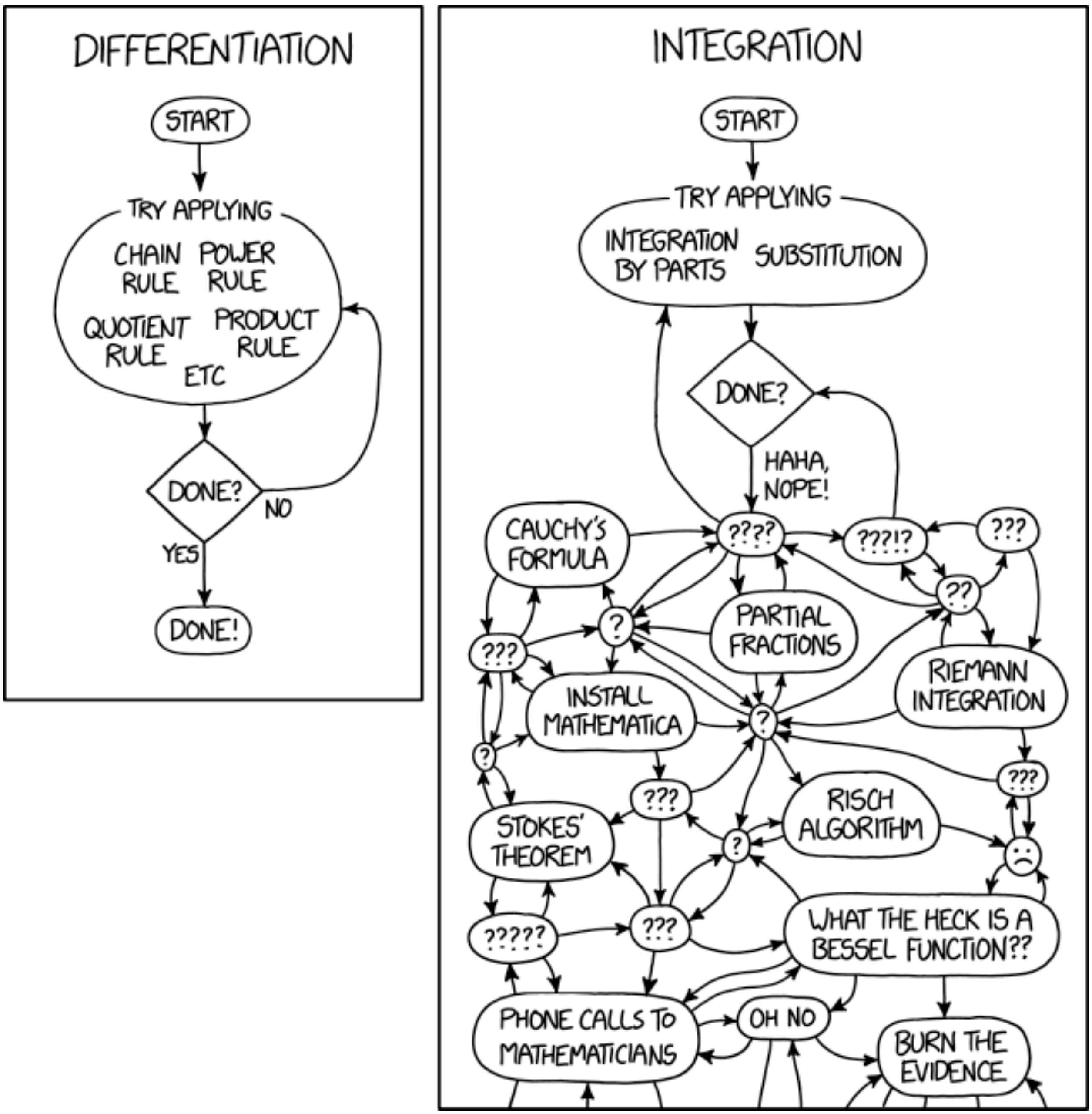
Chapter 6: Monte Carlo Integration

/ 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 expessed 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

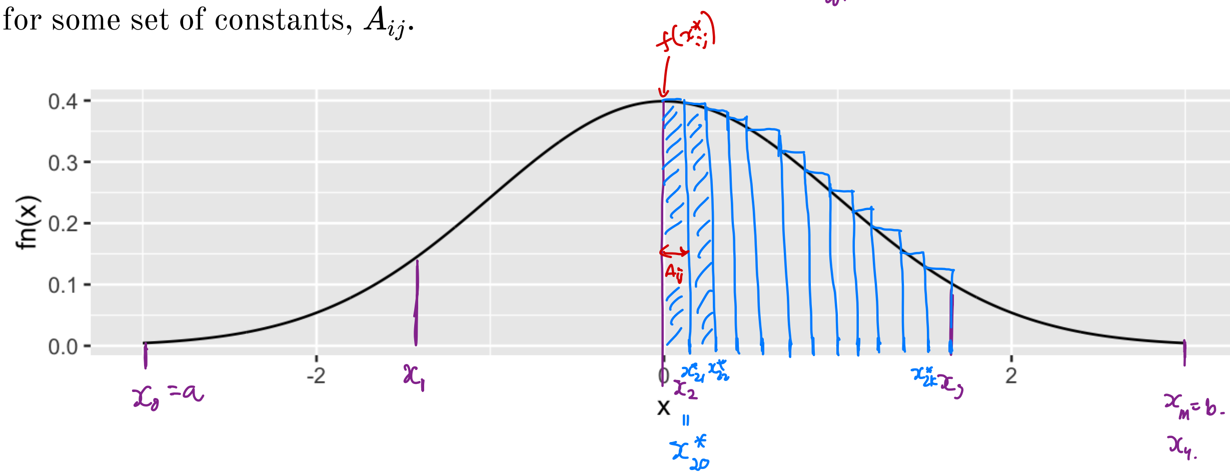
Idea: Approximate $\int_a^b f(x)dx$ via the sum of many polygons under the curve $f(x)$.

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

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

for some set of constants, A_{ij}.



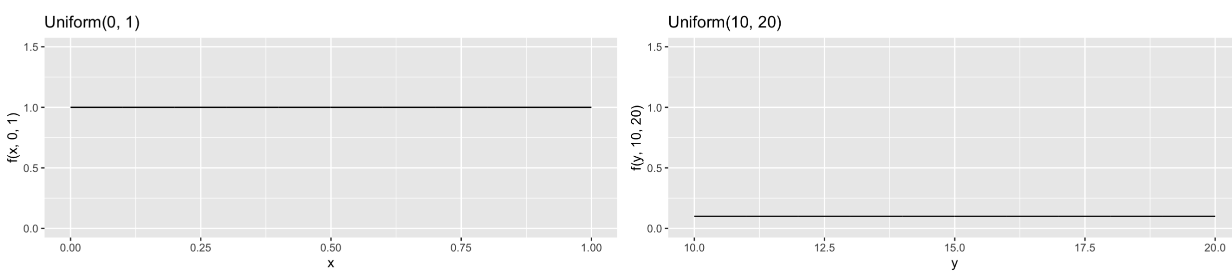
1.2 Monte Carlo Integration

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

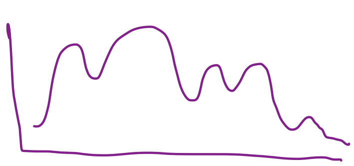
(exact)

$$f(y) = \begin{cases} \frac{1}{10} & 10 \le y \le 20 \\ 0 & \text{o.w.} \end{cases}$$

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

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

What about some other dsn?



?? Probably can't do this in closed form.

$\Rightarrow$ need approximation:

1.2.1 Notation

θ = parameter of interest (unknown).

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

Distribution of $\hat{\theta}$ = Sampling distribution.

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

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

estimated versions

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

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

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

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

1.2.2 Monte Carlo Simulation

What is Monte Carlo simulation?

Computer simulation that generates a large # of sample from a dsn.
The dsn characterizes the population from which a sample is drawn.

(sounds like Ch. 3).

1.2.3 Monte Carlo Integration

parameter characterizing population Thing we care about estimating!

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

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

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,

↳ by randomly generating a sample.

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 for 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.

$Y = X\beta + \varepsilon$, $\varepsilon \sim N(0, \sigma^2)$ $\quad \hat{\beta} = (X^TX)^{-1}X^TY$ in closed form.

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

(B) estimate quantiles of a dsn. $\Leftrightarrow$ Find y s.t. $\int_{-\infty}^{y} f(x)\,dx = .9$.

Why the mean?

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

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

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

and, by the strong law of large numbers,

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

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.

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

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

$\hat{Var}(g(X)) = \hat{E}(v(X))$. $\quad$ $= \hat{Var}(g(X))$.

don't know this! $\theta = E[g(X)]$

$\Rightarrow$ can replace with $\hat{\theta} = \frac{1}{m}\sum_{i=1}^{m} g(X_i)$.

Want to use this to estimate sampling variance of $\hat{\theta}$

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

approximate this using $\hat{Var}(g(X))$

When $Var(g(X))$ and is finite, CLT states..

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

Hence if m is large

we can use $\hat{Var}(g(X))$ plug in from prev. page.

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

we can use this to put confidence limits or error bounds on the MC estimate of any integral θ.

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.

↳ need m to be large.

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 a p-dimensional support region of f.
- does not require integrands to be smooth, does not require finite support.

1.2.4 Algorithm

The approach to finding a Monte Carlo estimator for $\theta = \int h(x)dx$ ~~$\int g(x)f(x)dx$~~ $= E[g(X)],\ X \sim f$ is as follows.

before we write code.

1. Rewrite $\int h(x)dx = \int g(x)f(x)\,dx$ where f is a pdf.
   Select g, f to define θ as an expected value.
2. derive the estimator s.t. $\hat{\theta}$ approximates $\theta = E[g(X)] = \int g(x)f(x)dx$.
   $\hat{\theta} = \frac{1}{m}\sum_{i=1}^{m} g(x_i),\ X_i \overset{iid}{\sim} f$.
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 $(0,1)$ density.

$$\theta = \int_0^1 h(x)dx = \int_0^1 h(x)\cdot \underbrace{1}_{\text{unif}(0,1)\text{ pdf}}\, dx \Rightarrow h(x) = g(x).$$

2. $\Rightarrow \hat{\theta} = \frac{1}{m}\sum_{i=1}^{m} g(x_i) = \frac{1}{m}\sum_{i=1}^{m} h(x_i),\quad X_i \overset{iid}{\sim} Unif(0,1).$

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

```
> x <- runif(m, 0, 1).
```

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

```
> mean(h(x)).
```

(g.)

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

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

$$\theta = \int_a^b h(x)\,dx = \int_a^b \underbrace{h(x)\cdot(b-a)}_{g(x)} \cdot \underbrace{\frac{1}{b-a}}_{f(x)}\,dx.$$

(2) $\hat{\theta} = \frac{1}{m}\sum_{i=1}^m (b-a)h(X_i)$ for $X_i \overset{iid}{\sim} Unif(a,b)$.

(3) Sample $X_1,\ldots,X_m \sim Unif(a,b)$ $\quad$ x <- runif(m, a, b).

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

Another approach:

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

But we care about $\int_a^b h(x)dx$

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

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

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

$\downarrow$ solve for x

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

$$\text{Now}\quad \theta = \int_a^b h(x)\,dx = \int_0^1 \underbrace{h\big(a + y(b-a)\big)(b-a)\cdot 1}_{g(y)\cdot f(y)}\,dy = E[g(Y)],\ Y \sim Unif(0,1).$$

support of Y.

To get a $\hat{\theta}$,

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

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

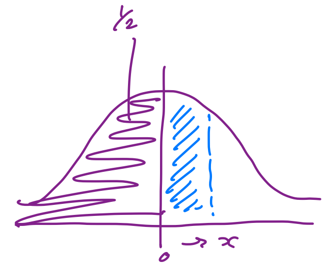
We can use this if limits of integration don't match any density!

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 $x > 0$.

Method 1 Note that for $x>0$, $\Phi(x) = \underbrace{\int_{-\infty}^{0} \phi(x)\,dx}_{>1/2} + \boxed{\int_0^x \phi(x)\,dx}$.

change of variables

Support of $y \sim \text{Unif}(0,1)$ is $(0,1)$. We want a function that maps $t \in [0,x)$ to $y \in [0,1]$.

linear transformation

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

$$\Downarrow$$

$$t = x \cdot y.$$

$$dt = x \cdot dy.$$

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

$g(y)$, $y \sim \text{Unif}(0,1)$.

So a MC estimate could be obtained by:

1. Sample $Y_1, \dots, Y_m \sim \text{Unif}(0,1)$.
2. Compute $\hat{\theta} = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x\cdot Y_i)^2}{2}\right) x + \frac{1}{2}$ for $x > 0$.

Method 2 Could generate $X \sim \text{Unif}(0,x)$.

clever choice of Unif

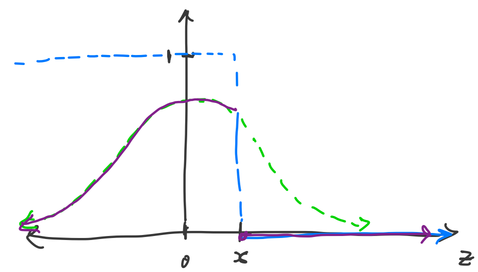
Homework.

Method 3

Indicator function

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

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



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

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

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

$$= \Phi(x)$$

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 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

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

So, we can construct confidence intervals for our MC estimator

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

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

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

Then $Var(\hat{\theta}) = Var\left[\frac{1}{m}\sum_{i=1}^{m} g(X_i)\right] = \frac{1}{m^2}\sum_{i=1}^{m} Var[g(X_i)] = \frac{Var\, g(X_1)}{m} = \frac{\sigma^2}{m}$

(an MC estimator). estimated variance of sampling dsn of $\hat{\theta}$

So $\hat{Var}(\hat{\theta}) = \frac{\hat{\sigma}^2}{m} = \frac{1}{m}\overbrace{\left[\frac{1}{m}\sum_{i=1}^{m}[g(X_i) - \underset{\uparrow\ \hat{E}g(X)}{\hat{\theta}}]^2\right]}^{\hat{\sigma}^2} = \frac{1}{m^2}\sum_{i=1}^{m}(g(X_i) - \hat{\theta})^2$

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

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

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

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

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

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

$$= 10{,}000$$

We would need $10000 \times m$ samples to acheive $se(\hat{\theta}) = .0001$!

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

probably.