Homework 1; Psych 186B, Winter 2018

Marshall Briggs

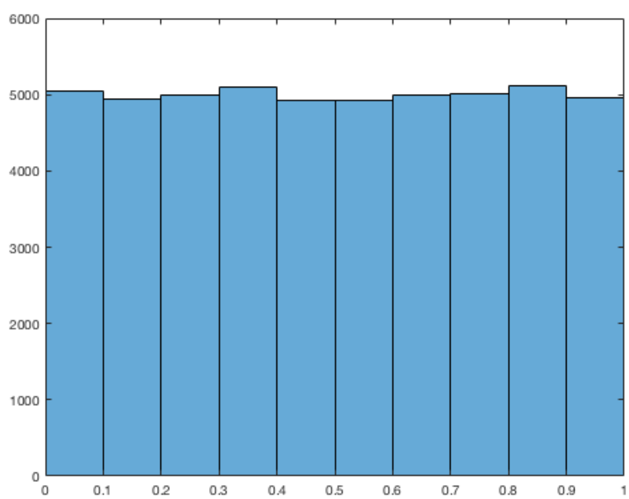
20 January 2017

1 Problem 1 (3 points) - problem1.m

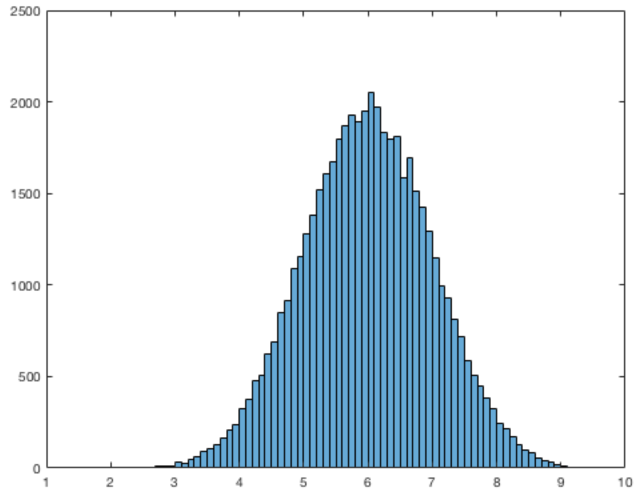
(a) I used the built-in Matlab pseudo-random number generation function, rand().

(b) Generate at least 50,000 random numbers

(c) Plot the histogram



2 Problem 2 (5 points) - problem2.m



These values do appear to be similar to the normal distribution, as the central limit theorem would suggest. In comparison to a figure of the true normal distribution, the mean is as expected (mean = 12 * 0.5 = 6.002), as is the standard deviation (0.997). Overall, there is less uniformity across the curve (mine has a couple of frequency spikes where the traditional normal distribution does not).

3 Problem 3 (15 points) - problem3.m

(a) (1 point) Geometrically, the dot product of two vectors is equal to:

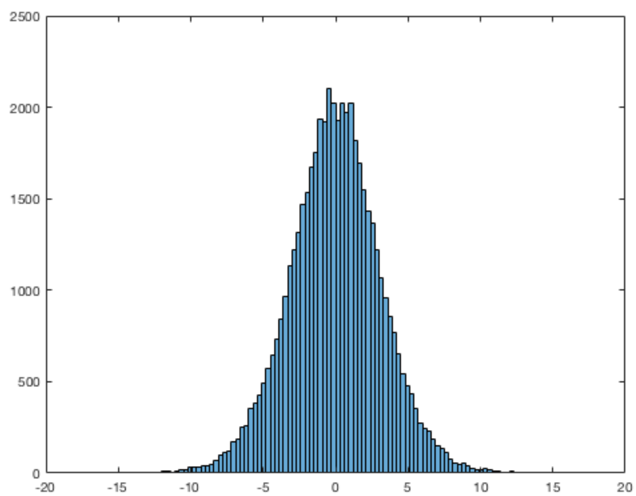
$$a * b = ||a||||b||cos(\theta)$$

In this case, because these vectors are normalized (length = 1), the dot product of two products is

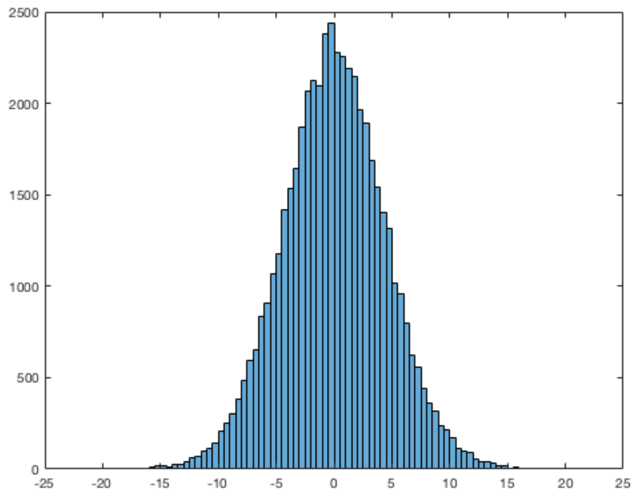
$$a * b = cos(\theta)$$

(b) (4 points) Generate a histogram of dot products and compute mean and standard deviation of the dot product. Use several dimensionalities: 10, 20, 50, 100 250, 500, 1000 and 2000.)

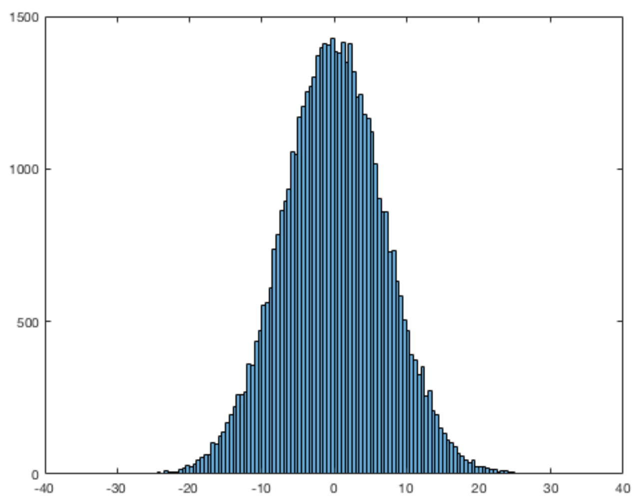
(i.) Dimensionality: 10; mean: -0.030312, stdev: 3.185221



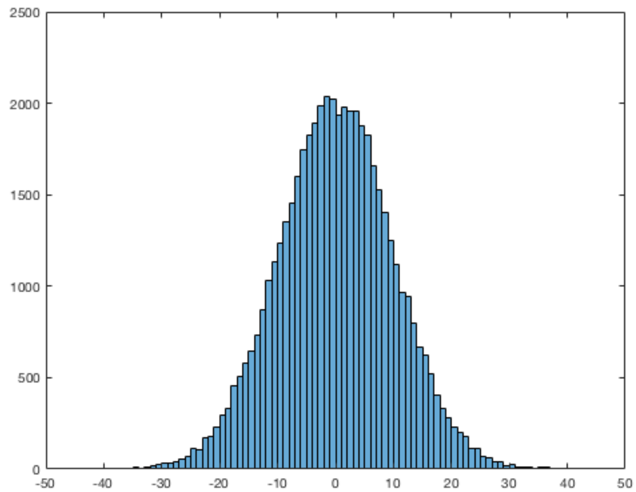
(ii.) Dimensionality: 20; mean: 0.003293, stdev: 4.443287



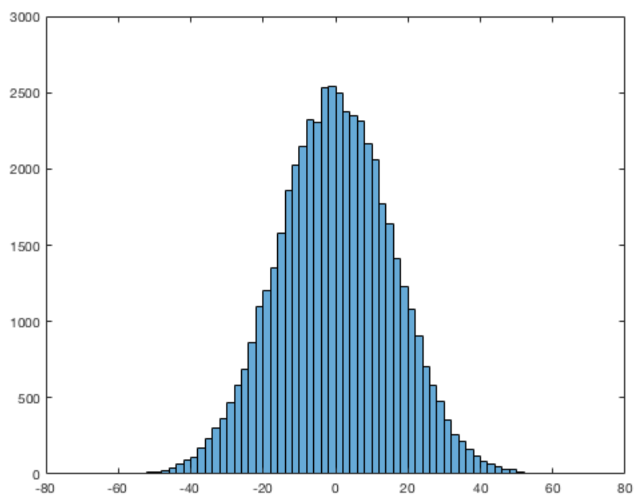
(iii.) Dimensionality: 50; mean: 0.013764, stdev: 7.078484



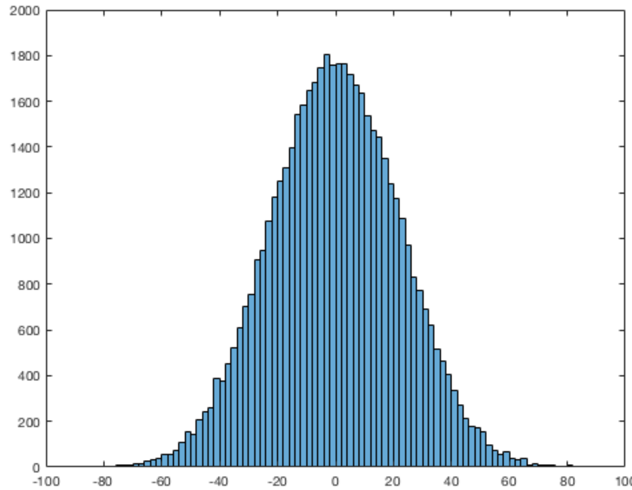
(iv.) Dimensionality: 100; mean: -0.016419, stdev: 9.979615



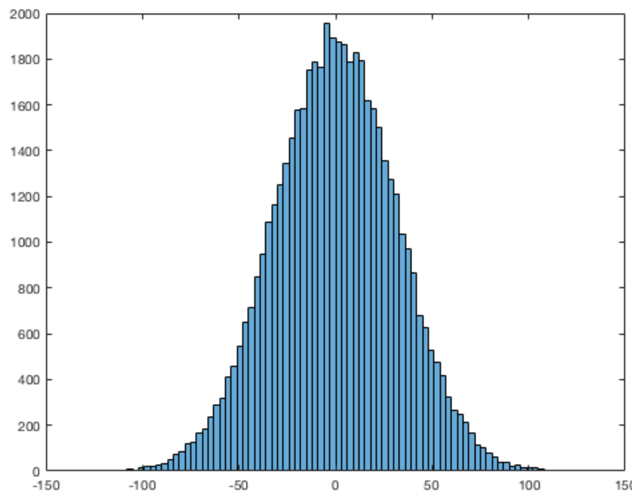
(v.) Dimensionality: 250; mean: -0.008430, stdev: 15.793571



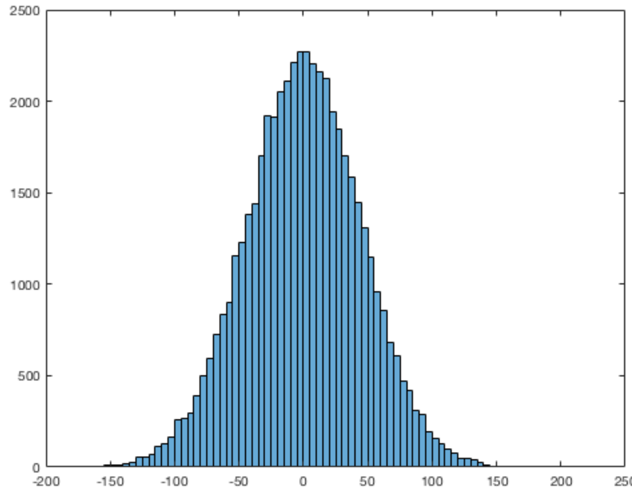
(vi.) Dimensionality: 500; mean: 0.031750, stdev: 22.400060



(vii.) Dimensionality: 1000; mean: 0.147033, stdev: 31.546584



(viii.) Dimensionality: 2000; mean: 0.538257, stdev: 44.786477



(c) Since the dot product of two normalized vectors is the cosine of the theta between them, the range of values the dot product can take is between [-1, 1], meaning the mean dot product should be 0.

(d) – Skip –

(e) After adding two more dimensionalities to the mix (3000, 5000) and attempting to eyeball a regression comparing the standard deviations among the various dimensionalities, they appear to increase as the dimensionality increases. There appears to be a negative quadratic relationship between dimensionality and standard deviation of the dot products.

$$-2.2613610^{-}6x^2 + 0.0238899x + 6.90872$$

One might assume that, as the dimensionality increased, standard deviation would decrease; however, this assumption would be a mistake. Standard error (the distribution of the sample means) decreases on an increasing sample size; this result is intuitive because we would assume regression to the mean would correct for any extreme sample means we would see. Standard deviation, however, will measure the variability of the sample overall, and that variability will only increase as the sample size / dimensionality increases (as further multiplications and combinations of extreme values will inevitability exacerbate those extreme values, producing values further from the mean than at lower dimensionalities).

4 Problem 4 (3 points) - problem4.m

The results of this simulation hovered around 3.14, as expected. I tested my pi simulation across various sample sizes:

n = 1000; pi = 3.16
n = 10000; pi = 3.13
n = 50000; pi = 3.139
n = 100000; pi = 3.142