MAS472/6004 Computational Inference

Richard Wilkinson

r.d.wilkinson@sheffield.ac.uk

Computational Inference

Simple computational tools for solving hard statistical problems.

- Monte Carlo/simulation
- MC and simulation in frequentist inference
- Random number generation/ simulating from probability distributions
- Further Bayesian computation

Methods implemented via simple programs in R.

Chapter 1: Monte Carlo methods

Problem 1: estimating probabilities

A particular site is being considered for a wind farm. At that site, the log of the wind speed in m/s on day t is known to follow an $AR(2)$ process:

$$\text{log Wind} = Y_t = 0.6Y_{t-1} + 0.4Y_{t-2} + \varepsilon_t, \tag{1}$$

with $\varepsilon_t \sim N(0, 0.01)$.

$\mathbb{P}(A) = \mathbb{E}\,\mathbb{I}_A$ (expected value)

If $Y_1 = Y_2 = 1.5$, what is the **probability** that the wind speed $\exp(Y_t)$ will be below 15 kmh for more than 10 days in a 100 day period?

indicator fnc$^{\text{n}}$.

$$\mathbb{P}(\underbrace{>10 \text{ days} < 15\,\text{km/h}}) = \int \mathbb{I}_{(\sum \mathbb{I}_{y_i < \log 15}) > 10}\, \pi(y)\, dy$$

$$\mathbb{I}_A = \begin{cases} 1 & \text{if event } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

Problem 2: estimating variances

Given a sample of 5 standard normal random variables $X_1, \ldots, X_5$, what is the **variance** of

$$\text{range}(X_i) = \max_i\{X_i\} - \min_i\{X_i\}$$

$$\text{Var}(Z) = \mathbb{E}(Z^2) - (\mathbb{E}Z)^2$$

So calculating variances is also acquires us to compute integrals

Problem 3: Estimating percentiles

The concentration of pollutant at any point in region following release from point source can be describe by the model

$$C(x, y, z) = \frac{Q}{2\pi u_{10}\sigma_z\sigma_y} \exp\left[-\frac{1}{2}\left\{\frac{y^2}{\sigma_y^2} + \frac{(z-h)^2}{\sigma_z^2}\right\}\right], \qquad (2)$$

C: air concentration of pollutant, Q: release rate, u_{10}: wind speed at $10m$ above ground, σ_y, σ_z: diffusion parameters in horizontal and vertical directions, h:release height, (x, y, z): coordinates along wind direction, cross wind and above ground.

Given $Q = 100$, $h = 50$m, but u, σ_z, σ_y uncertain. If

$\log u_{10} \sim N(2, .1)$ $\qquad$ $\log \sigma_y^2 \sim N(10, 0.2)$ $\qquad$ $\log \sigma_z^2 \sim N(5, 0.05)$

What is the **95th percentile** of $C(100, 100, 40)$?

Problem 4:Estimating expectations

A hospital ward has 8 beds

- The number of patients arriving each day is uniformly distributed between 0 and 5 inclusive.
- The length of stay for each patient is also uniformly distributed between 1 and 3 days inclusive.

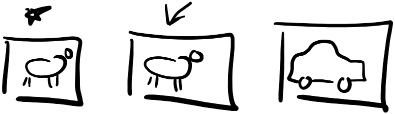
6/6/8

If all 8 beds are free initially, what is the **expected** number of days before there are more patients than beds?

an integral.

Problem 5: Optimal decisions

The Monty Hall Problem



On a game show you are given the choice of three doors.

- Behind one door is a car; behind the others, goats.

The rules of the game are

- After you have chosen a door, the game show host, Monty Hall, opens one of the two remaining doors to reveal a goat.
- You are now asked whether you want to stay with your first choice, or to switch to the other unopened door.

What is the **optimal strategy**? And what is the resulting probability of winning?

We want to compute $\mathbb{P}(\text{Win} \mid \text{switch})$ & $\mathbb{P}(\text{Win} \mid \text{stick})$ $\rightsquigarrow$ integrals.

These 5 problems are all either hard or impossible to tackle analytically. However, the **Monte Carlo method**, can be used to obtain approximate answers to all of them.

> *Monte Carlo methods are a broad class of computational algorithms relying on repeated random sampling to obtain numerical results. They use randomness to solve problems that might be deterministic in principle.*

$\mathbb{E}, \mathbb{P}, \mathrm{Var}$ – non-random/deterministic

Essentially Monte Carlo is just a form of numerical integration.

Some useful results

probability density function

Monte Carlo is primarily used to calculate integrals. For example

pdf

- Expectation of a random variable $X \sim f(\cdot)$, or a function of it

$$\mathbb{E}g(X) = \int g(x)f(x)\mathrm{dx}$$

- Variance

$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2. \tag{3}$$

- Probability $\mathbb{P}(X < a)$ is the expectation of $\mathbb{I}_{X<a}$, the indicator function which is 1 if $X < a$ and otherwise is 0. Then

$$\begin{aligned}\mathbb{P}(X < a) &= 1 \times \mathbb{P}(X < a) + 0 \times \mathbb{P}(X \geq a) \\ &= \mathbb{E}\{\mathbb{I}(X < a)\} = \underline{\int \mathbb{I}_{X<a} f(X)\mathrm{dx}}\end{aligned}$$

Monte Carlo Integration - I

Suppose we are interested in the integral

$$I = \mathbb{E}(g(X)) = \int g(x)f(x)\mathrm{dx} \qquad X_i \overset{iid}{\sim} f(\cdot)$$

Let $X_1, X_2, \ldots, X_n$ be independent random variables with pdf $f(x)$. Then a Monte Carlo approximation to I is

$$\hat{I}_n = \frac{1}{n}\sum_{i=1}^{n} g(X_i). \quad \text{— } \underline{\text{a random quantity}} \tag{4}$$

Example: $\mathbb{E}X = \int x f(x)\,dx \approx \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$

$$\mathbb{P}(X<a) = \int \mathbb{I}_{x<a} \cdot f(x)\,dx \approx \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{X_i<a}$$

Number of elements in the set.

$$= \frac{1}{n} \times \#\{X_i : X_i < a\}$$

$$= \text{observed prop}^{\underline{n}} \text{ of } X_1, \ldots, X_n < a.$$

Monte Carlo Integration - II

Some properties of $\hat{I}$.

(1) $\hat{I}_n$ is an unbiased estimator of I. **Proof:**

$\hat{I}_n$ is unbiased <u>if</u> $\mathbb{E}\hat{I}_n = I$

$$\mathbb{E}\hat{I}_n = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} g(X_i)\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}g(X_i)$$ as $\mathbb{E}$ is a linear operator.

$$= \frac{1}{n}\sum_{i=1}^{n} I = \frac{1}{n} nI = I$$

$\blacksquare$

Monte Carlo Integration - III

i.e. $\hat{I}_n$ is consistent

(2) $\hat{I}_n$ converges to I as $n \to \infty$.

Proof:

The strong law of large numbers says if $\{X_i\}$ are iid random variables with mean μ then $\frac{1}{n}\sum X_i \to \mu$ as $n \to \infty$ with probability 1.

Thus $\frac{1}{n}\sum g(X_i) \to I$ w.p. 1.

Monte Carlo Integration - IV

The SLLN tells us $\hat{I}_n$ converges, but not how fast. It doesn't tell us how large n must be to achieve a certain error.

$$\text{Mean Square Error} = \text{MSE} = \mathrm{Var}(\hat{I}) = \mathbb{E}\left((\hat{I} - \mathbb{E}\hat{I})^2\right) = \mathbb{E}\left((\hat{I} - I)^2\right)$$

$$= \mathbb{E}[(\hat{I}_n - I)^2] = \frac{\sigma^2}{n} \qquad (3)$$

where $\underline{\sigma^2 = \mathbb{V}\mathrm{ar}(g(X))}$. Thus the 'root mean square error' (RMSE) of $\hat{I}_n$ is

$$\boxed{\mathrm{RMSE}(\hat{I}_n) = \frac{\sigma}{\sqrt{n}} = O(n^{-1/2}).}$$

Thus, our estimate is more accurate as $n \to \infty$, and is less accurate when σ^2 is large. σ^2 will usually be unknown, but we can estimate it:

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (g(X_i) - \hat{I}_n)^2 = \text{sample variance of } g(X_1), \ldots, g(X_n)$$

Typo

We call $\cancel{\hat{\sigma}}\ \frac{\hat{\sigma}}{\sqrt{n}}$ the *Monte Carlo standard error*.

Monte Carlo Integration - V

To reduce error by factor of 2 needs 4n ~~to~~ samples.

We write[1]

$$\text{RMSE}(\hat{I}_n) = O(n^{-1/2})$$

to emphasise the rate of convergence of the error with n.

To get 1 - factor of 10 digit more accuracy requires a 100-fold increase in n. A 3-digit improvement would require us to multiply n by 10^6.

Consequently Monte Carlo is not usually suited for problems where we need a very high accuracy. Although the error rate is low (the RMSE decreases slowly with n), it has the nice properties that the RMSE

- does not depend on $d = \dim(x)$
- does not depend on the smoothness of f

– Unlike most numerical integration methods.

Consequently Monte Carlo is very competitive in high dimensional problems that are not smooth.

Monte Carlo Integration - VI

- sum of iid r.v.s

$\frac{1}{n}\sum g(X_i)$

$g(X_i)$ has mean I & variance σ^2

In addition to the rate of convergence, the central limit theorem tells us the asymptotic[2] distribution of $\hat{I}_n$

$= (\hat{I} - E\hat{I}) / [Var(g(X))/n]^{1/2}$

(4)
$$\frac{\sqrt{n}(\hat{I}_n - I)}{\sigma} \to N(0,1) \text{ in distribution as } n \to \infty$$

Informally, $\boxed{\hat{I}_n \text{ is approximately } N(I, \frac{\sigma^2}{n})}$ for large n.

This allows us to calculate confidence intervals for I.

See the R code on MOLE.

A 95% approximate CI for I is $\hat{I} \pm 1.96 \frac{\hat{\sigma}}{\sqrt{n}}$

- If we require $E\{f(X)\}$, random observations from distribution of $f(X)$ can be generated by generating $X_1, \ldots, X_n$ from distribution of X, and then evaluating $f(X_1), \ldots, f(X_n)$.
- Preceding results can be applied when estimating variances or probabilities of events.
- Percentiles estimated by taking the sample percentile from the generated sample of values $X_1, \ldots, X_n$.
- We expect the estimate to be more accurate as n increases. Determining a percentile is equivalent to inverting a CDF. If wish to know the 95th percentile, we must find ν such that

$$P(X \leq \nu) = 0.95, \tag{5}$$

e.g. to compute median of $f(X)$, find 50^{th} percentile of $f(X_1), \ldots, f(X_n)$

Monte Carlo solutions to the example problems

Question 1

Define E: the event that in 100 days the wind speed is below 15kmh for more than 10 days.
To estimate $\mathbb{P}(E)$, generate lots of individual time series, and count proportion of series in which E occurs

FOR $i = 1, \dots, N$

1. Generate ith realisation of the time series process:
   For $t = 3, 4, \dots, 100$:
   - Set $Y_t \leftarrow 0.6Y_{t-1} + 0.4Y_{t-2} + N(0, 0.01)$

   Generate random sequence $Y_1, \dots, Y_{150}$
2. Count number of elements of $\{Y_1, \dots, Y_{100}\}$ less than $\log 15 = 4.167$:
   - Set $X_i \leftarrow \sum_{t=1}^{100} I\{Y_t < 4.167\}$

   " log 15
3. Determine if event E has occured for time series i:
   - Set $E_i \leftarrow I\{X_i > 10\}$

   $\mathbb{I}$

   Calculate $g(X)$
4. Estimate $\mathbb{P}(E)$ by $\frac{1}{N}\sum_{i=1}^{N} E_i$

Question 2

Define Z to be the difference between max and min of 5 standard normal random variables. Estimate the variance of Z.

$$Z = \max_{i=1,\dots,5} X_i - \min_{i=1,\dots,5} X_i$$

$$X_1, \dots, X_5 \sim N(0,1)$$

$$Var(Z) = \mathbb{E}Z^2 - (\mathbb{E}Z)^2$$

For $n = 1, \dots, N$

1) Simulate $X_1^{(n)}, \dots, X_5^{(n)} \sim N(0,1)$
2) Calculate $Z^{(n)} = \max_{i=1,\dots,5} X_i^{(n)} - \min_{i=1,\dots,5} X_i^{(n)}$

Estimate $Var(Z)$ by $\frac{1}{N}\sum Z^{(n)2} - \left(\frac{1}{N}\sum Z^{(n)}\right)^2$

Question 3

Transformation of a random variable:
Given random variables $X_1, \ldots, X_d$ we want to know the distribution of $Y = f(X_1, \ldots, X_d)$.

- The Monte Carlo method can be used
  - Sample unknown inputs from their distributions,
  - evaluate the function to obtain output value from its distribution.
- Given suitably large sample, 95th percentile from distribution of $C(100, 100, 40)$ can be estimated by the 95th percentile from sample of simulated values of $C(100, 100, 40)$.

For $i = 1, 2, \ldots, N$:

1. Sample a set of input values:
   - Sample $u_{10,i}$ from $\log N(2, .1)$
   - Sample $\sigma^2_{y,i}$ from $\log N(10, 0.2)$
   - Sample $\sigma^2_{z,i}$ from $\log N(5, 0.05)$
2. Evaluate the model output C_i:
   - Set $C_i \leftarrow \frac{100}{2\pi u_{10,i}\sigma_{z,i}\sigma_{y,i}} \exp\left[-\frac{1}{2}\left\{\frac{40^2}{\sigma^2_{y,i}} + \frac{100}{\sigma^2_{z,i}}\right\}\right]$
3. Return the 95th percentile of $C_1, C_2, \ldots, C_N$.

Question 4

- Define W to be the number of days before the first patient arrives to find no available beds.
- The question has asked us to give $E(W)$.
- If we can generate $W_1, \ldots, W_n$ from the distribution of W, we can then estimate $E(W)$ by $\bar{W}$.

See the R code on MOLE for a way to simulate this process.

The Monty Hall Problem

- Simulate N separate games by randomly letting x take values in $\{1, 2, 3\}$ with equal probability. x represents which door the car is behind.
- Simulate the contestant randomly picking a door by choosing a value y in $\{1, 2, 3\}$ (it doesn't matter how we do this, we can always choose 1 if you like, the results are the same).
- Now the game show host will open the door which hasn't been picked that contains a goat. For each of the N games, record the success of the two strategies
  1. stick with choice y
  2. change to the unopened door.
- Calculate the success rate for each strategy.

Example 1

Consider the probability p that a standard normal random variable will lie in the interval $[0, 1]$. This can be written as an integral

$$p = \int_0^1 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx. = \mathbb{P}\left(0 \le N(0,1) \le 1\right) \tag{6}$$

Two methods for estimating/evaluating this probability are

1. numerical integration/quadrature, e.g., trapezium rule, Simpson's rule etc
2. given a sample of standard normal random variables $Z_1, \ldots, Z_n$, look at the proportion of Z_is occurring in the interval $[0, 1]$.

$$p = \mathbb{E}_{X \sim N(0,1)}\left(\mathbb{1}_{0 \le X \le 1}\right) = \int_{-\infty}^{\infty} \mathbb{1}_{0 \le x \le 1} \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

where $X_i \sim N(0,1)$

$$\hat{p} = \frac{1}{n} \sum \mathbb{1}_{0 \le X_i \le 1}$$

Example 1: An alternative method

1. Y is a RV with $f(Y)$ any function of Y. To generate a random value from the distribution of $f(Y)$, generate a random Y from the distribution of Y, and then evaluate $f(Y)$.
2. Providing $\mathbb{E}\{f(Y)\}$ exists, given a sample $f(Y_1), \ldots, f(Y_n)$,

$$\frac{1}{n}\sum_{i=1}^{n} f(Y_i)$$

   is an unbiased estimator of $\mathbb{E}\{f(Y)\}$.
3. Let X be a random variable with a $U[0,1]$ distribution. For an arbitrary function $f(X)$, what is the expectation of $f(X)$?

$$\mathbb{E}_{X\sim U[0,1]}\big(f(X)\big) = \int f(x)\, \underbrace{g(x)}_{\text{pdf of } U[0,1]}\, dx \qquad g(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$= \int_0^1 f(x)\, dx$$

4 Now choose f to be the function $f(X) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{X^2}{2}\right)$. Then if $X \sim U[0,1]$

$$\mathbb{E}\{f(X)\} = \int_0^1 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) 1\mathrm{dx} \tag{8}$$

Given a sample $f(X_1), \ldots, f(X_n)$ from the distribution of $f(X)$, we can estimate $E\{f(X)\}$ by the *unbiased* **Monte Carlo** estimator $\hat{p}$

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{n} f(X_i), \tag{9}$$

where X_i is drawn randomly from the $U[0,1]$ distribution.

Key idea

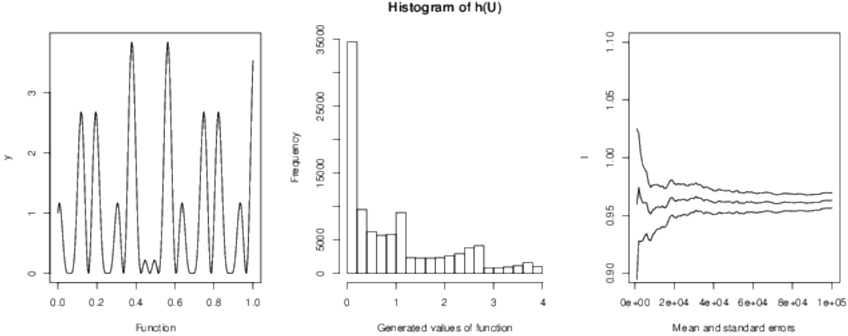
re-express the integral of interest (6) as an *expectation*.

Example 2

Consider the integral $\int_0^1 h(x)\mathrm{dx}$ where

$$h(x) = [\cos(50x) + \sin(20x)]^2$$

Generate $X_1, \ldots, X_n$ from $U[0,1]$ and estimate with $\hat{I}_n = \frac{1}{n}\sum h(X_i)$. $X_i \sim U[0,1]$



The general framework

$$R = \int f(x)dx \tag{10}$$

Let $g(x)$ be some density function that is easy to sample from. How do we re-write (10) as the expectation of a function of a random variable X with density function $g(x)$?

$$R = \int \frac{f(x)}{g(x)} g(x) dx = \mathbb{E}_{X \sim g} h(X)$$

where

$$h(x) = \frac{f(x)}{g(x)}$$

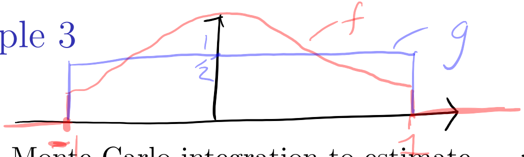
So we now have $R = E\{h(X)\}$, where X has the density function $g(x)$. If we now sample $X_1, \ldots, X_n$ from $g(x)$, then evaluate $h(X_1), \ldots, h(X_n)$,

$$\boxed{\hat{R} = \frac{1}{n}\sum_{i=1}^{n} h(X_i)} \tag{12}$$

— a Monte Carlo estimator of R.

is an unbiased estimator of R.

Example 3



Use Monte Carlo integration to estimate $f(x) = e^{-x^2}$

$$R = \int_{-1}^{1} \exp(-x^2)dx. = \int \frac{f(x)}{g(x)} g(x)\, dx \tag{13}$$

$$g(x) = \begin{cases} \frac{1}{2} & -1 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

We'll consider two different choices for $g(x)$.

1. A uniform density on $[-1, 1]$: $g(x) = 0.5$ for $x \in [-1, 1]$.
   We sample $X_1, \ldots, X_n$ from $U[-1, 1]$, and estimate R by

$$\boxed{\hat{R} = \frac{1}{n}\sum_{i=1}^{n} \frac{\exp(-X_i^2)}{g(X_i)}} = \frac{1}{n}\sum_{i=1}^{n} 2\exp(-X_i^2). \tag{14}$$

$$R = \int_{-1}^{1} \frac{e^{-x^2}}{1/2} \cdot \frac{1}{2}\, dx = \mathbb{E}_{X \sim g}\, e^{-X^2} \approx \frac{1}{n}\sum_{i=1}^{n} \frac{e^{-X_i^2}}{1/2}$$

where $X_i \sim U[-1, 1]$

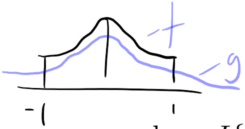
2 A normal density function $N(0, 0.5)$. $g(x) = \frac{1}{\sqrt{\pi}} e^{-x^2}$

Note: sampled value X from $g(x)$ not constrained to lie in $[-1, 1]$.

Re-write R as

$R = \int_{-1}^{1} \frac{e^{-x^2}}{\frac{1}{\sqrt{\pi}} e^{-x^2}} \cdot \frac{1}{\sqrt{\pi}} e^{-x^2} dx = \int_{-\infty}^{\infty} \sqrt{\pi}\, g(x)\, \mathbb{1}_{-1<x<1}\, dx$

$= \mathbb{E}\left(\sqrt{\pi}\, \mathbb{1}_{-1<X<1}\right)$



$$R = \int_{-\infty}^{\infty} I\{-1 \le x \le 1\} \exp(-x^2) dx, \tag{15}$$

$\approx \frac{1}{n} \sum_{i=1}^{n} \sqrt{\pi}\, \mathbb{1}_{-1 \le X_i \le 1}$

where $I\{\}$ denotes the indicator function.

We now sample $X_1, \ldots, X_n$ from $N(0, 0.5)$ and estimate R by

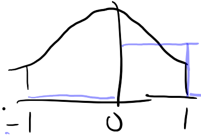
where $X_i \sim N(0, \frac{1}{2})$

$$\hat{R} = \frac{1}{n} \sum_{i=1}^{n} \frac{I\{-1 \le X_i \le 1\} \exp(-X_i^2)}{g(X_i)} = \frac{1}{n} \sum_{i=1}^{n} \pi^{1/2} I\{-1 \le X_i \le \tag{16}$$

Key idea

$g(x)$ needs to mimic $f(x)$ as closely as possible. Consider again $R = \int_{-1}^{1} \exp(-x^2)\mathrm{dx}$.

Two terrible choices of g:



1. A uniform density on $[0,1]$: $g(x) = 1$ for $x \in [0,1]$.

   not a valid choice.

$$R = \int_{-\infty}^{\infty} I\{-1 \le x \le 1\} \exp(-x^2) dx, \tag{17}$$

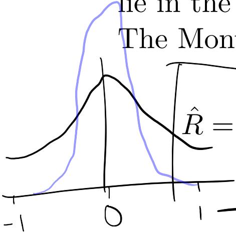
   For $x \in [-1,0)$, we have $f(x) > 0$ and $g(x) = 0$. Must have $g(x) > 0$ for all x where $f(x) > 0$.

   $\mathrm{supp}(f) \not\subseteq \mathrm{supp}(g)$!

2. A normal density $N(0, 0.09)$.
   In this case, we have $g(x) > 0$ for $x \in [-1,1]$, but we when we sample x from g, we expect around 95% of the values to lie in the range $(-0.6, 6)$.

   $\mathrm{supp}(f) \subseteq \mathrm{supp}(g)$ - So g is valid but unwise.

   The Monte Carlo estimate of R is given by



$$\boxed{\hat{R} = \frac{1}{n} \sum_{i=1}^{n} I\{-1 \le X_i \le 1\} \frac{\exp(-X_i^2)\sqrt{0.18\pi}}{\exp(-5.56 X_i^2)}.} \tag{18}$$

Exercise

Convergence

$\equiv$ $\mathrm{supp}(f) \subseteq \mathrm{supp}(g)$ then

- Provided $f(x) > 0 \Rightarrow g(x) > 0$, $\hat{R}$ will converge to R as $n \to \infty$.
- Use the central limit theorem to derive a confidence interval for $\hat{R}$:

$$\hat{R} \sim N\left(R, \frac{\sigma^2}{n}\right), \quad \sigma^2 = \mathrm{Var}\, h(X) \tag{19}$$

$X \sim g(\cdot)$

where we estimate σ^2 by

So we want to choose g so that σ^2 is small.

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left\{ h(X_i) - \hat{R} \right\}^2 \tag{20}$$

- We can then report the confidence interval as

$$\hat{R} \pm Z_{1-\alpha/2} \sqrt{\hat{\sigma}^2 / n}, \tag{21}$$

3

- Estimates of σ^2 in the example: $U[-1,1] : 0.16$, $N(0, 0.5) : 0.42$, $N(0, 0.09) : 6.81$.

Comparison of Monte Carlo with numerical integration

Mid-ordinate rule

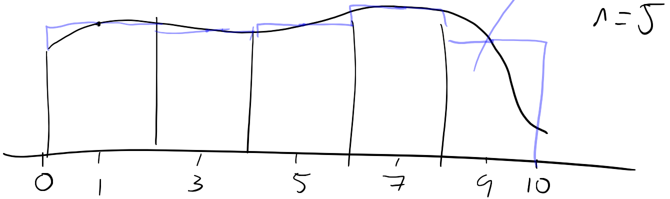
Consider finding $I = \int_0^1 f(x)\,\mathrm{dx}$. There are many different numerical integration schemes we might use.

For example, the mid-ordinate rule is one of the simplest methods, and approximates I by a sum

$$\boxed{\tilde{I}_n = \frac{1}{n}\sum_{i=1}^{n} f(x_i),}$$

$f(x_i)h$

The points $x_i = (i - \frac{1}{2})h$ are equally spaced at intervals of $h = 1/n$.



Comparison of MC with numerical integration II

Mid-ordinate rule error analysis

For smooth 1-d functions the error rates for quadrature rules can be much better than Monte Carlo

For example, if $f : [0,1] \to \mathbb{R}$ and $f''(x)$ is continuous, then

$$|I - \tilde{I}_n| \leq \frac{1}{24n^2} \max_{0 \leq x \leq 1} |f''(x)|$$

For Monte Carlo

$RMSE(\hat{I}_{MC}) = O\left(\frac{1}{\sqrt{n}}\right)$

So

$$\boxed{\mathrm{RMSE}(\tilde{I}) = O(n^{-2})}$$

i.e., it is a second order method. Other rules achieve higher errror rates. For example, Simpson's rule is a fourth order method.

This is much faster than Monte Carlo: to get an extra digit of accuracy we only need multiply n by a factor of $\sqrt{10} = 3.2$

Comparison of MC with numerical integration III

$\int f(x)\,dx$

Curse of dimensionality

1 dimensional problems

Classial quadrature methods work well for smooth 1d problems. But for d-dimensional integrals we have a problem. Suppose

$$I = \int_0^1 \int_0^1 \dots \int_0^1 f(x_1, \dots, x_d) \mathrm{d}x_1 \dots \mathrm{d}x_d$$

We can use the same N point 1-d quadrature rules on each of the d integrals.

This uses $n = N^d$ evaluations of f. The 1d mid-ordinate rule has error $O(N^{-2})$, so the d-dimensional mid-ordinate rule has error

$$|I - \tilde{I}| = O(N^{-2}) = O(n^{-2/d}) = O\left(\frac{1}{n^{2/d}}\right)$$

For $d = 4$ this is the same as Monte Carlo. For larger d it is worse.

In addition, we require f to be smooth ($f''(x)$ to be continuous) for the method to work well.

Monte Carlo has the same $O(n^{-1/2})$ error rate regardless of $\dim(x)$ or $f''(x)$