1. Redo the problem 4 in HW1 in python (use only the distribution objects, like norm, binom), make sure you print conclusions from two methods(using clssic statistics method and p-value)

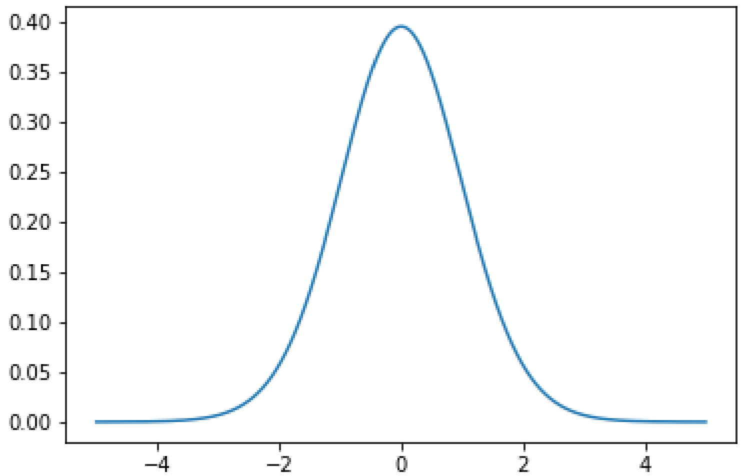
In [25]:

```python
import numpy as np
import scipy as sp
#from scipy.stats import ttest_ind
#from scipy.stats import norm
#from scipy.stats import binom
from scipy import stats

N = 100.0
second = 28.0
first = 0.37
alpha = 0.01

z = ((second/N) - first) * np.sqrt(N / (first*(1-first)))
z = round(z, 2)
#pcomp = sp.stats.t.sf(np.abs(z), (first*(1-first)-1)*2
print 'z = ', z
#Studnt, Single tail
#Should be -2.33
cv = -stats.t.ppf(1-alpha, 999)
cv = round(cv, 2)
if z < cv:
    print z, ' < ', cv
    print 'Reject the null hypothesis'
else: #should print this
    print z, ' > ', cv
    print 'Fail to reject the null hypothesis'
    print 'There is not enough evidence to conclude that the cars have improve
d'

print

p = sp.stats.norm.cdf(z)
p = round(p, 3)
print 'p-value = ', p
if p < alpha:
    print p, ' < ', alpha
    print 'We reject the null hypothesis'
else: #should print this
    print p, ' > ', alpha
    print 'We fail to reject the null hypothesis'
    print 'There is not enough evidence to conclude that the cars have improve
d'
```

```
z =  -1.86
-1.86  >  -2.33
Fail to reject the null hypothesis
There is not enough evidence to conclude that the cars have improved

p-value =  0.031
0.031  >  0.01
We fail to reject the null hypothesis
There is not enough evidence to conclude that the cars have improved
```

2. Plot the pdf of a t distribution with degree of freedom 30

In [26]:
```
#sample size = 31
# plot the t-distribution pdf
from numpy import arange
from matplotlib import pyplot
from scipy.stats import t
# define the distribution parameters
sample_space = arange(-5, 5, 0.001)
#dof = len(sample_space) - 1
dof = 30
# calculate the pdf
pdf = t.pdf(sample_space, dof)
# plot
pyplot.plot(sample_space, pdf)
pyplot.show()
```



3.Anomaly detection Assume we have a datasets, it represents the daily number of clicks of a small website for 3 months. We assume it follows a poisson distribution. We also assume the the data that exceeds the 95% confidence interval of the sample mean is not likely to happen. We consider them as anomalies. Write a function to extract those anomalies:

In [89]:
```
def anomaly_detection_poisson(values):
    anomalies = []
    margin = 1.96 #margin of error associated with 95% confidence interval
    counter = 0
    num = len(values)
    mean = np.mean(values)
    std = np.std(values)
    se = std / (np.sqrt(num))
    upper = mean + (margin * se)
    lower = mean - (margin * se)
    for i in values:
        if i/1. > upper or i/1. < lower:
            anomalies.append(i)
            counter += 1
    #print counter
    return anomalies

#748, 863
```

The following is a test function to test your detection function. If your function is correct, you should get the following result: `[731, 745, 868, 713, 863, 877, 867, 875, 867, 748, 900, 747, 717]`

In [90]:
```
value=[796, 756, 827, 779, 787, 731, 847, 829, 794, 790, 831, 820, 822, 792, 7
67, 757, 777, 775, 799, 756, 849, 808, 800, 823, 798, 809, 783, 800, 848, 772,
853, 759, 745, 868, 713, 863, 819, 829, 799, 777, 785, 812, 877, 839, 783, 867
, 875, 829, 818, 813, 763, 829, 812, 867, 790, 748, 822, 774, 900, 830, 781, 7
95, 825, 799, 795, 839, 840, 765, 781, 781, 797, 821, 852, 836, 811, 771, 800,
752, 776, 755, 798, 839, 821, 794, 846, 834, 825, 825, 830, 814, 839, 760, 786
, 747, 803, 717, 801, 819, 789, 824, 835, 806, 858, 774, 848, 793, 826, 796, 7
98, 773, 779, 775, 779, 806, 768, 787, 788, 822, 843, 780]
print anomaly_detection_poisson(value)
```

```
[796, 756, 827, 779, 787, 731, 847, 829, 794, 790, 831, 820, 822, 792, 767, 7
57, 777, 775, 756, 849, 823, 783, 848, 772, 853, 759, 745, 868, 713, 863, 81
9, 829, 777, 785, 812, 877, 839, 783, 867, 875, 829, 818, 813, 763, 829, 812,
867, 790, 748, 822, 774, 900, 830, 781, 795, 825, 795, 839, 840, 765, 781, 78
1, 797, 821, 852, 836, 811, 771, 752, 776, 755, 839, 821, 794, 846, 834, 825,
825, 830, 814, 839, 760, 786, 747, 717, 819, 789, 824, 835, 858, 774, 848, 79
3, 826, 796, 773, 779, 775, 779, 768, 787, 788, 822, 843, 780]
```