MAE 384 Homework #1

1. (a) Find the *largest positive number* and the *smallest positive number* that are allowable on your calculator. Describe the process through which you find these two limits. (b) Provide an example of an operation that leads to underflow on your calculator. For example, on my calculator the operation "$10^{-99} \div 2$" leads to "0", instead of the correct answer of 0.5×10^{-99}. If possible, please provide the brand and model of your calculator. (Mine is CASIO *fx*-260.) **1 point**

2. (Modified from Prob 1.8) Write the number 256.1875 in *binary floating point representation*. Please provide the detail of your procedure, not just the final answer. **1 point**

> Hint: The answer to this question has the form, $1.\text{bbbbb...} \times 2^{\text{ppp...}}$, in which both the mantissa (bbbbbb...) and the exponent (ppp...) are expressed in binary form. For example, the number 50 is 1.1001×2^{101} in binary floating point representation.

3. (Modified from Prob 1.17) Taylor series expansion of the exponential function $f(x) = e^x$ is

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \ldots$$

Use this formula to evaluate e^{-2} for the following cases. (a) Use the first 4 terms. (b) Use the first 6 terms. (c) Use the first 8 terms. (Please keep whatever number of digits your calculator gives you and don't worry about the specific way of rounding as the original Prob. 1.17 demanded.) (d) Directly evaluate e^{-2} using the built-in exponential function on your calculator. Treat the outcome as the true value and use it to evaluate the truncation errors for (a)-(c). **1 point**

> Hint: The factorial function, N!, is defined as
>
> $$N! = N \times (N-1) \times (N-2) \ldots 3 \times 2 \times 1$$
>
> For example, $3! = 3 \times 2 \times 1 = 6$.

4. Solve the equation,

$$x^3 - 1 = 0 ,$$

using Bisection method. Choose (0.5, 2) as the initial interval and perform 5 iterations (i.e., do the "cutting" of the intervals 5 times). Compare the numerical result with the exact solution, $x = 1$; What are the true error and true relative error after 5 iterations? Please provide the detail of your procedure. 2 **points**