Let's learn the T-distribution! Note: Learn about the normal distribution first!

For previous distributions the sample size was assumed large (N>30). For sample sizes that are less than 30, otherwise (N<30). Note: Sometimes the t-distribution is known as the student distribution.

The t-distribution allows for use of small samples, but does so by sacrificing certainty with a margin-of-error trade-off. The t-distribution takes into account the sample size using n-1 degrees of freedom, which means there is a different t-distribution for every different sample size. If we see the t-distribution against a normal distribution, you'll notice the tail ends increase as the peak get 'squished' down.

It's important to note, that as n gets larger, the t-distribution converges into a normal distribution.

To further explain degrees of freedom and how it relates tothe t-distribution, you can think of degrees of freedom as an adjustment to the sample size, such as (n-1). This is connected to the idea that we are estimating something of a larger population, in practice it gives a slightly larger margin of error in the estimate.

Let's define a new variable called t, where :

$$t = \frac{\overline{X} - \mu}{s}\sqrt{N-1} = \frac{\overline{X} - \mu}{s/\sqrt{N}}$$

which is analogous to the z statistic given by

$$z = \frac{\overline{X} - \mu}{\sigma/\sqrt{N}}$$

The sampling distribution for t can be obtained:

$$f(t) = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{v\pi}\,\Gamma(\frac{v}{2})}(1 + \frac{t^2}{v})^{-\frac{v+1}{2}}$$

Where the gamma function is:

$$\Gamma(n) = (n-1)!$$

And v is the number of degrees of freedom, typically equal to N-1.

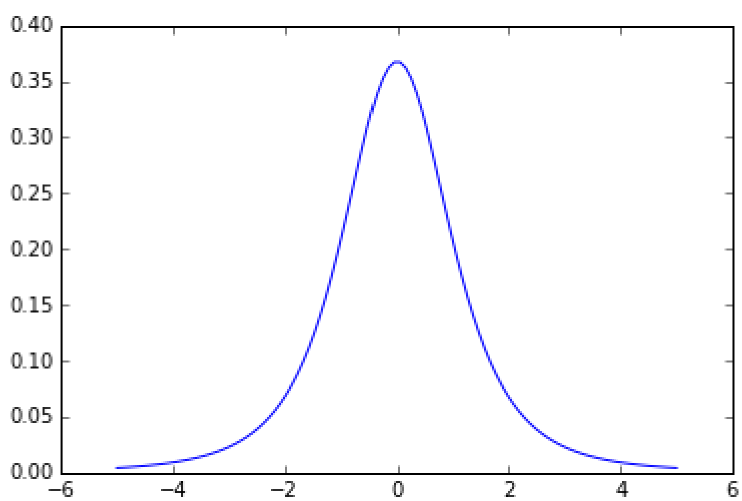
Similar to a z-score table used with a normal distribution, a t-distribution uses a t-table. Knowing the degrees of freedom and the desired cumulative probability (e.g. P(T >= t)) we can find the value of t. Here is an example of a lookup table for a t-distribution:

http://www.sjsu.edu/faculty/gerstman/StatPrimer/t-table.pdf (http://www.sjsu.edu/faculty/gerstman/StatPrimer/t-table.pdf)

Now let's see how to get the t-distribution in Python using scipy!

In [5]:
```
#Import for plots
import matplotlib.pyplot as plt
%matplotlib inline

#Import the stats library
from scipy.stats import t

#import numpy
import numpy as np

# Create x range
x = np.linspace(-5,5,100)

# Create the t distribution with scipy
rv = t(3)

# Plot the PDF versus the x range
plt.plot(x, rv.pdf(x))
```

Out[5]:
```
[<matplotlib.lines.Line2D at 0x15d9e5c0>]
```



Additional resources can be found here:

1.) http://en.wikipedia.org/wiki/Student%27s_t-distribution (http://en.wikipedia.org/wiki/Student%27s_t-distribution)

2.) http://mathworld.wolfram.com/Studentst-Distribution.html (http://mathworld.wolfram.com/Studentst-Distribution.html)

3.) http://stattrek.com/probability-distributions/t-distribution.aspx (http://stattrek.com/probability-distributions/t-distribution.aspx)

In [ ]:

```

```