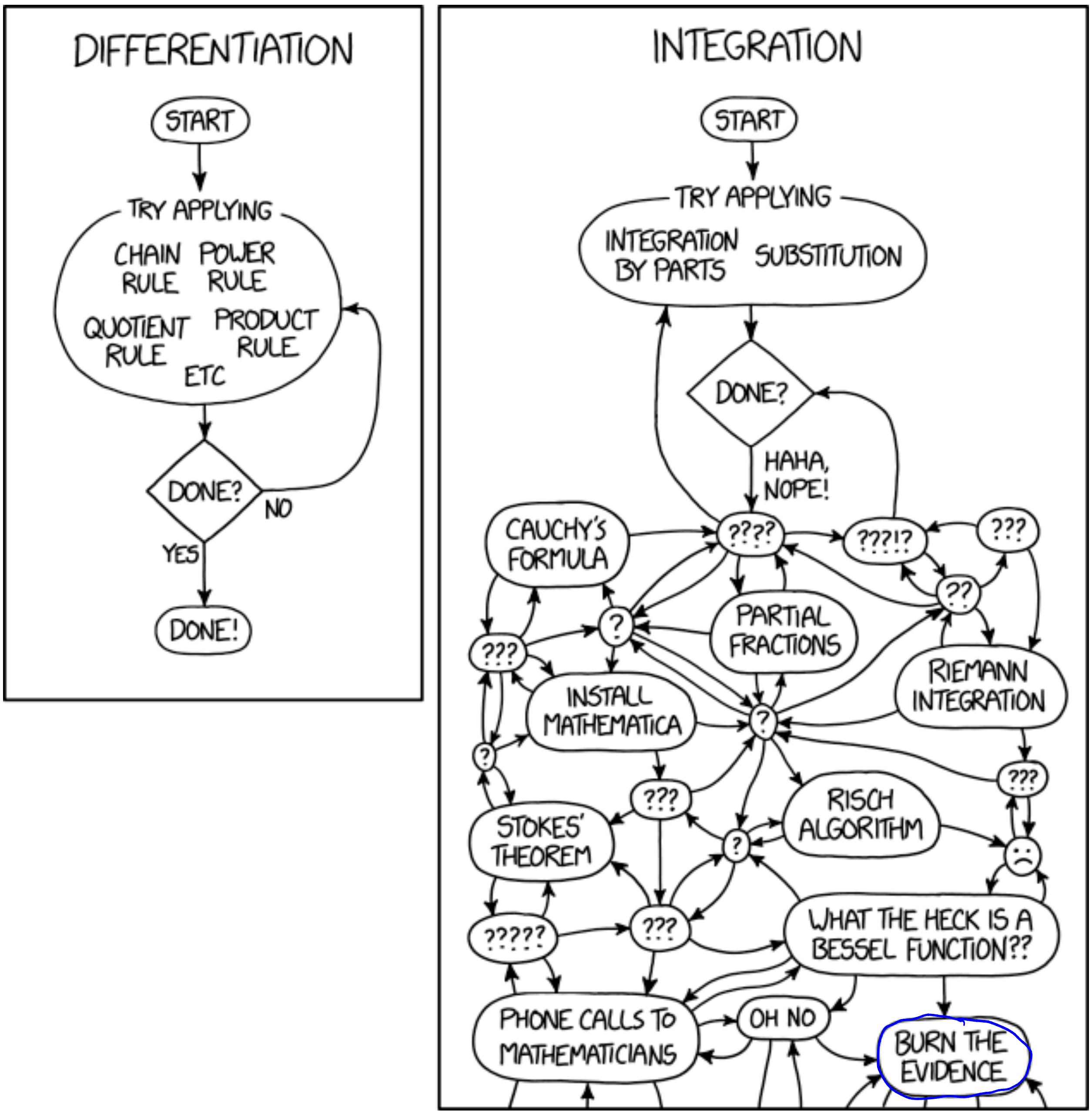
Chapter 6: Monte Carlo Integration

ch. 3

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.

$$\int_a^b \underbrace{f(x)}_{\text{integrand.}} dx$$

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?

Many quantities of interest in statistics can be written as the expectation of a function of a random variable

$$E[g(X)] = \int_{\mathcal{X}} \underbrace{g(x)f(x)}_{\text{integrand}} dx$$

1.1 Numerical Integration

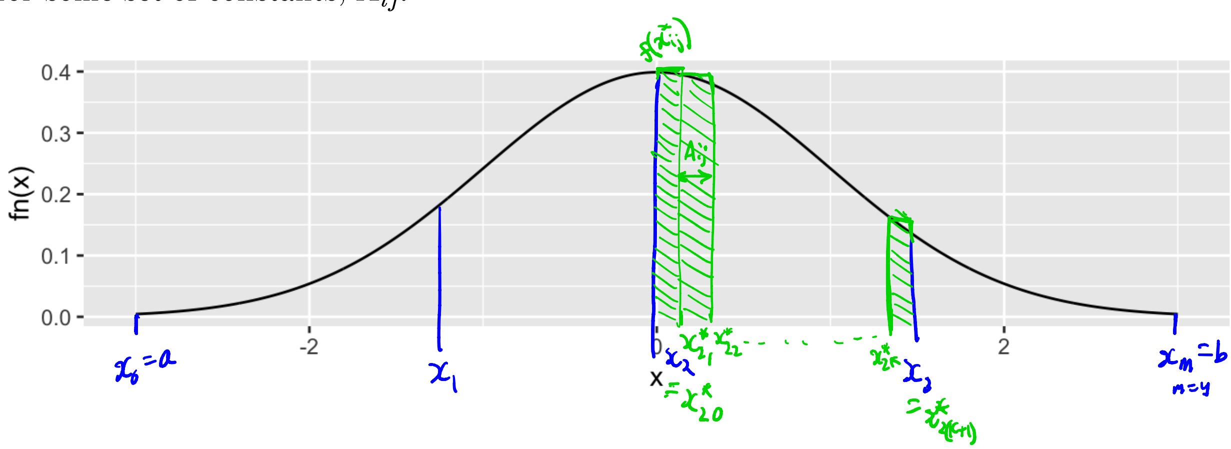
Idea: Approximate $\int_a^b f(x)dx$ via the sum of many polygons under the curve $f(x)$.

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

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

for some set of constants, A_{ij}.



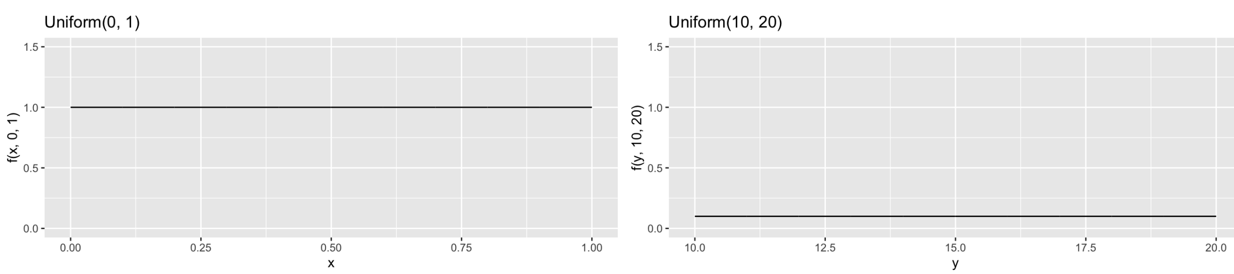
1.2 Monte Carlo Integration

How do we compute the mean of a distribution?

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

```r
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
(exact)

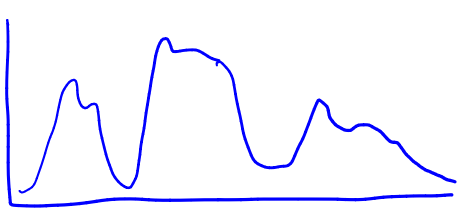
$$EX = \int_0^1 x f(x)\,dx = \int_0^1 x \cdot 1\,dx = \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2}$$

$$EY = \int_{10}^{20} y f(y)\,dy \quad \text{where } f(y) = \begin{cases} \frac{1}{10} & 10 \le y \le 20 \\ 0 & \text{o.w.} \end{cases}$$

$$= \int_{10}^{20} y \cdot \frac{1}{10}\,dy = \frac{1}{10}\left[\frac{y^2}{2}\right]_{10}^{20} = 15.$$

How about some other dsn?

$x \sim$



want to find EX

Probably can't do this in closed form

$\Rightarrow$ need an approximation.

1.2.1 Notation

θ = parameter (unknown)

$\hat{\theta}$ = estimator of θ, statistic (sometimes we use $\bar{X}$, s^2, etc. instead of $\hat{\theta}$).

Distribution of $\hat{\theta}$ = sampling distribution

$\hat{\theta}$ is a function of random variables ⇒ a random variable.

$E[\hat{\theta}]$ = on average, what is the value of $\hat{\theta}$?
theoretic mean of the sampling distribution of $\hat{\theta}$.

$Var(\hat{\theta})$ = theoretical variance of $\hat{\theta}$
variance of the sampling distribution of $\hat{\theta}$

estimated versions.

$\hat{E}[\hat{\theta}]$ = estimated mean of distribution of $\hat{\theta}$

$\hat{Var}(\hat{\theta})$ = estimated variance of dsn of $\hat{\theta}$

$se(\hat{\theta}) = \sqrt{Var(\hat{\theta})}$ theoretical se of $\hat{\theta}$ = sd of sampling dsn of $\hat{\theta}$

$\hat{se}(\hat{\theta}) = \sqrt{\hat{Var}(\hat{\theta})}$ estimated se of $\hat{\theta}$ = estimated sd of sampling dsn of $\hat{\theta}$

1.2.2 Monte Carlo Simulation

What is Monte Carlo simulation?

Computer simulation that generates a large quantity of samples from a distribution. The distribution characterizes the population from which the sample is drawn.

(sounds a lot like ch. 3).

1.2.3 Monte Carlo Integration

parameter characterizes the population. Thing we care about!

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

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} X_i \approx EX$$

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,

① draw $X_1, \dots, X_m \sim Unif(0,1)$

② Compute $\hat{\theta} = \frac{1}{m} \sum_{i=1}^{m} X_i$

① draw $Y_1, \dots, Y_m \sim Unif(10,20)$

② Compute $\hat{\theta} = \frac{1}{m} \sum_{i=1}^{m} Y_i$

This is useful when we can't compute EX in closed form. Also useful to approximate other integrals.

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

parameter of interest

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

The Monte Carlo approximation of θ could then be obtained by

1. Draw $X_1, \dots, X_m \sim f$

2. $\hat{\theta} = \frac{1}{m} \sum_{i=1}^{n} g(X_i)$.

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 distribution with support over the range of integration.

Example 1.3
(A) parameter estimation: Linear models vs. generalized linear models
$Y = X\beta + \varepsilon \quad \varepsilon \sim N(0, \sigma^2), \quad \hat{\beta} = (X^TX)^{-1}X^TY$ closed form solution

GLM: $Y \sim Binom(p)$
$logit(p) = \beta_0 + \beta_1 X$ no estimate for β_0, β_1 in closed form.

(B) estimate quantiles of a dsn. Find y s.t. $0.9 = \int_{-\infty}^{y} f(x)\,dx$.

Why the mean?

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

$$E(\hat{\theta}) = E\left[\frac{1}{m}\sum_{i=1}^{m} g(X_i)\right] = \frac{1}{m}\sum_{i=1}^{m} E(g(X_i)) = \frac{1}{m}[\underbrace{\theta + \cdots + \theta}_{m \text{ times}}] = \theta$$

so $\hat{\theta}$ obtained from MC integration approach is unbiased

and, by the strong law of large numbers,

$$\hat{\theta} = \frac{1}{m}\sum_{i=1}^{m} g(X_i) \xrightarrow{p} E[g(X)] = \theta.$$

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.

$$\hat{Var}[g(X)] = \hat{E}[v(X)]$$

(1) Sample $X_1, \ldots, X_m$ from f

(2) Compute $\frac{1}{m}\sum_{i=1}^{m}[g(X_i) - \theta]^2$ don't know this! can replace with $\hat{\theta} = \frac{1}{m}\sum_{i=1}^{m} g(X_i)$

Use to estimate sampling variance. $\Rightarrow$ $Var\,\hat{\theta} = Var\left[\frac{1}{m}\sum_{i=1}^{m} g(X_i)\right]$
$= \frac{1}{m^2}\sum Var\, g(X_i) = \frac{1}{m} Var\, g(X)$ can estimate using $\hat{Var}\, g(X)$.

estimate $se(\hat{\theta})$ by $\sqrt{\frac{1}{m}\hat{Var}\, g(X)}$

When $\mathrm{Var}\, g(X)$ exists and is finite, the CLT states

$$\frac{\hat{\theta} - \underbrace{E\hat{\theta}}_{=\theta}}{\sqrt{\mathrm{Var}\,\hat{\theta} = \frac{\mathrm{Var}\, g(X)}{m}}} \xrightarrow{d} N(0,1) \quad \text{as } m \to \infty.$$

Hence if m is large,

$$\hat{\theta} \,\dot{\sim}\, N\left(\theta, \frac{\boxed{\mathrm{Var}\, g(X)}}{m}\right)$$

can plug in estimate $\widehat{\mathrm{Var}}\, g(X)$ from above.

We can use this knowledge to create confidence limits or error bounds on the MC estimate of the integral $\hat{\theta}$.

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 with a smooth integrand, Monte Carlo does not suffer these weaknesses.

numeric integration cannot say the same.

- MC integration does not attempt a systematic exploration of the p-dimensional support region of f. (curse of dimensionality).
- MC doesn't require integrand to be smooth, does not require finite support!

1.2.4 Algorithm

$\int h(x)dx$

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

before R

1. Select f, g to define θ as an expected value.
2. derive estimator s.t. $\hat{\theta}$ approximates $\theta = E[g(x)] = \int_{-\infty}^{\infty} g(x) f(x) dx = \int_{-\infty}^{\infty} h(x) dx$.

in R.

3. Sample $X_1, \dots, X_m$ from f
4. Compute $\hat{\theta} = \frac{1}{m}\sum_{i=1}^{m} g(X_i)$.

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

(1) Let f be the Unif $(0,1)$ density $\Rightarrow g(x) = h(x)$.

(2) Then $\theta = \int_0^1 h(x)dx = \int_0^1 g(x) \cdot \underset{\uparrow \text{ unif } (0,1) \text{ density}}{1}\, dx = E[g(X)]$ ✓

(3) Sample $X_1, \dots, X_m$ from f

> X <- runif(m, 0, 1).

(4) Compute $\hat{\theta} = \frac{1}{m}\sum_{i=1}^{m} g(X_i)$

write this function in R.

> mean(g(x))

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

(1) Choose $f \equiv \text{Unif}(a,b) \Rightarrow f(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & \text{o.w.} \end{cases}$

Then $g(x) = (b-a) \cdot h(x)$.

(2) So that $\theta = \int_a^b h(x)dx = \int_a^b (b-a)h(x) \cdot \frac{1}{b-a}\, dx = \int_a^b g(x) f(x)\, dx = E g(X)$, $X \sim \text{Unif}(a,b)$

(3) Sample $X_1, \dots, X_m$ from $\text{Unif}(a,b)$. > x <- runif(m, a, b).

(4) Compute $\hat{\theta} = \frac{1}{m}\sum_{i=1}^{m} (b-a) \cdot h(x_i)$ > (b-a) mean(h(x)).

Another approach:

(a,b) maps $(0,1)$.

What if I chose $Y \sim \text{Unif}(0,1)$ instead?

Then $f(y) = \begin{cases} 1 & 0 \le y \le 1 \\ 0 & \text{o.w.} \end{cases}$

But we care about $E[g(Y)] = \int_{\mathcal{Y}} g(y) f(y)\, dy$

$\mathcal{Y}$: support of f.

We want to integrate from (a,b), but support of dsn is $(0,1)$. So we need a change of variable to use MC integration.

Need a function to map $x \in (a,b)$ to $y \in (0,1)$. We will use linear transformation.

$(y \to x)$ $\quad \frac{x-a}{b-a} = \frac{y-0}{1-0} \Rightarrow \frac{x-a}{b-a} = y$.

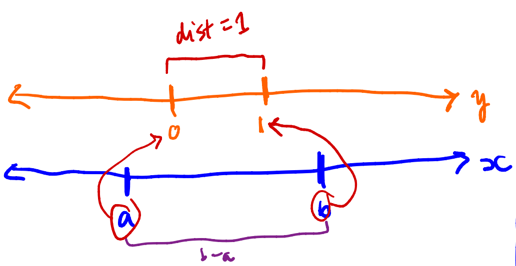
solve for x

$x = a + (b-a)y$.

$dx = (b-a)dy$.

$$\theta = \int_a^b g(x)\,dx = \int_0^1 \underbrace{g(a+(b-a)y)\cdot(b-a)}_{g}\, dy.$$

$f(y) = 1$.



To get $\hat{\theta}$,

(1) Simulate $Y_1, \dots, Y_m$ from $\text{Unif}(0,1)$.

(2) $\hat{\theta} = \frac{1}{m}\sum_{i=1}^{m} \{ g(a + y_i(b-a))(b-a) \}$

We can use this if the limits of integration don't match any density!

Example 1.7 Monte Carlo integration for the standard Normal cdf. 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 estimator

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!**