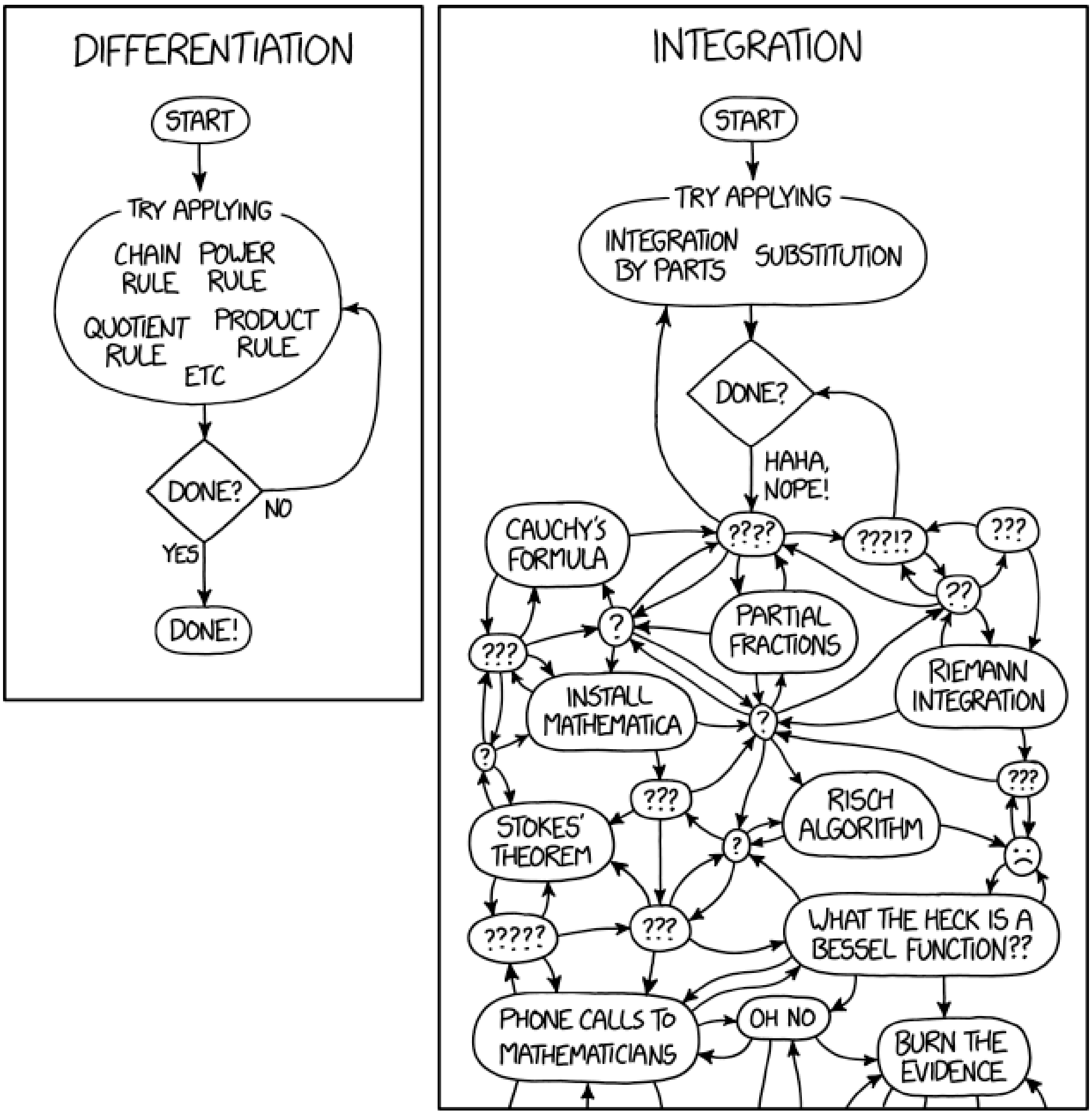
Chapter 6: Monte Carlo Integration

Monte Carlo integration is a statistical method based on random sampling in order to approximate integrals. This section could alternatively be titled,

> "Integrals are hard, how can we avoid doing them?"



https://xkcd.com/2117/

1 A Tale of Two Approaches

Consider a one-dimensional integral.

The value of the integral can be derived analytically only for a few functions, f. For the rest, numerical approximations are often useful.

Why is integration important to statistics?

1.1 Numerical Integration

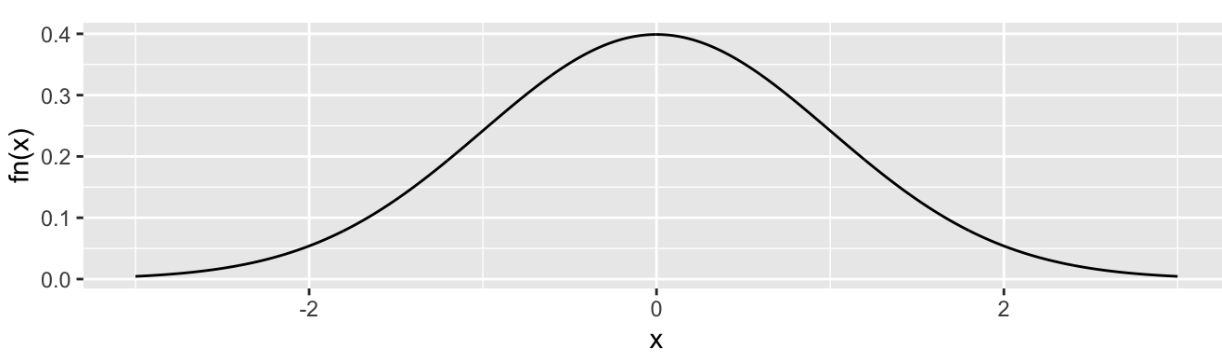
Idea: Approximate $\int_a^b f(x)dx$ via the sum of many polygons under the curve $f(x)$.

To do this, we could partition the interval $[a, b]$ into m subintervals $[x_i, x_{i+1}]$ for $i = 0, \dots, m-1$ with $x_0 = a$ and $x_m = b$.

Within each interval, insert $k+1$ nodes, so for $[x_i, x_{i+1}]$ let x_{ij}^* for $j = 0, \dots, k$, then

$$\int\limits_a^b f(x)dx = \sum_{i=0}^{m-1} \int\limits_{x_i}^{x_{i+1}} f(x)dx \approx \sum_{i=0}^{m-1} \sum_{j=0}^{k} A_{ij} f(x_{ij}^*)$$

for some set of constants, A_{ij}.



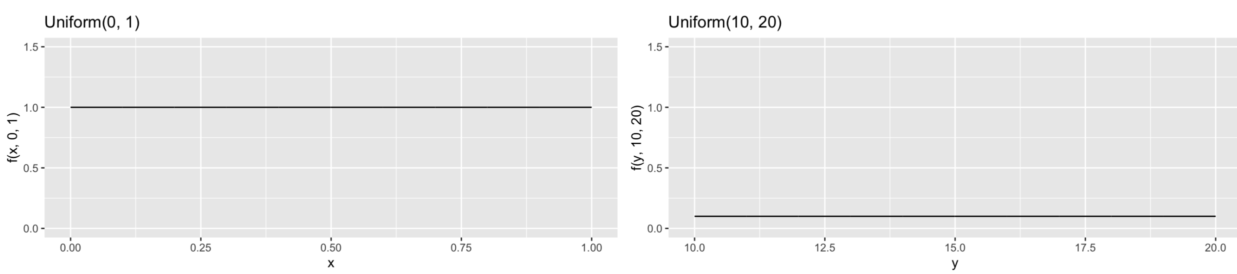
1.2 Monte Carlo Integration

How do we compute the mean of a distribution?

Example 1.1 Let $X \sim Unif(0, 1)$ and $Y \sim Unif(10, 20)$.

```
x <- seq(0, 1, length.out = 1000)
f <- function(x, a, b) 1/(b - a)
ggplot() +
  geom_line(aes(x, f(x, 0, 1))) +
  ylim(c(0, 1.5)) +
  ggtitle("Uniform(0, 1)")

y <- seq(10, 20, length.out = 1000)
ggplot() +
  geom_line(aes(y, f(y, 10, 20))) +
  ylim(c(0, 1.5)) +
  ggtitle("Uniform(10, 20)")
```



Theory

1.2.1 Notation

θ

$\hat{\theta}$

Distribution of $\hat{\theta}$

$E[\hat{\theta}]$

$Var(\hat{\theta})$

$\hat{E}[\hat{\theta}]$

$V\hat{a}r(\hat{\theta})$

$se(\hat{\theta})$

$\hat{se}(\hat{\theta})$

1.2.2 Monte Carlo Simulation

What is Monte Carlo simulation?

1.2.3 Monte Carlo Integration

To approximate $\theta = \int x f(x) dx$, we can obtain an iid random sample $X_1, \dots, X_n$ from f and then approximate θ via the sample average

Example 1.2 Again, let $X \sim Unif(0, 1)$ and $Y \sim Unif(10, 20)$. To estimate $E[X]$ and $E[Y]$ using a Monte Carlo approach,

Now consider $E[g(X)]$.

$$\theta = E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx.$$

The Monte Carlo approximation of θ could then be obtained by

1.

2.

Definition 1.1 *Monte Carlo integration* is the statistical estimation of the value of an integral using evaluations of an integrand at a set of points drawn randomly from a distirbution with support over the range of integration.

Example 1.3

Why the mean?

Let $E[g(X)] = \theta$, then

and, by the strong law of large numbers,

Example 1.4 Let $v(x) = (g(x) - \theta)^2$, where $\theta = E[g(X)]$, and assume $g(X)^2$ has finite expectation under f. Then

$$Var(g(X)) = E[(g(X) - \theta)^2] = E[v(x)].$$

We can estimate this using a Monte Carlo approach.

Monte Carlo integration provides slow convergence, i.e. even though by the SLLN we know we have convergence, it may take us a while to get there.

But, Monte Carlo integration is a **very** powerful tool. While numerical integration methods are difficult to extend to multiple dimensions and work best ith a smooth integrand, Monte Carlo does not suffer these weaknesses.

-

-

1.2.4 Algorithm

The approach to finding a Monte Carlo estimator for $\int g(x)f(x)dx$ is as follows.

1.

2.

3.

4.

Example 1.5 Estimate $\theta = \int_0^1 g(x)dx$.

Example 1.6 Estimate $\theta = \int_a^b g(x)dx$.

Another approach:

Example 1.7 Monte Carlo integration for the standard Normal cdf ($N(0,1)$). Let $X \sim N(0,1)$, then the pdf of X is

$$\phi(x) = f(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right), \qquad -\infty < x < \infty$$

and the cdf of X is

$$\Phi(x) = F(x) = \int_{\infty}^{x} \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t^2}{2}\right)dt.$$

We will look at 3 methods to estimate $\Phi(x)$ for $x > 0$.

1.2.5 Inference for MC Estimators

The Central Limit Theorem implies

So, we can construct confidence intervals for our estimates

1.

2.

But we need to estimate $Var(\hat{\theta})$.

So, if $m \uparrow$ then $Var(\hat{\theta}) \downarrow$. How much does changing m matter?

Example 1.8 If the current $se(\hat{\theta}) = 0.01$ based on m samples, how many more samples do we need to get $se(\hat{\theta}) = 0.0001$?

Is there a better way to decrease the variance? **Yes!**