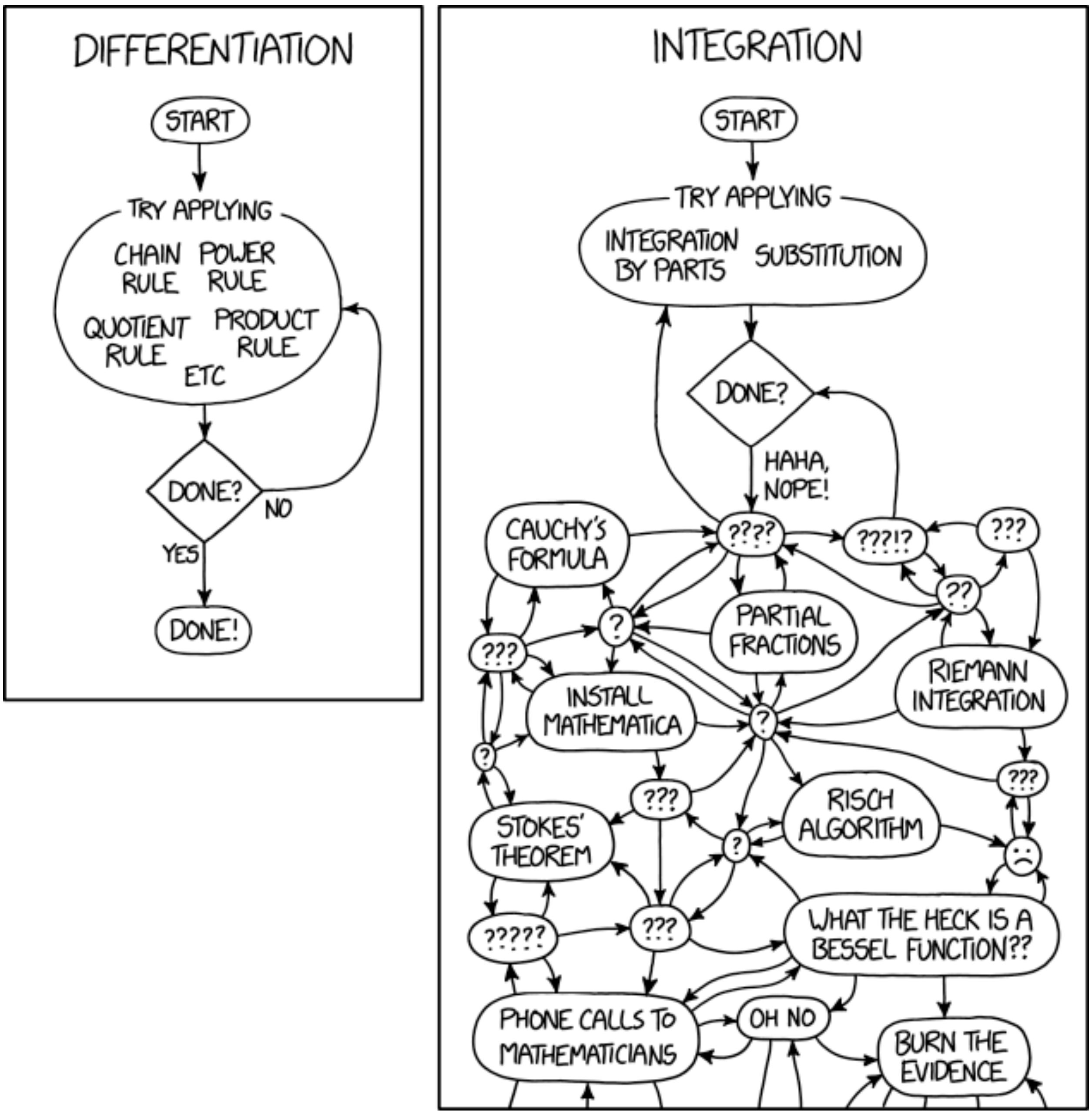
Chapter 6: Monte Carlo Integration

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 inferential statistics can be expressed as the expectation of a function of a random variable. $E[g(X)] = \int \underbrace{g(x)f(x)}_{\text{integrand.}} dx$

1.1 Numerical Integration

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

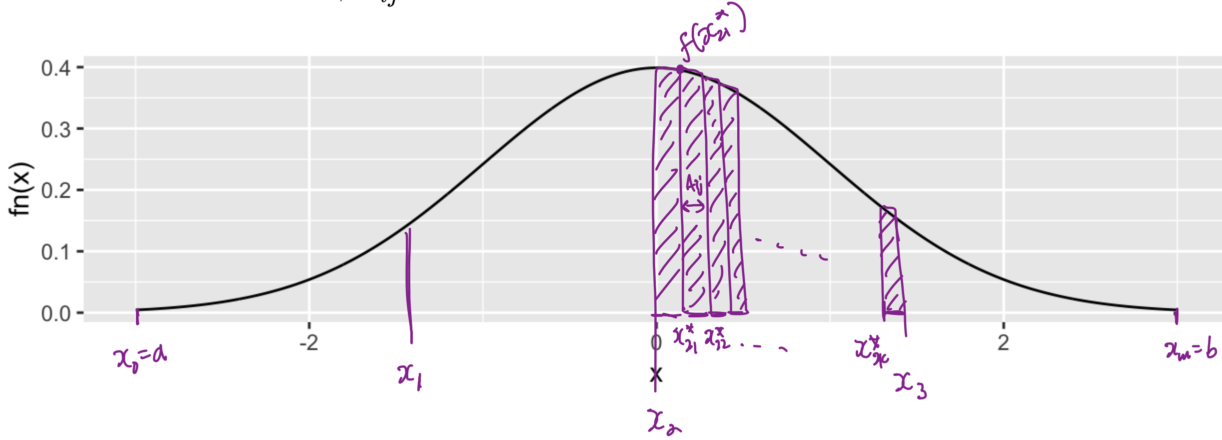
To do this, we could partition the interval $[a, b]$ into m subintervals $[x_i, x_{i+1}]$ for $i = 0, \dots, m-1$ with $x_0 = a$ and $x_m = b$.

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

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

width height

for some set of constants, A_{ij}.



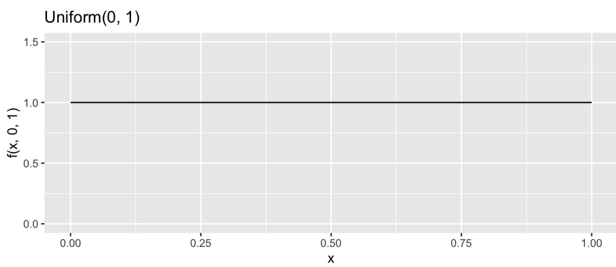
1.2 Monte Carlo Integration

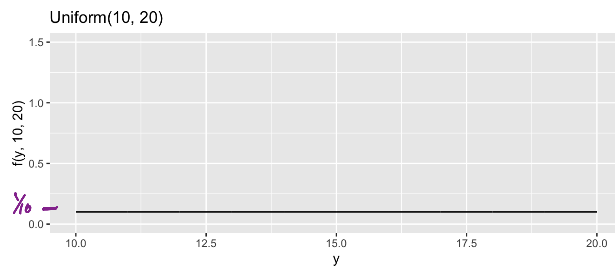
How do we compute the mean of a distribution?

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

$$E(X) = \int_0^1 x f(x)\,dx = \int_0^1 x \cdot 1\,dx = \frac{x^2}{2}\Big|_0^1 = \frac{1}{2}$$

$$E(Y) = \int_{10}^{20} y f(y)\,dy = \int_{10}^{20} y \frac{1}{10}\,dy = \frac{1}{10}\left[\frac{y^2}{2}\right]_{10}^{20} = 15$$

How about some other dsn?

??

probably can't do this is closed form $\Rightarrow$ need approximation.

1.2.1 Notation

θ = parameter (unknown)

$\hat{\theta}$ = estimator of θ, statistic (sometime we write $\bar{X}$, S^2 instead of $\hat{\theta}$).

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

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

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

estimated versions

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

$V\hat{a}r(\hat{\theta})$ = estimated variance of dsn of $\hat{\theta}$

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

$\hat{se}(\hat{\theta})$ = $\sqrt{V\hat{a}r(\hat{\theta})}$ estimated S.E. of $\hat{\theta}$

1.2.2 Monte Carlo Simulation

What is Monte Carlo simulation?

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

(Sounds a like Ch. 3).

1.2.3 Monte Carlo Integration

parameter characterizing population

Thing we care about!

1

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

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

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,

$X \sim Unif(0,1)$, estimate EX:

1. drawing $X_1, \ldots, X_m \overset{iid}{\sim} Unif(0,1)$
2. Compute $\hat{\theta} = \frac{1}{m}\sum_{i=1}^{m} X_i$

$Y \sim Unif(10,20)$, estimate EY:

1. drawing $Y_1, \ldots, Y_m \overset{iid}{\sim} Unif(10,20)$
2. Computing $\hat{\theta} = \frac{1}{m}\sum_{i=1}^{m} Y_i$

This is useful when we can't compute EX in closed form.

Also useful to approximate other integrals...

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

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

The Monte Carlo approximation of θ could then be obtained by

1. Draw $X_1, \ldots, X_m \sim f$
2. Compute $\hat{\theta} = \frac{1}{m}\sum_{i=1}^{m} g(X_i)$.

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 model: $Y = X\beta + \varepsilon \quad \varepsilon \sim N(0, \sigma^2) \implies \hat{\beta} = (X^T X)^{-1} X^T Y$ closed form.

GLM: $Y \sim \text{Binom}(p)$

$\text{logit}(p) = \beta_0 + \beta_1 X$ no estimate for β_0, β_1 in closed form.

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

Why the mean?

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

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

So $\hat{\theta}$ is unbiased.

and, by the strong law of large numbers,

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

"consistency"

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.

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

(1) Sample $X_1, \dots, X_m \sim f$

(2) Compute $\frac{1}{m}\sum_{i=1}^{m} (g(X_i) - \theta)^2$

We don't know this! can replace with $\hat{\theta} = \frac{1}{m}\sum_{i=1}^{m} g(X_i)$.

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

If $\operatorname{Var} g(X) < \infty \implies$ CLT states

$$\frac{\hat{\theta} - E(\hat{\theta})^{=\theta}}{\sqrt{\operatorname{Var}(\hat{\theta})}} \xrightarrow{d} N(0,1) \quad \text{as } m \to \infty.$$

$$\hookrightarrow \operatorname{Var}(\hat{\theta}) = \operatorname{Var}\left(\frac{1}{m}\sum_{i=1}^{m} g(X_i)\right) = \frac{1}{m^2}\sum_{i=1}^{m} \operatorname{Var} g(X_i)$$

$$= \frac{1}{m}\operatorname{Var} g(X).$$

Hence if m is large

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

$$\hat{\theta} \overset{\circ}{\sim} N\left(\theta, \frac{\operatorname{Var} g(X)}{m}\right)$$

We can use this to put 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 does not attempt systematic exploration of the p-dimensional support of f.
- MC does not require integrand to be smooth, does not require finite support.

1.2.4 Algorithm

$\int h(x)dx = \theta$

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

Before R

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

in R.

3. Sample $X_1, \dots, X_m \sim f$
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 Uniform $(0,1)$. Find g s.t. $f \cdot g = h$. $\Rightarrow$ Let $g(x) = h(x)$.

② Then $\theta = \int_0^1 h(x)dx = \int_0^1 g(x)\cdot 1\,dx = E(g(X))$ $X \sim Unif(0,1)$. Unif(0,1) density

③ Sample $X_1, \dots, X_m$ from Unif $(0,1)$.

```
> x <- runif(m).
```

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

```
> mean(g(x))
```

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

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

Then $g(x) = (b-a)h(x)$

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

(3) Sample $X_1, \ldots, X_m \sim$ 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:

change of variable approach.

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

What if I chose $Y \sim$ Unif$(0,1)$ instead? Then $f_Y(y) = \begin{cases} 1 & 0 \le y \le 1 \\ 0 & \text{o.w.} \end{cases}$

But we care about $E(g(X)) = \int_a^b g(x) f(x) dx$.

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.

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

$$\frac{x-a}{b-a} = \frac{y-0}{1} \implies \frac{x-a}{b-a} = y.$$

$\Downarrow$ solve for x $(x \to y)$.

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

Now $\theta = \int_a^b g(x) f(x) dx = \int_0^1 g(a + y(b-a)) f_Y(y) (b-a) dy.$

$\uparrow$

$E[\tilde{g}(Y)]$, $Y \sim$ Unif$(0,1)$,

$\tilde{g}(y) = g(a + y(b-a)) \cdot (b-a)$.

To get $\hat{\theta}$,

(1) Simulate $Y_1, \ldots, Y_m$ from 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 named density.

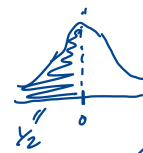
Example 1.7 Monte Carlo integration for the standard Normal cdf. Let $X \sim N(0,1)$, then the pdf of X is

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

and the cdf of X is

$P(X \le x)$ $\phi(x)$

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



why? because now limits of integration are $(0, x)$.

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

Method 1: Note that for $x>0$ $\Phi(x) = \underbrace{\int_{-\infty}^{0} \phi\, dx}_{= 1/2} + \int_0^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt$

change of variables approach

Support of $Y \sim Unif(0,1)$ is $(0<y<1)$. So want function that maps $t \in (0,x)$ to $y \in (0,1)$.

linear transformation!

$\frac{y-0}{1-0} = \frac{t-0}{x-0}$

Let $y = \frac{t}{x}$ so if $t=0 \Rightarrow y=0$ ✓, $t=x \Rightarrow y=1$ ✓

$\Rightarrow t = xy \qquad dt = x\,dy$

Then $\int_0^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt = \int_0^1 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2y^2}{2}\right) x\,dy$

$\Rightarrow$ Want to estimate $\theta = E_Y\left[\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2Y^2}{2}\right) \cdot x\right]$ where $Y \sim Unif(0,1)$.

So a MC estimate could be obtained by:

(1) Sample $Y_1, \ldots, Y_m \sim Unif(0,1)$.

(2) $\hat{\Phi}(x) = 0.5 + \frac{1}{m}\sum_{i=1}^{m}\left\{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2 Y_i^2}{2}\right) \cdot x\right\}$ for $x>0$.

fixed number.

Method 2 Could instead have chosen to $Y \sim Unif(0, x)$

Homework.

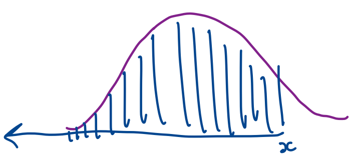
Method 3

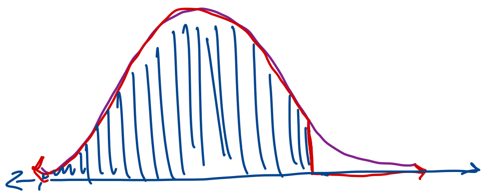
Let $\mathbb{I}$ be an indicator function.

$$\mathbb{I}(Z \le z) = \begin{cases} 1 & \text{if } Z \le z \\ 0 & \text{o.w.} \end{cases}$$

Let $Z \sim N(0,1)$. Then

$$\Phi(x) = \int_{-\infty}^{x} \phi(t)\,dt = \int_{-\infty}^{\infty} \underbrace{\mathbb{I}(t \le x)}_{g(t)}\, \underbrace{\phi(t)}_{f(t)}\,dt$$





So a MC estimator of $\Phi(x)$ is

① Generate $Z_1, \dots, Z_m \sim N(0,1)$

② $\hat{\Phi}(x) = \frac{1}{m} \underbrace{\sum_{i=1}^{m} \mathbb{I}(Z_i \le x)}_{\text{count of } \# Z_i\text{'s} \le x.}$

Notes:

① We can show that Method 3 has less bias in tails. and Method 2 has less bias in the center.

② Method 3 works for any dsn to approximate cdf (change f accordingly).

1.2.5 Inference for MC Estimators

The Central Limit Theorem implies

MC estimator

$$\frac{\hat{\theta} - E(\hat{\theta})}{\sqrt{Var(\hat{\theta})}} \xrightarrow{d} N(0,1) \quad \text{as } M \to \infty$$

$se(\hat{\theta})$

So, we can construct confidence intervals for our estimator

plug in an estimate.

1. 95% CI for $\underset{}{\overset{\checkmark \theta \text{ (unbiased)}}{E(\hat{\theta})}}$: $\hat{\theta} \pm \underbrace{1.96}_{\text{qnorm }(.975)} \sqrt{\hat{Var}(\hat{\theta})}$
2. (HW) 95% CI for $\Phi(2)$: $\hat{\Phi}(2) \pm 1.96 \sqrt{\hat{Var}(\hat{\Phi}(2))}$

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

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

let $\sigma^2 = Var[g(X)]$.

Then $Var[\hat{\theta}] = Var\left[\frac{1}{m} \sum_{i=1}^{m} g(X_i)\right] = \frac{1}{m^2} \sum_{i=1}^{m} Var(g(X_i)) = \frac{\sigma^2}{m}$.

we can estimate this w/ MC integration.

So, $\hat{Var}[\hat{\theta}] = \frac{\hat{\sigma}^2}{m} = \frac{1}{m} \underbrace{\left[\frac{1}{m} \sum_{i=1}^{m} \left[(g(X_i) - \hat{\theta})^2\right]\right]}_{\hat{\sigma}^2} = \frac{1}{m^2} \sum_{i=1}^{M} (g(x_i) - \hat{\theta})^2$

$\hat{E} g(X)$

and $\hat{se}(\hat{\theta}) = \sqrt{\hat{Var}(\hat{\theta})}$

a MC estimator of variance of sampling dsn of $\hat{\theta}$

Recall that we usually use $s^2 = \frac{1}{m-1} \sum_{i=1}^{m} (X_i - \bar{X})^2$ to estimate σ^2.

Why not use s^2 w/ $\frac{1}{m-1}$ instead of $\hat{\sigma}^2$ w/ $\frac{1}{m}$?

For MC integration, m is large so $\frac{1}{m-1} \approx \frac{1}{m}$

Ex: if $m = 1000$, $\frac{1}{m-1} - \frac{1}{m} = 1 \times 10^{-6}$

Some books use $\frac{1}{m-1}$ so $\hat{Var}(\hat{\theta}) = \frac{1}{m(m-1)} \sum_{i=1}^{m} (g(X_i) - \hat{\theta})^2$

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 $se(\hat{\theta}) = 0.0001$?

current $se(\hat{\theta}) = \sqrt{\frac{\sigma^2}{m}} = .01$

$\sqrt{\frac{\sigma^2}{m \cdot a}} = .0001$ find a?

$\frac{\sigma^2}{m} \cdot \frac{1}{a} = (.0001)^2$

$(.01)^2 \cdot \frac{1}{a} = (.0001)^2$

$\left(\frac{.01}{.0001}\right)^2 = a$

$= 10,000$

We would need $10,000 \times m$ samples to achieve $se(\hat{\theta}) = .0001$!

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