Integration in $\mathbb{R}^d$:

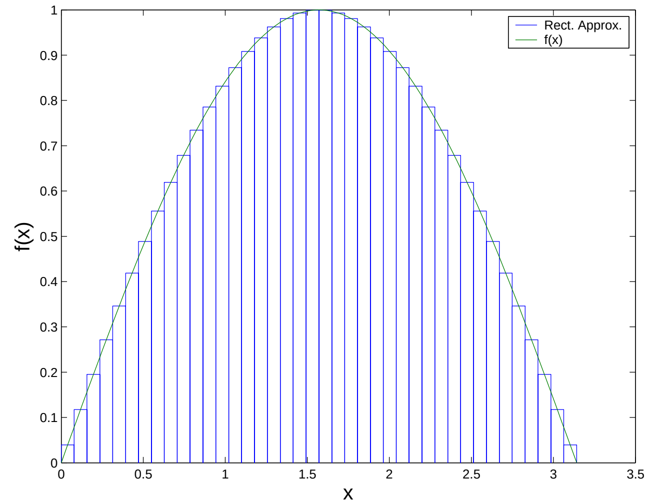
Stochastic and Deterministic Methods

By **Robert L. Wolpert**
Duke University ISDS

Let's consider how to evaluate the integral of a function $f(x)$—first over a one-dimensional interval $[a, b]$, later over a square in $\mathbb{R}^2$ or a cube in $\mathbb{R}^3$ and eventually over a (hyper-)cube in $\mathbb{R}^d$ for a LARGE dimension d. This problem comes up all the time in modern statistics; we'll see some examples below. We begin in one dimension. Let's find ways to approximate the integral:

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

1. Rectangle Method



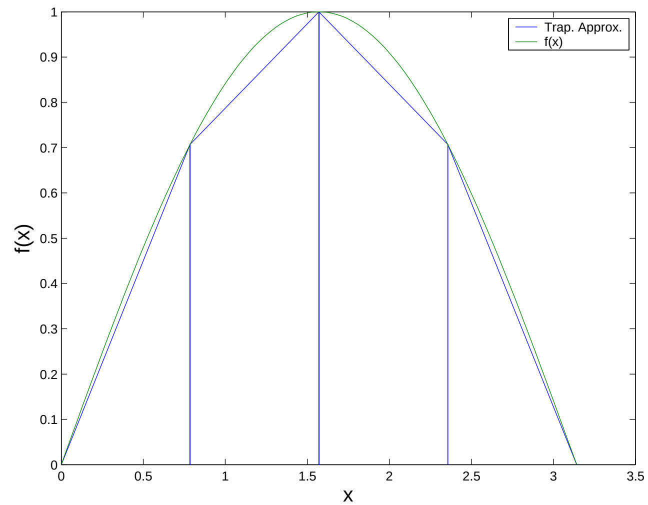

```
function rv = int_R(N, a,b, fun)
% int_R(): Numerical integration, rectangle method

    x  = linspace(a,b, N+1);  % Evenly-spaced

    dx = diff(x,1);           % Interval Widths
    mx = x(1:end-1)+dx/2;     % Interval Midpoints
    fx = feval(fun,mx);       % F(x) at midpoints

    rv = dx * fx';            % Rectangle Integral approx
```


2. Trapezoid Method



```
function rv = int_T(N, a,b, fun)
% int_T(): Numerical integration, trapezoid method

    x  = linspace(a,b, N+1);  % Evenly-spaced
    y  = feval(fun,x);

    dx = diff(x,1);           % Interval Widths
    fx = mean(feval(fun,x([1:end-1;...  % Avg of f(x)
                           2:end])));   % at endpoints
    rv = dx * fx';             % Trapezoid Integral approx
```


Which is Better? Why?

- For specified number N of function evaluations, which is more accurate?
- For specified accuracy, which uses fewer function evaluations?
- What if $f(x)$ is smooth?
- What if $f(x)$ is *not* smooth?
- Are there methods better than *both* of these?

Let's look at some evidence.

$$\begin{aligned} I(0,\pi/2,\sin) &= \int_0^{\pi/2} \sin(x)\,dx \\ &= -\cos(\pi/2)+\cos(0) \\ &= 1 \end{aligned}$$

$N =$	2	4	6	8	10
$R(N,0,\pi/2,\sin) =$	1.0262	1.0065	1.0029	1.0016	1.0010
$T(N,0,\pi/2,\sin) =$	0.9481	0.9871	0.9943	0.9481	0.9979

Evidently the Rectangle method is about twice as accurate. Let's see why.

Consequences of Taylor's Theorem

Look at the integral with $N = 1$ on a short interval $[-\epsilon, \epsilon]$:

$$\begin{array}{rlllll} f(x) = & f(0) & +xf'(0) & +\frac{1}{2}x^2 f''(0) & +\frac{1}{6}x^3 f'''(0) & +O(x^4) \\ f(\pm\epsilon) = & f(0) & \pm\epsilon f'(0) & +\frac{1}{2}\epsilon^2 f''(0) & \pm\frac{1}{6}\epsilon^3 f'''(0) & +O(\epsilon^4) \\ \int_{-\epsilon}^{\epsilon} f(x) = & 2\epsilon\, f(0) & +0 & +\frac{\epsilon^3}{3} f''(0) & +0 & +O(\epsilon^5) \end{array}$$

$$\begin{array}{lll} R(1,f) & = 2\epsilon f(0) & = 2\epsilon\, f(0) + O(\epsilon^5) \\ & & = I(f) - \frac{\epsilon^3}{3} f''(0) + O(\epsilon^5) \end{array}$$

$$\begin{array}{lll} T(1,f) & = \epsilon[f(-\epsilon) + f(\epsilon)] & = 2\epsilon\, f(0) + \epsilon^3 f''(0) + O(\epsilon^5) \\ & & = I(f) + \frac{2\epsilon^3}{3} f''(0) + O(\epsilon^5) \end{array}$$

Thus $R(f)$ *is* about twice as good. What if $N > 1$?

Simpson's Clever Idea

$$R(N,f) = I(f) - \frac{\epsilon^2}{6}\int_a^b f''(x)\,dx + O(\epsilon^4)$$
$$T(N,f) = I(f) + \frac{\epsilon^2}{3}\int_a^b f''(x)\,dx + O(\epsilon^4)$$

Simpson had the good idea to look at the weighted average $S(N,f) \equiv [2\,R(N,f) + T(N,f)]/3$:

$$S(N,f) = I(f) \qquad\qquad + O(\epsilon^4)$$

How much does it matter whether the error is $O(\epsilon^2)$ or $O(\epsilon^4)$?

How many evals do we need?

Since $R(N,f) = I(f) + O(\epsilon^2)$ and $T(N,f) = I(f) + O(\epsilon^2)$, with either method we need order $N > \delta^{-1/2}$ function evaluations for the error to be smaller than δ; for an accuracy of m digits, so that $\delta = 10^{-m}$, the number of evaluations needed (and hence the time to evaluate the integral) grows like $N \approx c\,10^{m/2}$.

With Simpson's method, $S(N,f) = I(f) + O(\epsilon^4)$ so we need only $N \approx c\,10^{m/4}$. Look what a difference it makes:

	$\delta =$	0.1	0.01	0.001	10^{-4}	10^{-6}
$R(N,0,\pi/2,\sin):$	$N =$	2	4	10	33	321
$T(N,0,\pi/2,\sin):$	$N =$	2	5	14	46	454
$S(N,0,\pi/2,\sin):$	$N =$	1	1	2	3	7

What about d dimensions?

It takes $N \approx \left(\frac{b-a}{dx}\right)^2$ points to fill the square $[a,b]^2 \subset \mathbb{R}^2$ with a grid of points spaced dx apart— so, if we try to approximate the two-dimensional integral

$$I(f) = \int_a^b \left(\int_a^b f(x,y)\,dx\right) dy$$
$$\approx R\Big(M,a,b,R(M,a,b,f)\Big)$$

by iterating the Rectangle (or Trapezoid) method it will take $N = M^2$ function evaluations to achieve an accuracy of $\delta \propto M^{-1/2} = N^{-1}$, so errors fall off only as $1/N$ and it takes about $N \approx c\,\delta^{-1}$ evaluations to achieve accuracy δ.

Integration in d dimensions

Iterating Simpson's rule

$$I(f) = \int_a^b \left(\int_a^b f(x,y)\,dx\right) dy$$
$$\approx S\Big(M,a,b,S(M,a,b,f)\Big)$$

errors fall off at rate $1/N^2$ so it takes about $N \approx c\,10^{-m/2}$ evaluations to achieve m decimals of precision in 2 dimensions.

In d dimensions it will take $N = M^d$ evaluations to achieve $\delta \propto M^{-4} = N^{-4/d}$ for Simpson's method, so $N \approx 10^{-dm/4}$ evaluations are needed to achieve m decimals of precision. For dimensions $d > 5$ or 10 this is simply impractical.

Monte Carlo Integration

If X_i are independent random variables from the uniform $\mathsf{Un}(a,b)$ distribution, then

$$\begin{aligned} \mathsf{E}[f(X_i)] &= \frac{1}{b-a}\int_a^b f(x)\,dx \\ &= I(f)/(b-a); \end{aligned}$$

by the Strong Law of Large Numbers (SLLN),

$$= \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} f(X_i).$$

Thus we can approximate $I(f) \approx (b-a)\overline{f(X_i)}$; how good is this approximation?

Central Limit Theorem

The random variables $Y_i = f(X_i)$ are independent with means and variances

$$\begin{aligned} \mu &= \frac{1}{b-a}\int_a^b f(x)\,dx = \frac{I(f)}{b-a}; \\ \sigma^2 &= \frac{1}{b-a}\int_a^b (f(x)-\mu)^2\,dx \\ &= \frac{1}{b-a}\int_a^b f(x)^2\,dx - \mu^2 = \frac{I(f^2)}{b-a} - \frac{I(f)^2}{(b-a)^2}. \end{aligned}$$

By the Central Limit Theorem (CLT),

$$\overline{f(X_i)} = \frac{1}{N}\sum_{i=1}^{N} f(X_i) \approx \mathsf{No}(\mu, \sigma^2/N).$$

Stochastic *vs.* Quadrature

Amazingly, this holds in *any* number d of dimensions; thus the approximate error

$$|I(f) - (b-a)^d\overline{f(X_i)}| \approx (b-a)^d\sigma/\sqrt{N}$$

falls off like $N^{-1/2}$. This is faster than Simpson's method if $N^{-1/4} \le N^{-2/d}$, *i.e.*, $d > 7$, and is faster than the Rectangle or Trapezoid method if $d > 3$.

Free lunch?

Does this *really* work in *any* number d of dimensions? Statisticians must integrate functions in **hundreds** of dimensions; does Monte Carlo make this practical?

"Not yet" is a good answer... the constant σ in the error bound

$$|I(f) - (b-a)^d\overline{f(X_i)}| \approx \sigma(b-a)^d\,N^{-1/2}$$

can be huge.