PS#14 - The normal distribution!

The well-known "68-95-99.7 rule" states that approximately 68% of values of a normally-distributed random variable lie within one standard deviation of the mean. In this extended problem, we will use Taylor series to calculate this percentage directly.

Here is the equation for the "normal probability density function" with mean $\mu = 0$ and standard deviation $\sigma = 1$:

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right).$$

You may know this function better by another name: the bell curve!

Throughout this problem set, when you are asked to write down a Taylor series for some function, you may either use Σ notation or write out the first few terms of the series. You may in particular wish to write *both* versions in each part, because frequently one is more useful than the other.

1. Draw a graph of $p(x)$. You can use Desmos, but in a minute I'm going to ask you to shade something in, so you may want something you can draw on.
2. We want to find the proportion of values that lie within one standard deviation ($\sigma = 1$) of the mean ($\mu = 0$); in other words, we want to find the proportion of values that lie less than 1 distance away from 0 in either direction. Shade in this region on your graph.
3. Areas are definite integrals; we've just shaded in $\int_a^b p(x)\,dx$.
   What should our bounds of integration a and b be? Why?
4. If we find an antiderivative of $p(x)$, which we might call $P(x)$, then our answer will be $P(b) - P(a)$. But unfortunately, we can't immediately write down an antiderivative $P(x)$. Why not?
5. Write down the Taylor series for $f(u) = e^u$.
6. Use the Taylor series for e^u to write down $T_p(x)$, the Taylor series for $p(x)$.
   (Don't forget the $\frac{1}{\sqrt{2\pi}}$!)
7. Use the Taylor series for $p(x)$ to write down the Taylor series for its antiderivative $P(x)$, which would therefore be called $T_P(x)$.
8. Safety check: if a and b were outside the interval of convergence for $T_P(x)$, that would be bad. Therefore, determine the interval of convergence for $T_P(x)$. (You can use the ratio test, or you can explain how you know what the interval of convergence is in some other way.)
9. Use the first four nonzero terms of $T_P(x)$ to approximate $P(b)$.
   Report at least 7 decimal places in your answer.
10. Use the first four nonzero terms of $T_P(x)$ to approximate $P(a)$.
    Again, report at least 7 decimal places.
11. Use your values from the two previous parts to approximate $P(b) - P(a)$.
    (Your answer should be close to 68%.)

    (Problem continues on other side of this page!)

12. Now let's think about accuracy. How many terms in $T_P(x)$ would be necessary to ensure that your approximation to $P(b) - P(a)$ is accurate to four decimal places? **Notes:**

    - By "accurate to four decimal places," we mean that if we rounded both our approximation and the actual value to the fourth decimal place, they'd round to the same thing. How far off can we be?
    - You can't compare your computed value directly to the actual value, because the actual value is in principle unknown!
    - This is an alternating series, so we can use the alternating series error bound. (The Lagrange error bound for this particular series is, uh, difficult.)
    - We're actually approximating two different values, $P(b)$ and $P(a)$, so to figure out the total error, we'll need to think separately about the error in each approximation.

13. Conclude by telling me why Taylor series are really awesome for dealing with yucky integrals.

14. **(Bonus!)** Similarly, check the "95" and "99.7" parts of the rule, which talk about the proportion of values within 2 and 3 standard deviations of the mean. (What will you have to change in your previous work?)