Analysis and documentation of practice 5A: Numerical integration with Simpson's rule and the t distribution

Jaime Ivan Avila Muñoz

November 2022

Abstract

During the development of this document the details to be highlighted will be explained based on practice 5A. The steps for the calculation of the numerical integration with Simpson's rule as well as for the T-distribution will be explained, together with their respective tests and highlights during their application.

Keywords: Simpson´s rule, Numerical Integration, T distribution, formula.

1 Introduction

1.1 Objectives

Calculation of the numerical integration applying Simpson's rule, together with the T distribution, checking the results with the tests suggested in the "issues" located in the Hampty5A repository in github.

1.2 Problem Statement

Within numerical analysis, numerical integration constitutes a wide range of algorithms for computing the numerical value of a definite integral and, by extension, the term is sometimes used to describe numerical algorithms for solving differential equations. In probability and statistics, the t-distribution is a probability distribution that arises from the problem of estimating the mean of a normally distributed population when the sample size is small and the population standard deviation is unknown. During this practice, Simpson's rule will help us to give a better resolution to the numerical integration, as well as the T-distribution will be used to reduce the range of error that can be made in finding the "area under the curve". Applying the necessary formulas for its calculation, it is expected to achieve the proposed tests.

2 State of the art

2.1 Numerical Integration

There are several reasons for performing numerical integration. The main one may be the impossibility of performing the integration analytically. That is to say, integrals that would require a great knowledge and handling of advanced mathematics can be solved in a simpler way by means of numerical methods. There are even functions that can be integrated but whose primitive cannot be calculated, being the numerical integration of vital importance. The analytical solution of an integral would give us an exact solution, while the numerical solution would give us an approximate solution. The error of the approximation, which depends on the method used and how fine it is, can be so small that it is possible to obtain a result identical to the analytical solution in the first decimal places. Numerical integration is the process of determining the area "under" some function. Numerical integration calculates this area by dividing it into vertical "strips" and summing their individual areas. The key is to minimize the error in this approximation.

Integrating *a function*
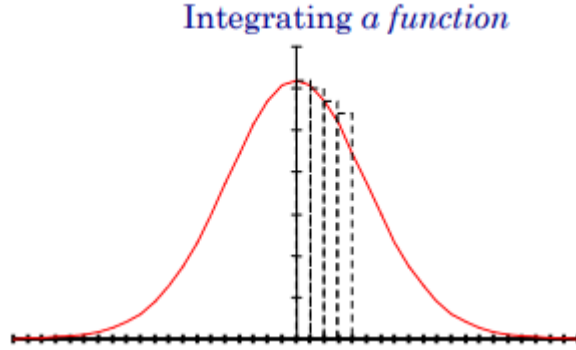
Simpson's rule is a method that uses parabolas to approximate the curve instead of drawing linear segments. Simpson's rule can be used to integrate a symmetrical statistical distribution function over a specified range (e.g., from 0 to some value x). The formula to apply it is:

$$p = \frac{W}{3}\left[F(0) + \sum_{i=1,3,5\ldots}^{num\_seg-1} 4F(iW) + \sum_{i=2,4,6\ldots}^{num\_seg-2} 2F(iW) + F(x)\right]$$

Where:

1. num_seg = initial number of segments, an even number

2. W = x/num_seg, the segment width

3. E = the acceptable error, e.g., 0.00001

2.2 T distribution

The t distribution is a very important statistical tool. It is used instead of the normal distribution when the true value of the population variance is not known and must be estimated from a sample. The shape of the t distribution is dependent on the number of points in your dataset. As n gets large, the t distribution approaches the normal distribution. For lower values, it has a lower central "hump" and fatter "tails." In other words, the t-distribution is a probability distribution that estimates the value of the mean of a small sample drawn from a population that follows a normal distribution and for which we do not know its standard deviation.
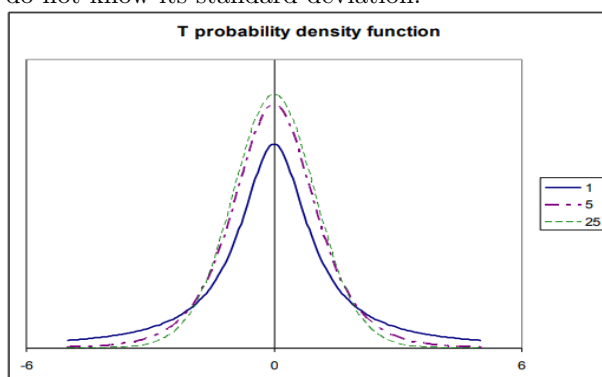
When numerically integrating the t distribution with Simpson's rule, use the following function:

$$F(x) = \frac{\Gamma\left(\dfrac{dof+1}{2}\right)}{(dof * \pi)^{1/2}\,\Gamma\left(\dfrac{dof}{2}\right)}\left(1+\frac{x^2}{dof}\right)^{-(dof+1)/2}$$

Where:

- dof = degrees of freedom

- Γ is the gamma function

The gamma function is
$\Gamma(x) = (x - 1)\,\Gamma(x - 1)$, where:

- $\Gamma(1) = 1$

- $\Gamma(1/2) = \sqrt{\pi}$

3 Methodology

3.1 General architecture

For the development of the practice, the formulas mentioned above were used and applied to the following tests in order to correctly arrive at the "area under the curve" as accurately as possible:

We apply the tests corresponding to the practice. For Simpson's rule we will apply the following ones:

- Should return p=16 if f(x)=2x, x0=0, x1=4, num_seg=4, dof=0

- Should return p=0.33333 if f(x)= x*x, x0=0, x1=1, num_seg=4, dof=0

- Should return p=1.386 if f(x)= 1/x, x0=1, x1=4, num_seg=6, dof=0

We apply the tests now for the T-distribution, which are as follows:

- Should return P=0.35006 if f(x)x0=0, x1=1.1, num_seg=10, dof=9

- Should return P=0.33333 if f(x)x0=0, x1=1.1812, num_seg=10, dof=10

- Should return p=1.386 if f(x)x0=0, x1=2.750, num_seg=10, dof=30

4 Results

As a final result we should obtain the data given in the following table:

Test		Expected Value	Actual Value
x	**dof**	**p**	
0 to x= 1.1	9	0.35006	
0 to x= 1.1812	10	0.36757	
0 to x= 2.750	30	0.49500	