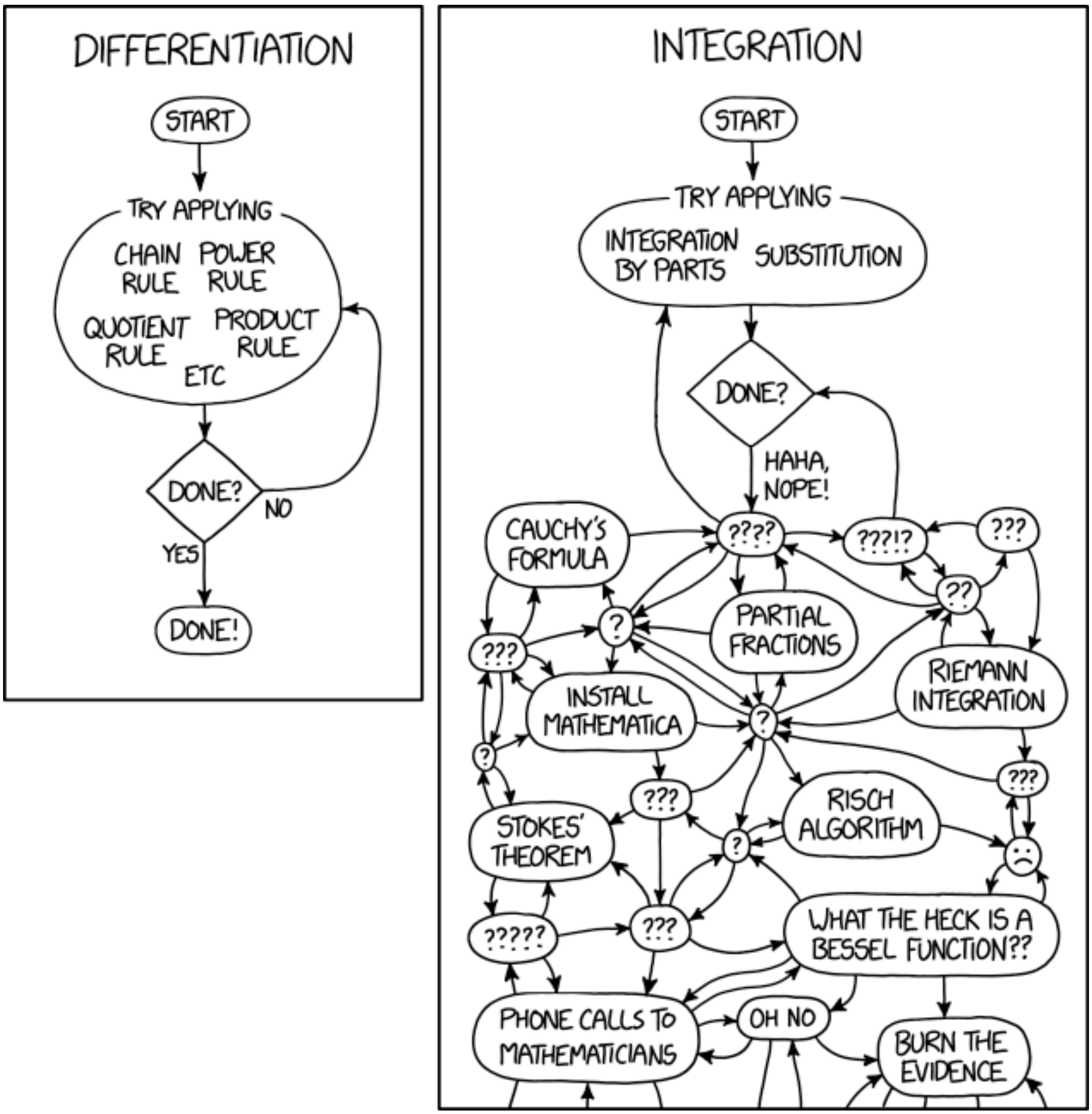
Chapter 6: Monte Carlo Integration

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

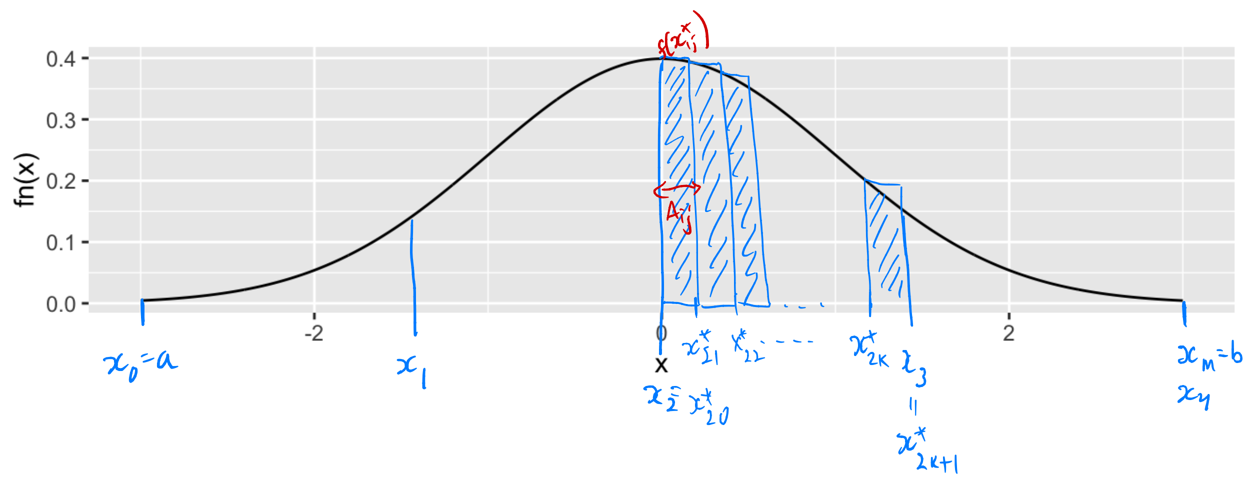
Idea: Approximate $\int_a^b f(x)dx$ via the sum of many polygons under the curve $f(x)$.

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

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

for some set of constants, A_{ij}.



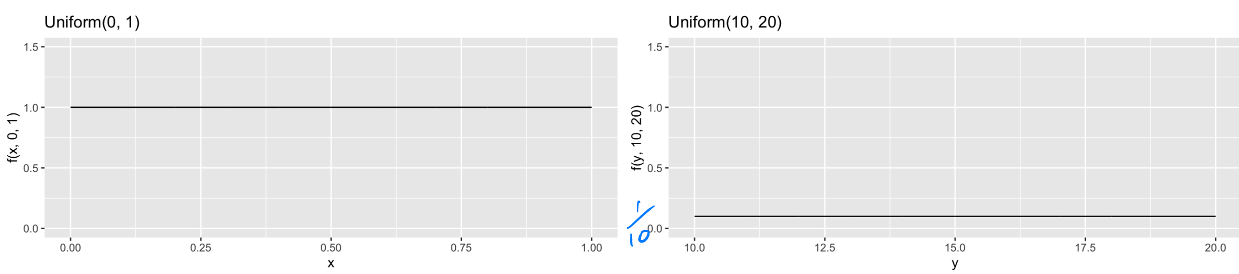
1.2 Monte Carlo Integration

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)")
```



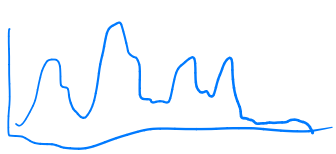
Theory
(exact)

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

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

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

How about some other dsn?



???

Probably can't do this in closed form $\Rightarrow$ need approximation.

1.2.1 Notation

θ = parameter (unknown)

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

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

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

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

estimated versions.

$\hat{E}[\hat{\theta}]$ = estimated mean of dsn 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 number 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 characterizing population. Thing we care about!

To approximate $\theta = \int xf(x)dx$ (EX =), we can (1) obtain an iid random sample $X_1, \ldots, X_n$ from f and then (2) 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,

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

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

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

(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 $E[g(X)]$.

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

The Monte Carlo approximation of θ could then be obtained by

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

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

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 models vs. generalized linear models

$y = X\beta + \varepsilon \quad \varepsilon \sim N(0, \sigma^2) \qquad \hat{\beta} = (X^T X)^{-1} X^T y$ 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 the dsn. Find y s.t. $0.9 = \int_{-\infty}^{y} f(x)\,dx$.

Why the mean?

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

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

So $\hat{\theta}$ is an 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

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

to estimate sampling variance.

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

We can estimate this using a Monte Carlo approach.

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

Sample $X_1, \dots, X_n$ from f

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

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

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

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

$$E\hat\theta = \theta \qquad \mathrm{Var}\,\hat\theta = \frac{\mathrm{Var}\, g(X)}{m}$$

Hence if m is large,

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

can use $\widehat{\mathrm{Var}}\, g(X)$ plugin from above.

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 ~~in~~ 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 region of f
- does not require integrand to be smooth, does not require finite support

1.2.4 Algorithm

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 the estimator s.t. $\hat{\theta}$ approximates $\theta = E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\, dx$.

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 $\theta = \int_0^1 \cancel{g}\, h(x)dx$.

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

② Then $\theta = \int_0^1 g(x)\,dx = \int_0^1 g(x)\cdot \underset{\substack{\uparrow \\ \text{Unif}(0,1)\text{ density.}}}{1}\, dx = Eg(X)$.

③ Sample $X_1, \dots, X_m$ from f

> x <- runif(m)

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

> mean(g(x)).

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

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

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

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

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 to $(0,1)$.

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 $E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x)\, dx$

support of f

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

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

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

$dy = \frac{1}{b-a} dx$

solve for x

$(x \to y)$ $\quad x = a + y(b-a)$

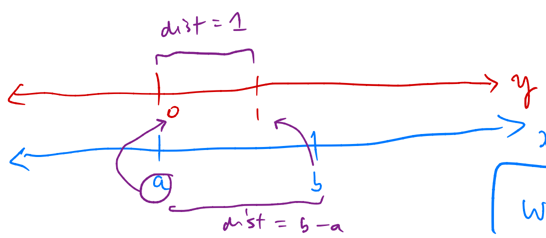
$dx = (b-a)\, dy$

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

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 ~~(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

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

why do this?
because now support of t is $[0, x]$.

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

Method 1

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

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

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

change of variables

then $dy = \frac{1}{x} dt$

$\Rightarrow t = xy$ and $dt = x\, dy$.

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

$\Rightarrow$ want to estimate $\theta = E_Y\left[\frac{1}{\sqrt{2\pi}} x \exp\left(-\frac{(xY)^2}{2}\right)\right]$ 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. $\hat{\Phi}(x) = 0.5 + \frac{1}{m} \sum_{i=1}^{m} \frac{1}{\sqrt{2\pi}} x \exp\left\{-\frac{(xY_i)^2}{2}\right\}$ for $x > 0$.

Method 2

Could generate $X \sim Unif(0, x)$.

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

$$E_Z\left[\mathbb{I}(Z \le x)\right] = \int_{-\infty}^{\infty} \mathbb{I}(Z \le x) f(z)\, dz$$

$$= \int_{-\infty}^{x} f(z)\, dz$$

$$= \Phi(x).$$

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

1. Generate $Z_1, \dots, Z_m \sim N(0,1)$.
2. $\hat{\Phi}(x) = \frac{1}{m} \underbrace{\sum_{i=1}^{m} \mathbb{I}[Z_i \le x]}_{\text{counts \# of } Z_i\text{'s} \le x}$

Notes

1. Can show Method 3 has less bias in the tails and Method 2 has less bias in the center.

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

1.2.5 Inference for MC Estimators

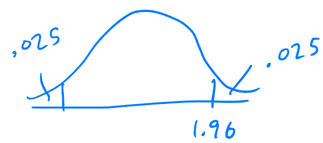
The Central Limit Theorem implies

$$\frac{\hat{\theta} - E(\hat{\theta})}{\sqrt{Var(\hat{\theta})}} \xrightarrow{d} N(0,1).$$

$se(\hat\theta) \longrightarrow \sqrt{Var(\hat\theta)}$

So, we can construct confidence intervals for our estimator

1. 95% CI for $E(\hat\theta)$: $\hat\theta \pm \underset{\text{qnorm}(.975)}{1.96} \sqrt{\hat{Var}(\hat\theta)}$



2. (HW) 95% CI for $\Phi(2)$ $\quad \hat\Phi(2) \pm 1.96 \sqrt{\hat{Var}(\hat\Phi(2))}$

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

recall

Assume $\theta = E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\, dx$
$\sigma^2 = Var[g(X)]$

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

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

$\hat{E}g(X)$

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

estimated variance of the sampling dsn of $\hat\theta$
(a MC estimator)

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

$$se(\hat{\theta}) = \frac{\sigma}{\sqrt{m}}$$

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

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

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

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

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

$$= 10,000$$

So we need $10{,}000 \times m$ samples to acheive $se(\hat{\theta}) = .0001$!

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