Start, as usual with your standard packages:

\[
\text{with(student);
with(plots);
readlib(unassign)};
\]

An important function, called the standard error function and denoted \( \text{erf}(x) \) is actually defined by an integral:

\[
\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt
\]

or, in Maple:

\[
\text{erfdef} := x \rightarrow (2/sqrt(Pi))*int(exp(-t^2), t=0..x); 
\]

Check the values \( \text{erf}(x) \) at zero and infinity

\[
\text{erfdef}(0);
\text{erfdef}(\text{infinity});
\]

and plot the graph

\[
\text{plot( erfdef(x), x=-3..3)};
\]

This was only for illustration. This function is important enough so that Maple builds it in as \( \text{erf}(x) \). For example

\[
\text{plot( erf(x), x=-3..3)}; 
\]

shows exactly the same graph as you did above. The main reason for
the importance of this function is the Central Limit Theorem of Statistics. This theorem states that if we take repeated samples (in an independent manner) \( X_1, X_2, \ldots X_n \) from a fixed probability distribution with mean \( \mu \) and variance \( \sigma^2 \), and if we compute the sample average \( \bar{X}_n \) where

\[
\bar{X}_n = (\sum_{i=1}^{n} X_i) / n
\]

and, if we let \( Z_n \) be the normalized mean

\[
Z_n = \frac{\bar{X}_n - \mu}{(\sigma / \sqrt{n})}
\]
then $Z_n$ tends to a Standard Normal (also called Gaussian) distribution $Z$ with mean 0 and variance 1. This is very helpful because

1. It doesn’t matter what probability distribution we started out with (as long as the sampling was independent),

2. Probabilities can be found by computing definite integrals.

The Standard Normal distribution has a probability density function of the form

$$f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$

and the probability that $Z$ is between any two numbers $a$ and $b$ is

$$P\{a < Z < b\} = \int_a^b \frac{1}{\sqrt{2\pi}}e^{-(x^2/2)}dx$$

Define the probability density function for the Standard Normal distribution in Maple

```maple
f := x -> exp(-(1/2)*x^2)/sqrt(2*Pi);
```

and compute some probabilities

```maple
int(f(x),x=-1..1);
evalf(%);
evalf(int(f(x),x=-2..2));
evalf(int(f(x),x=-3..3));
```

In general, one can show by substitution that

$$P\{a < Z < b\} = (1/2) \text{erf}(b/\sqrt{2}) - (1/2) \text{erf}(a/\sqrt{2}) \quad (***)$$

So that Maple will respond in terms of the standard error function when one types

```maple
int(f(x),x=a..b);
```

We want to show that one can approximate the cumulative distribution function $F(x)$ of the Standard Normal Distribution by picking random numbers and combining them in accordance with the Central Limit Theorem. We first need a way to generate a random fraction between 0.0 and 1.0. We can do this in Maple by

```maple
evalf(rand() / 99999999999999999);
```

(note there are 12 nines) since `rand()`, by itself, gives a 12-digit random
positive integer. Since the mean $\mu$ of the uniform distribution on $[0, 1]$ is $1/2$ and since the variance $\sigma^2$ is $(1/12)$ we need to automate the computation of

$$Z_n = \frac{X_n - \mu}{(\sigma/\sqrt{n})} = \frac{X_n - (1/2)}{\sqrt{12n}}$$

We do this by writing the following Maple procedure `rfrac()` (is is easiest to do this on a single line and let Maple wrap the display)

```maple
iterate := 100;
rfrac := proc() local i, rsum; rsum := 0;
for i from 1 to iterate do:
    rsum := rsum + evalf(rand()/999999999999):
    od;
    evalf(((rsum/iterate)
    - 0.5)*sqrt(12*iterate));
end;
```

We made `iterate` a variable because we can increase the accuracy of our approximation by increasing `iterate` (at the expense of longer computation times). Test that this procedure `rfrac` produces random real numbers $Z_{\text{iterate}}$ with a Standard Normal Distribution by typing

```maple
rfrac();
```

a dozen or more times. Note that virtually all of the values are between $-3$ and 3, which is to be expected.

Next we use the function `rfrac()` above to define a second procedure which estimates the cumulative distribution function of the Standard Normal distribution, that is, it approximates

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \quad (\ast)$$

We do this as follows (again typing all of this on one line)

```maple
cdf := proc(x) local i, rsum; rsum := 0;
for i from 1 to iterate do:
    if (rfrac() < x) then rsum := rsum + 1 fi:
    od; evalf(rsum/iterate);
end;
```

What is being done by `cdf` is to count up the fraction of times that the random variable `rfrac()` is less than $x$. This is an approximation to the integral in $(\ast)$ (why?) Since `cdf(x)` is an approximation to the cumulative distribution function $F(x)$ above, it must be the case that if we compare
cdf(x) and
\[
\lim_{a \to -\infty} \left( (1/2) \text{erf}(x/\sqrt{2}) - (1/2) \text{erf}(a/\sqrt{2}) \right) = (1/2) \text{erf}(x/\sqrt{2}) + (1/2)
\]
then the former must be an approximation to the latter. Plot the graphs (in separate windows) of both these functions in Maple. **Warning: note that the first graph will take 150-250 seconds of computation time** (with iterate equal to 100) Also note the the syntax in plot is different for a procedure (basically you don’t type any x’s).

\begin{verbatim}
plot(cdf,-3..3,numpoints=isqrt(iterate));
plot((1/2)*erf(x/sqrt(2)) + (1/2),x=-3..3);
\end{verbatim}

Print out both of these graphs. Highlight the latter one in red or another bold color and superimpose them to compare the effectiveness of the approximation.

**Assignment:**

1. Verify the assertion of equation (**) above by manually applying the substitution \( u = (x/\sqrt{2}) \) to the integral

\[
\int_{a}^{b} \frac{1}{\sqrt{2\pi}} e^{-(x^2/2)} \, dx
\]

2. Show that the variance of the uniform distribution on \([0,1]\) is \((1/12)\) (as claimed above) by evaluating the definite integral

\[
\int_{0}^{1} (x-\mu)^2 \, dx = \int_{0}^{1} (x-(1/2))^2 \, dx
\]

3. Explain why “counting up the fraction of times that the random variable \( \text{rfrac()} \) is less than \( x \) is an approximation to the integral”

\[
\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \, dt
\]

4. Increase the value of \( \text{iterate} \) to 225 (**this will take around 800-900 seconds of computation time**) and print both graphs
again, highlighting the latter one as before and superimposing the two graphs. This approximation should take noticeably longer but be visibly better than the approximation above. Turn in these two graphs along with your answers the problems 1, 2, and 3.