We will get a "recipe-style" overview of a hypothesis test. A fully formal and definite understanding of hypothesis testing is beyond current scope.

Outcomes of a Hypothesis Test.

"State of Nature"	Do not reject H_0	Reject H_0
H_0 is "true"	OK	Type I error
H_1 is "true"	Type II error	OK.

So, we want to reject H_0 when H_0 is true with a small probability (minimize Type I error). Similarly, we want to minimize Type II error probability as well.

P-value is one way to conduct desirable hypothesis test.

Evidence scale against the null hypothesis H_0 in terms of p-value: [Note: p-value is not the prob. that H_0 is true, p-value $\neq P(H_0 | \underset{\text{data}}{x})$)

P-value range.	Evidence.
$(0, 0.01]$	Very strong evidence against H_0
$(0.01, 0.05]$	strong evidence " "
$(0.05, 0.1]$	weak " " "
$(0.1, 1)$	little or no " " "

Let us look at the details of a hypothesis test called Permutation Test now.

Permutation Test is a non-parametric exact method for testing whether two distributions are the same based on two sets of samples from them.

because we do not impose any parametric assumptions.

works for any sample size.

Formally, we suppose that:

$$X_1, X_2, \cdots, X_m \overset{IID}{\sim} F^* \quad \text{and} \quad X_{m+1}, X_{m+2}, \cdots, X_{m+n} \overset{IID}{\sim} G^*$$

where $F^*, G^* \in \{\text{all DFs}\}$.

Now, consider the following hypothesis test:

$$H_0: F^* = G^* \qquad \text{Versus} \qquad H_1: F^* \neq G^*$$

Let our test statistic $T(X_1, \cdots, X_m, X_{m+1}, \cdots, X_{m+n})$ be some "sensible" one.

For instance, let us cook-up a test statistic that gets large when F^* is more and more different from G^*.

Say,
$$T = \text{abs}\left(\frac{1}{m}\sum_{i=1}^{m} X_i - \frac{1}{n}\sum_{i=m+1}^{m+n} X_i \right)$$

Aside

Remark:

Realize that this T will not get larger if F^* and G^* have the same population mean but different variances. Thus, it is important to realise the limitations of your test statistic.

The idea of a permutation test:

1) Let $N := m+n$ be the pooled sample size and consider all $N!$ permutations of the observed data

$$x_{obs} := (x_1, x_2, \dots, x_m, x_{m+1}, x_{m+2}, \dots, x_{\underbrace{m+n}_{=N}})$$

2) For each permutation of the data, compute the test statistic T(permuted data x) and denote these $N!$ values of T by:

$$t_1, t_2, t_3, \dots, t_{N!}$$

3) Under H_0, i.e., if H_0 were True,

$$H_0: X_1, X_2, \dots, X_m, X_{m+1}, \dots, X_{m+n} \overset{IID}{\sim} F^* = G^*$$

then each permutation of $x_{obs} = (x_1, \dots, x_m, x_{m+1}, \dots, x_{m+n})$ has the same joint density $\prod_{i=1}^{m+n} f(x_i)$

where $f(x_i) = dF^*(x_i) = dG^*(x_i)$

Therefore, the transformation of the data by our statistic T also has the same probability over the different values in $\{t_1, t_2, \dots, t_{N!}\}$

This is the <u>Permutation Distribution</u>

$$\mathbf{P}_0 \sim \text{Discrete Uniform Distribution on } \{t_1, \dots, t_{N!}\}$$

4) Let $t_{obs} := T(x_{obs})$, our <u>observed test statistic</u>

5) Assuming we reject H_0 when T is large, then the p-value $= \mathbb{P}_0(T \geq t_{obs})$

eg:- Guo-shen Experiment with coarse Venus shell diameters (2007 Stat 218 Project) on either side of New Brighton Pier.

Let us just take two shells from left of pier and one shell from right of pier.

$$H_0: X_1, X_2, X_3 \overset{IID}{\sim} F^* = G^*, \quad H_1: X_1, X_2 \overset{IID}{\sim} F^*$$
$$X_3 \overset{IID}{\sim} G^*, \quad F^* \neq G^*$$

$$T(X_1, X_2, X_3) = \text{abs}\left(\frac{1}{2}\sum_{i=1}^{2} X_i - \frac{1}{1}\sum_{i=2+1}^{3} X_i\right)$$

$$= \text{abs}\left(\frac{X_1 + X_2}{2} - \frac{X_3}{1}\right).$$

Observed data & test statistic:

$$x_{obs} = (x_1, x_2, x_3) = (52, 54, 58)$$
$$t_{obs} = \text{abs}\left(\frac{52+54}{2} - \frac{58}{1}\right) = \text{abs}(53-58) = \text{abs}(-5) = 5$$

Tabulate the permutation probabilities.

Permutation	t	$\mathbb{P}_0(T=t)$
(52, 54, 58)	5	1/6
(54, 52, 58)	5	1/6
(52, 58, 54)	1	1/6
(58, 52, 54)	1	1/6
(58, 54, 52)	4	1/6
(54, 58, 52)	4	1/6

$$
\begin{aligned}
\text{p-value} &= \mathbb{P}_0\,(T \geq t_{obs}) \\
&= \mathbb{P}_0\,(T \geq 5) \\
&= \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3} \approx 0.333.
\end{aligned}
$$

Therefore, there is little or no evidence against H_0.

———x———

<u>Remark.</u> The lowest possible p-value for pooled sample size of $N = m+n$ is $\frac{1}{N!}$.

Do you see why this is the case?

So, for previous example with $N = m+n = 2+1 = 3$ the smallest possible p-value $= \frac{1}{3!} = \frac{1}{3\times 2} = \frac{1}{6}$. Therefore, the pooled sample size N should be large enough for p-value to be allowed to range below $0.01 = \frac{1}{100}$ and give the "possibility of very strong evidence against H_0". Thus, since $\underbrace{5\times\underbrace{4\times\underbrace{3\times\underbrace{2\times 1}_{2}}_{6}}_{24}}_{120}$ it is good to have $N \geq 5$.

But, then the tabulation gets large as we have to work through $N!$ possible permutations.

We can use <u>Monte Carlo integration</u> to approximate p-value.

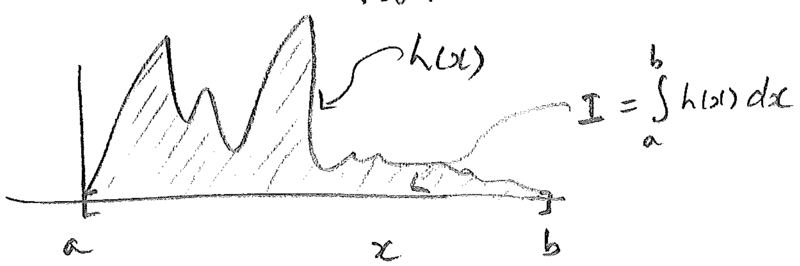
This is our next topic.

Basic Monte Carlo Integration (BMC)

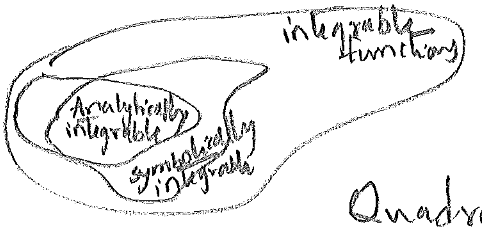
Suppose we want to evaluate the integral:

$$I = \int_a^b h(x)\,dx$$

for some function h.



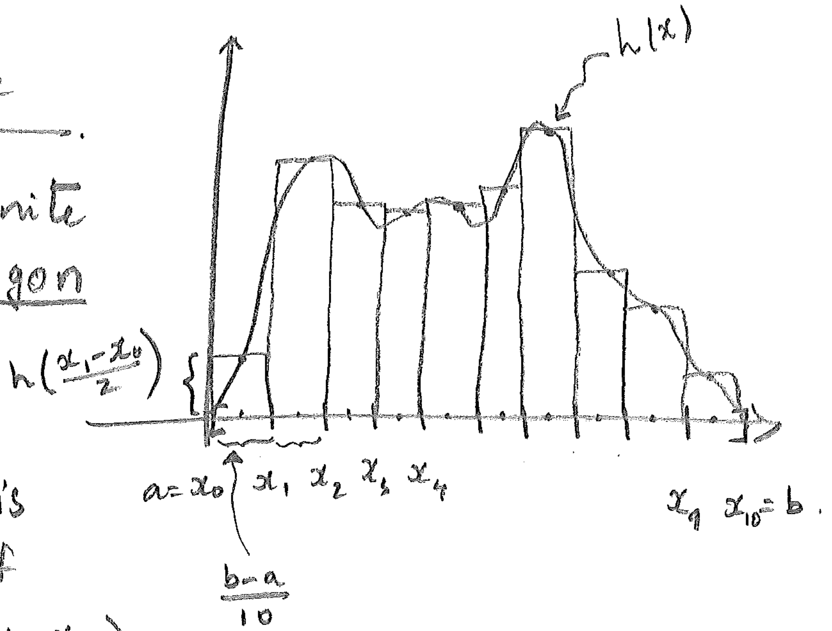
options : (i) Quadrature or Numerical Integration
(ii) Analytically by hand.
(iii) Symbolically (Maple / SAGE)



Quadrature

Idea is to use a finite "Riemann Sum" of polygon areas.

"many sided"



A simple polygon is a rectangle with base of width $\frac{b-a}{n}$ and height $h\left(\frac{x_i-x_{i-1}}{2}\right)$ from a equal sized partition of

width of each interval

mid-point

$[a,b]$ into n intervals: $[x_0,x_1],\dots,[x_{n-1},x_n]$.

This is called the mid-point Rectangle Quadrature Rule.

Then,
we can approximate $I = \int_a^b h(x)\,dx$ with a large n due to the limit:

$$\lim_{\substack{\#\text{ bins or intervals } n \to \infty \\ \text{or bin-width} \to 0}} \sum_{i=1}^{n} \left(\begin{array}{c}\text{interval width}\\ \text{for interval } i\end{array}\right) * \left(\text{height } h \text{ at } \begin{array}{c}\text{midpoint of}\\ \text{interval } i\end{array}\right)$$

<u>Variants on Mid-point Rectangle Rule</u> (with other polygons).

(i) Trapezoid Rule.
(ii) Simpson's Rule ($\frac{1}{3}$ trapezoid Rule + $\frac{2}{3}$ Midpoint Rule).

All these methods partition the domain $[a,b]$ into equal-sized intervals or bins.
Therefore, the suffer from the '<u>curse of dimensionality</u>' when we want to extend these methods to integrals over higher dimensional domains.

Suppose we have to evaluate

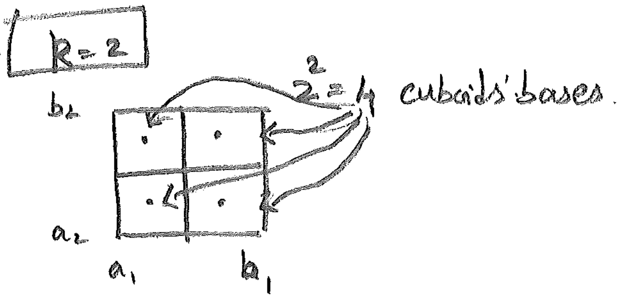
$$I = \int_{a_1}^{b_1}\int_{a_2}^{b_2} h(x_1, x_2)\, dx_2\, dx_1$$

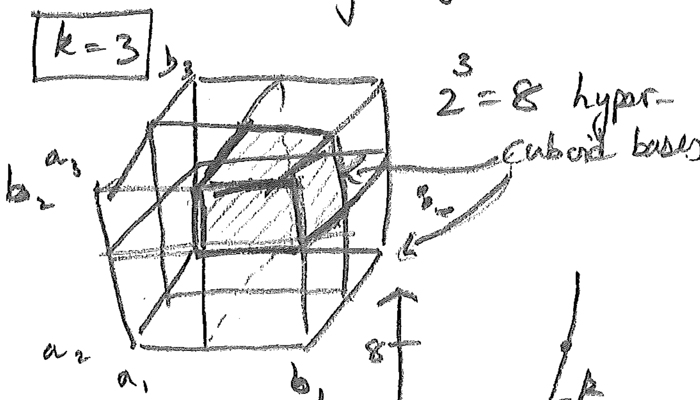
Or more generally, we want to evaluate

$$I = \int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_k}^{b_k} h(x_1, \ldots, x_k)\, dx_k\, dx_{k-1} \cdots dx_2\, dx_1$$

<u>SO</u> even if we were to chop each dimension only once to make two intervals in that dimension, you will

have to evaluate the volume of 2^k many "hyper-cuboids"

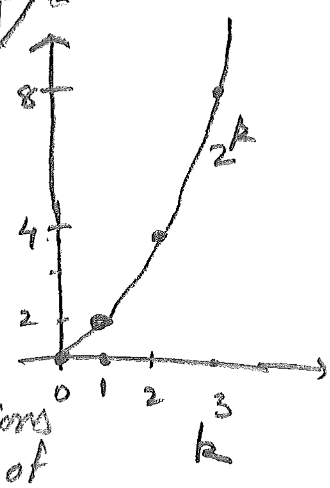




This is the curse of dimensionality since

2^k grows exponentially in k

\# of dimensions of the domain of integration



The problem of evaluating a multidimensional integral I becomes possible if we view it as a statistical estimation problem.

Let us begin by writing

$$I = \int_a^b h(x)\,dx = \int_a^b w(x) f(x)\,dx =: E(w(X)),$$

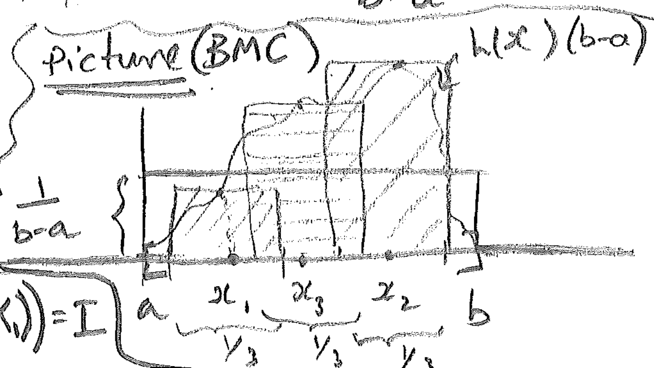
this expectation is with respect to RV X with pdf $f(x)$.

where $w(x) = \frac{h(x)}{f(x)} = h(x)(b-a)$; $f(x) = \frac{1}{b-a}$,

and $X \sim \text{Uniform}(a,b)$

Formally $X_1, \ldots, X_n \overset{\text{IID}}{\sim} \text{Uniform}(a,b)$

Then by WLLN, $\hat{I} = \frac{1}{n}\sum_{i=1}^{n} W(X_i) \xrightarrow{P} E(W(X_1)) = I$



and due to CLT we can compute the estimated standard error of the estimate:

$$\hat{se} = \frac{s}{\sqrt{n}}, \quad s^2 = \frac{\sum_{i=1}^{n} (w(x_i) - \hat{I})^2}{n-1}$$

← sample variance.

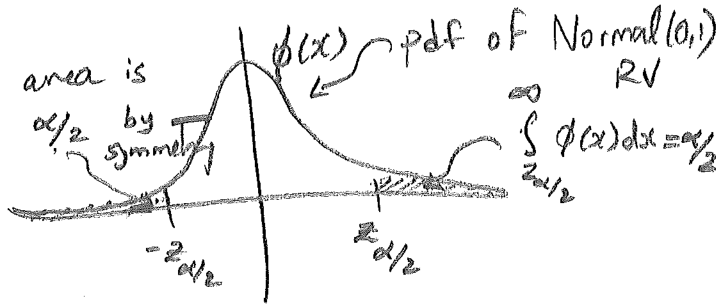
Thus, due to WLLN and CLT we get

$\hat{I} = \frac{1}{n}\sum_{i=1}^{n} w(x_i)$ as a point estimate of I

and $\hat{I} \pm z_{\alpha/2}\, \hat{se}$ as a $1-\alpha$ confidence interval of I.

what is $z_{\alpha/2}$?

$z_{\alpha/2} \cong 1.96$ if $\alpha = 0.05$



Example:

$$I = \int_0^1 x^3\,dx = \frac{1}{4}.$$

"Pretend" we don't know the value of I and use BMC integration to estimate I.

Based on $n = 10{,}000$ samples

$x_1, \ldots, x_{10{,}000} \overset{IID}{\sim}$ Uniform $(0,1)$

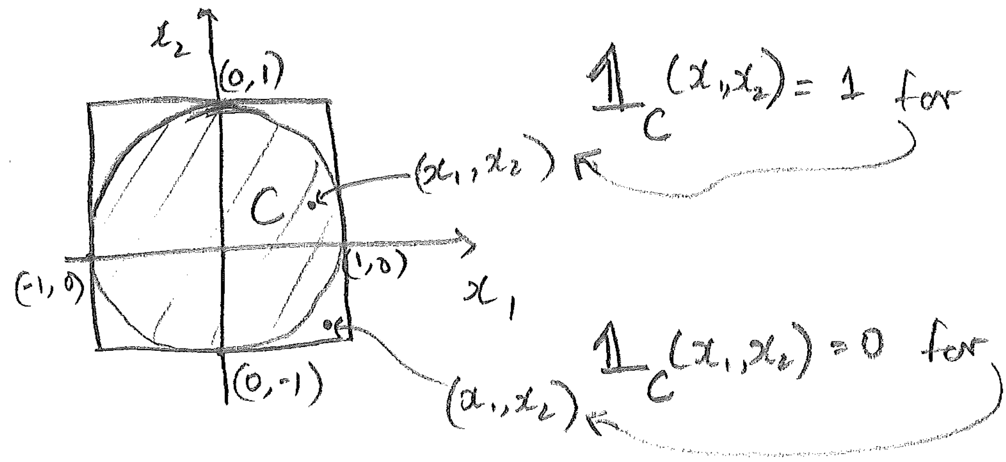
$$\hat{I} = \frac{1}{10000}\sum_{i=1}^{10{,}000} x_i^3$$

Exercise Do this in SAGE

Answer: $\hat{I} = 0.248$ $\quad \hat{se} = 0.0028$

95% Conf. Interval is $[0.248 - 0.0056, 0.248 + 0.0056]$ and it has trapped $I = 1/4$.

example (2D integral)



Suppose I am interested in the area of the circle centred at the origin:

$$I = \int_{-1}^{1}\int_{-1}^{1} \mathbb{1}_C(x_1,x_2)\, dx_1 dx_2 = \int_{-1}^{1}\int_{-1}^{1} \underbrace{\left(\mathbb{1}_C(x_1,x_2)\, 4\right)}_{w(x_1,x_2)} \underbrace{\frac{1}{4}}_{f(x_1,x_2)}\, dx_1 dx_2$$

Basic idea:

$$X_1, X_2, \ldots, X_n \overset{IID}{\sim} \text{Uniform}([-1,1]\times[-1,1])$$

(X_{11}, X_{12}), (X_{21}, X_{22}), ..., (X_{n1}, X_{n2})

Random vectors uniformly distributed in the square $[-1,1]\times[-1,1]$.

has PDF $f(x_1,x_2) = \frac{1}{4}$ since $[-1,1]\times[-1,1]$ has area $(1-(-1) * (1-(-1))) = 2 * 2 = 4$.

Exercise: If you understand this example and can conduct point and conf. set estimation of I and implement it in SAGE to see if you get closer to the right $I = \pi 1^2 = \pi$, then you have done well in this course. Good luck! — Raaz