$, to the differential equation

$$y' = f'/f$$

With $x_0 = 1$ and $f(t) = t$, $t \in (0, \infty)$, this is the usual log function. Can add a constant such that $\exp(\log f) = f$.

Meromorphic Functions

A ratio of analytic functions has a singularity when the denominator is zero.

Definition

A **pole** of a function $f(x)$ at a point p is a type of singularity such that as x approaches p, the function approaches infinity.

Definition

A function $f(x)$ which is holomorphic (complex analytic) on an open interval R *except* for a countable set of points corresponding to the poles of $f(x)$ is called a **meromorphic function**.

Every meromorphic function can be expressed as the ratio of two holomorphic (complex analytic) functions.

Example

e^x/x is meromorphic on the real line, as are rational functions, gamma function on complex plane. If f is meromorphic, then e^f and $\log f$ are, too.

Hand-Waving

The set of meromorphic functions on a (non-empty) open interval R is a **field**, and we can define a derivative operator on this field in the usual way.

Definition

If $f_1, \ldots, f_n$ are meromorphic functions, define $\mathbb{C}(f_1, \ldots, f_n)$ to be the set of all meromorphic functions h of the form

$$h = \frac{p(f_1, \ldots, f_n)}{q(f_1, \ldots, f_n)}, \quad q, f \text{ polynomials, } q \neq 0$$

Such a (differential) field $\mathbb{C}(f_1, \ldots, f_n)$, is the setting for Liouville's theorem.

Example

$K = \mathbb{C}(x, \sin x, \cos x) = \mathbb{C}(x, e^{ix})$

- more details: Math 416 (Complex Variables), Math 430 (Modern Algebra), Math 5031 (Algebra I), Abramowitz & Stegun's *Handbook of Mathematical Functions*

Statement of Liouville's theorem

Let f be an elementary function and let K be *any* elementary field containing f. The function f can be integrated in elementary terms if and only if there exist nonzero $c_1, \ldots, c_n \in \mathbb{C}$, nonzero $g_1, \ldots, g_n \in K$ and an element $h \in K$ such that

$$f = \sum c_j \frac{g_j'}{g_j} + h'.$$

This means that $\sum c_j \log(g_j) + h$ is an elementary integral of f.

- If an elementary function has an elementary integral, then the latter is itself an elementary function <u>plus</u> a finite sum of constant multiples of logarithms of elementary functions.

Example

Consider $f = e^{-x^2}$. This lies in the elementary field $K = \mathbb{C}(x, e^{-x^2})$.

- Liouville's theorem says an elementary anti-derivative of f must have the special form $\sum c_j \log g_j + h$ for some $h \in \mathbb{C}(x, e^{-x^2})$ and nonzero $c_j \in \mathbb{C}$ and $g_j \in \mathbb{C}(x, e^{-x^2})$.
- still not obvious how to prove that such h and g_j's do not exist, but we have at least severely constrained the set of elementary functions which may be considered as anti-derivatives of f

Other Non-Elementary Antiderivatives

- $\frac{\sin x}{x}$
- x^x
- $\frac{1}{\log x}$
- $\log(\log x)$
- $\exp(e^x)$
- the integrands of **elliptic integrals**

While no elementary antiderivative exists for these functions, some of the integrals can be expressed using special functions.

Special Functions

Special functions are simply functions which arise with sufficient frequency in mathematics to warrant being given a name.

Example

The indicator function, the sign function, absolute value, Hermite polynomials, Riemann zeta function, step function, beta function, gamma function, etc.

Some special functions are the non-elementary antiderivatives of elementary functions (and hence must be approximated).

Example

Exponential integral: $\mathrm{Ei}(x) = -\int_{-x}^{\infty} \frac{e^{-t}}{t} dt$

Error function: $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$

Risch Algorithm (1968)

The Risch algorithm is a method for deciding whether or not a function has an indefinite integral which is an elementary function, and if so, how to compute it.

- essentially a simplified implementation of Liouville's theorem
- many computer algebra programs use this for symbolic integration
- modern generalizations also use knowledge about (non-elementary) special functions and their derivatives
- `SymPy` (Symbolic Python) by default uses a faster Risch-Norman algorithm, which may fail to find antiderivatives
- some `R` libraries: `rSymPy`, `Ryacas`

Types of Integral Approximations

- asymptotic expansions
- deterministic numerical approximations (Newton-Cotes quadrature, Romberg integration, Gaussian quadrature)
- Monte Carlo integration – simulation-based numerical approximation utilizing randomness

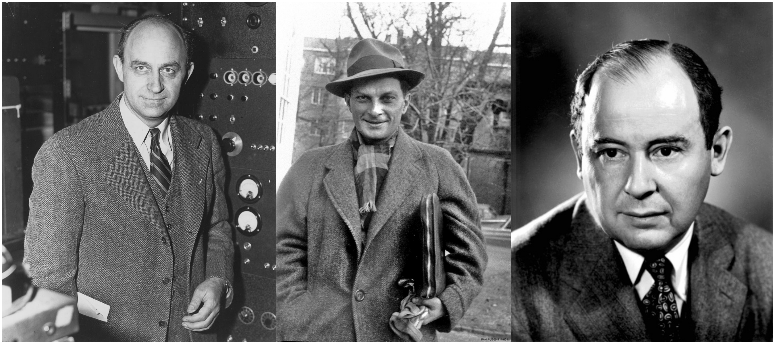
For *single*-dimension integrals, quadrature methods *can* yield convergence of order $O(n^{-2})$, but do not scale well to higher-dimensional integrals.

- Monte Carlo integration is slower with approximation error of order $O(n^{-1/2})$ *for any dimension*, but methods may require large samples for high-dimensions (to get an acceptable standard error).

Monte Carlo

Monte Carlo Methods

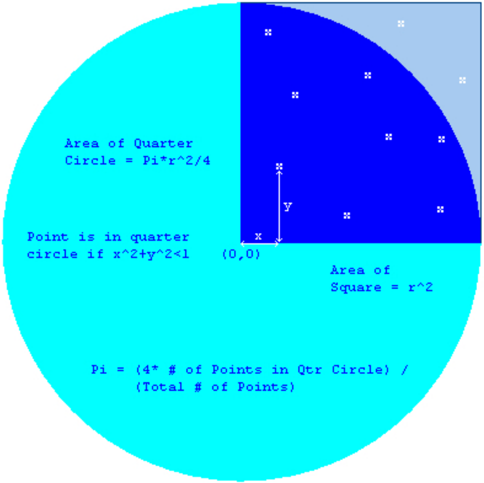
Monte Carlo methods are computational tools characterized by the use of random number generators to obtain a numerical approximation to an unknown quantity.



- Enrico Fermi (1901-1954, physicist)
- Stanslaw Ulam (1909-1984, mathematical physicist)
- John von Neumann (1904-1957, everything)

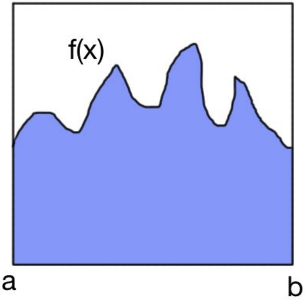
Example: Estimating π

Draw samples uniformly from unit square.



Numerical Approximation of Integral

Consider a single-dimensional integral, i.e. $A = \int_a^b f(x)dx$ =area under curve.



Simple approximation: sum over N points

$$A = \sum_{i=1}^{N} f(x_i)\Delta x = \frac{b-a}{N}\sum_{i=1}^{N} f(x_i)$$

where $\Delta x = \frac{b-a}{N}$ and $x_i = a + (i - 0.5)\Delta x$.

- takes the value of f at the midpoint of each subinterval
- can be made more accurate using Simpson's method, trapezoid rule, etc.

Generalizes to d dimensions with the hyperrectangle defined by the Cartesian product of the intervals $([a_1, b_1], [a_2, b_2], \ldots, [a_d, b_d])$; approximate the $(d+1)$-dimensional volume below the d-dimensional function $f(x)$ by

$$V^{(d+1)} = \frac{(b_1 - a_1)(b_2 - a_2)\cdots(b_d - a_d)}{N_1 N_2 \cdots N_d} \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \cdots \sum_{i_d=1}^{N_d} f(x_i)$$

where $x_i = (x_{i1}, x_{i2}, \ldots, x_{id})^T$ is a d-dimensional vector with each x_i defined as above.

- can be rewritten as

$$V^{(d+1)} = \frac{V^{(d)}}{N} \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \cdots \sum_{i_d=1}^{N_d} f(x_i) = V^{(d)} \frac{\sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \cdots \sum_{i_d=1}^{N_d} f(x_i)}{N}$$

  with $V^{(d)}$ the d-dimensional volume defining the integration area and N the total number of points
- the second term can be interpreted as taking the *average* over f in the interval in question, i.e. $V^{(d+1)} = V^{(d)} \langle f \rangle$ with $\langle f \rangle = \frac{\sum_{i=1}^{N} f(x_i)}{N}$

Order of error is $O(\{\Delta x\}^{2/d}) = O(N^{-2/d})$ as $N \to \infty$

Monte Carlo Approximation of Integrals

Monte Carlo integration is similar to the above, but **instead of sampling at regular intervals** Δx, points are chosen randomly and then the average is taken over those.

one-dimension pick N points x_i randomly in the interval $[a, b]$, then approximate $\int_a^b f(x)dx$ as

$$\frac{b-a}{N}\sum_{i=1}^{N} f(x_i)$$

d-dimensions pick vectors $x_i = (x_1, x_2, \ldots, x_d)^T$ randomly from the hyperrectangle $[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_d, b_d]$, and approximate the $(d+1)$-dimensional volume below the d-dimensional function $f(x)$ by

$$V^{(d+1)} \approx V^{(d)} \frac{\sum_{i=1}^{N} f(x_i)}{N} = V^{(d)}\langle f\rangle$$

Notice the similarity to the numerical integration above; <u>but</u> order of error is $O(N^{-1/2})$ for all d.

MC Approximation of Expectations

In statistics we often encounter a quantity expressed as the expected value of a function of a random variable, $E[h(X)]$.

- let f be the density of X, $\mu = E[h(X)]$ w.r.t. f
- given an i.i.d. sample $X_1, \ldots, X_n$ from f, we can approximate μ by the sample mean

$$\hat{\mu}_{\text{MC}} = \frac{1}{n} \sum_{i=1}^{n} h(x_i) \to \int h(x) f(x) dx = \mu \ \text{ a.s. by SLLN}$$

We are approximating $E[h(X)]$ by randomly sampling n observations from f and then plugging them in to an estimator for $E[h(X)]$.

More Examples

Example

Suppose we want to estimate the variance $\sigma^2(f)$. We can use

$$\widehat{\sigma^2(f)}_{\mathrm{MC}} = \frac{1}{n-1}\sum_{i=1}^{n}(h(x_i) - \hat{\mu}_{\mathrm{MC}})^2 \to \int (h(x) - \mu)^2 f(x)dx = \sigma^2(f).$$

Example

More generally, suppose we don't know the density, but we have the integral

$$\int_0^1 \frac{4}{1+x^2}dx.$$

A Monte Carlo approximation is found by generating n random numbers *uniformly* from the interval $[0,1]$ and then using the approximation

$$\frac{1}{n}\sum_{i=1}^{n}\frac{4}{1+x_i^2}.$$