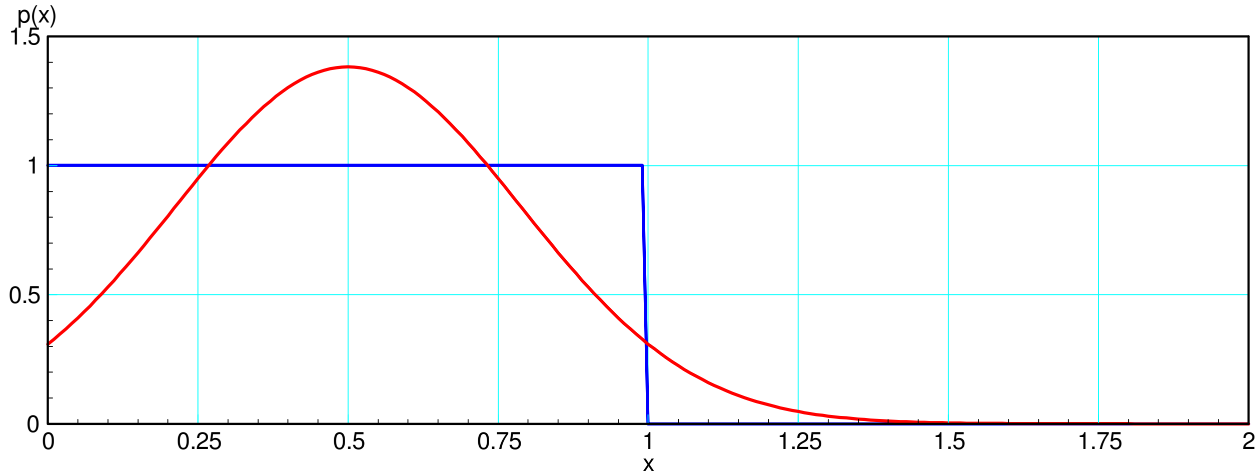
Central Limit Theorem

The Central Limit Theorem is probably one of the most important theorems in statistics. It basically says that all distributions coverage to a normal distribution.

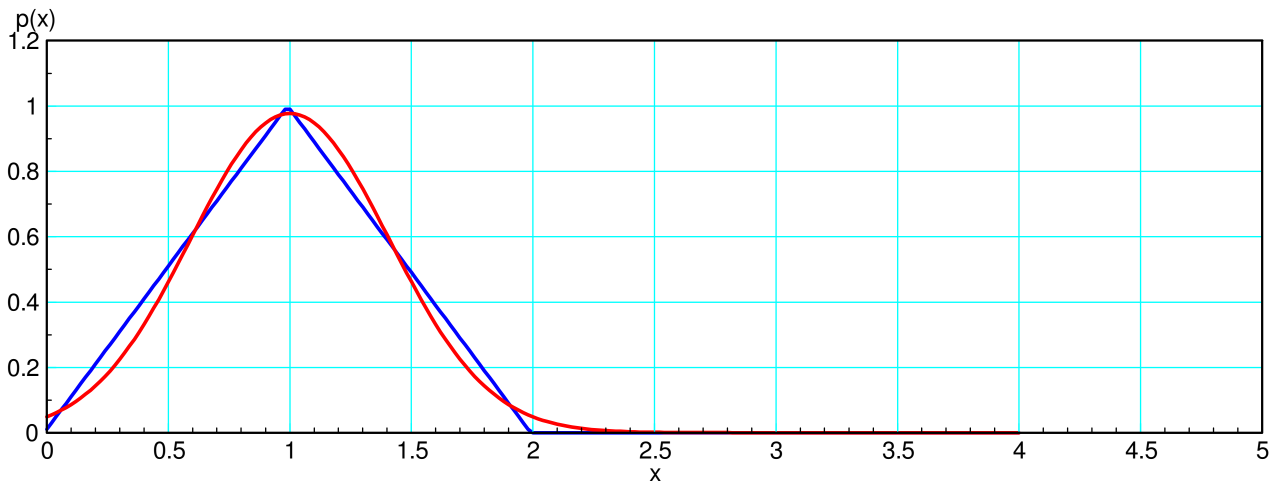
This is one of the reasons engineers tend to assume everything is described by a normal distribution (even when a Poisson distribution is more accurate). It also allows you to determine the probability for some fairly complex problems fairly accurately.

Example: Uniform Distribution.

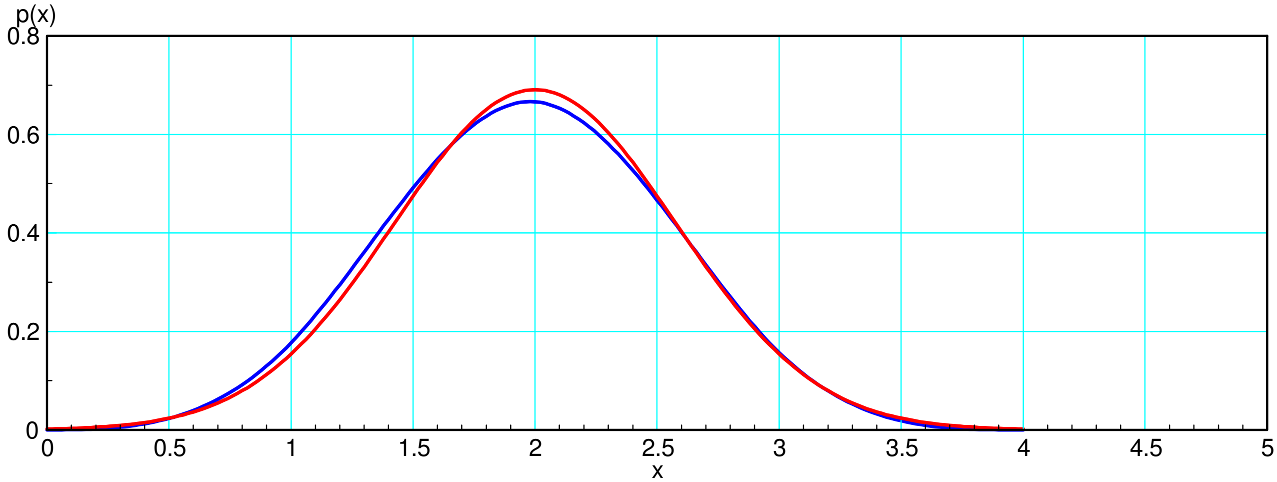
Let A be a uniform distribution over the range of (0, 1). If you plot this pdf along with a Normal distribution with the same mean and standard deviation, it is clear that the two are different.



Uniform Distribution (blue) and it's normal approximation (red)

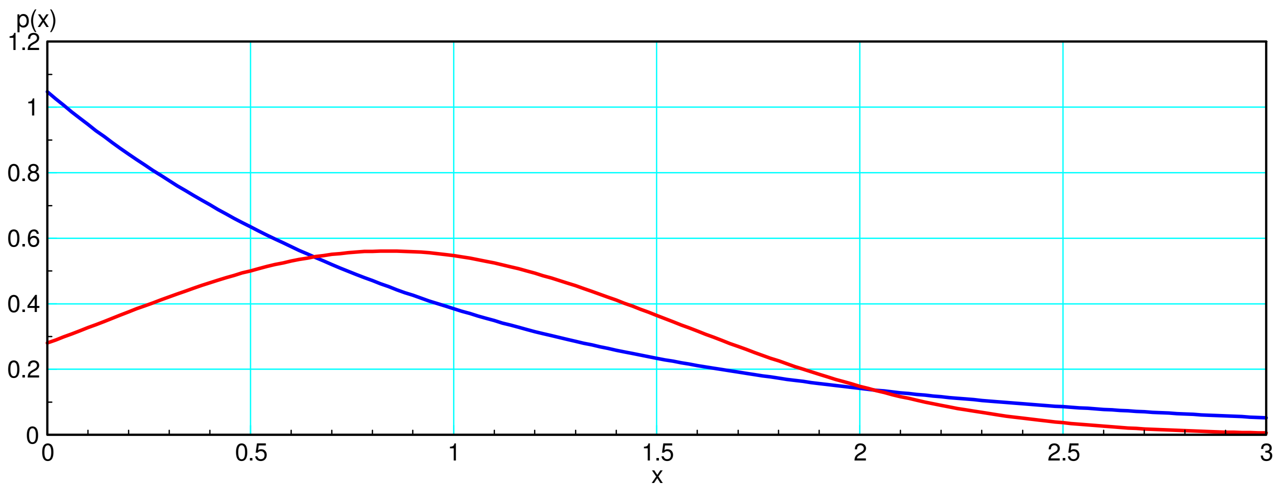


pdf for the sum of two uniform distributions (blue) and a normal distribution with the same mean and variance (red)

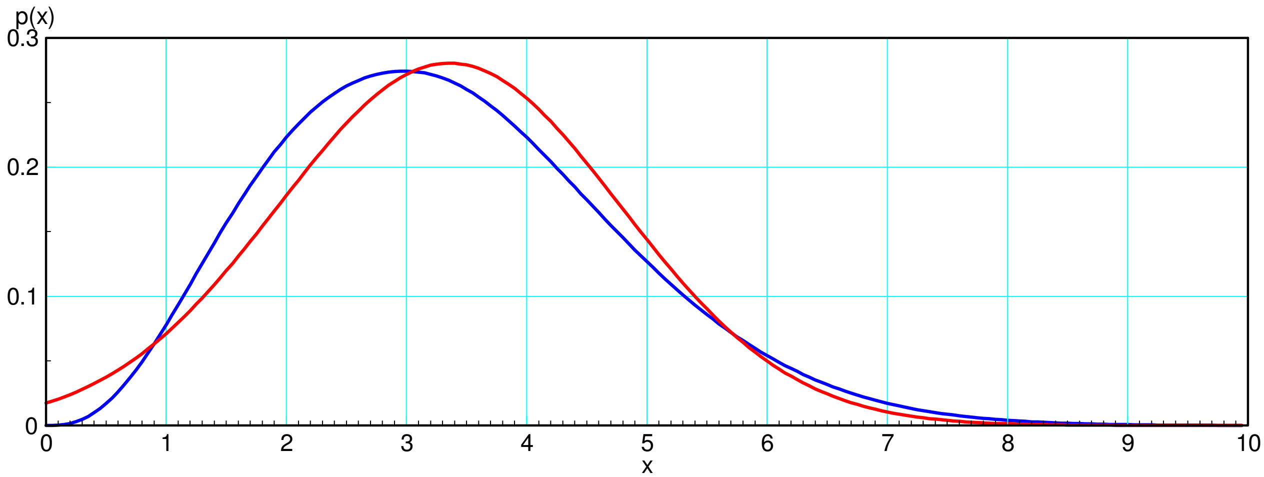


pdf for the sum of four uniform distributions (blue) and a normal distribution with the same mean and variance (red)

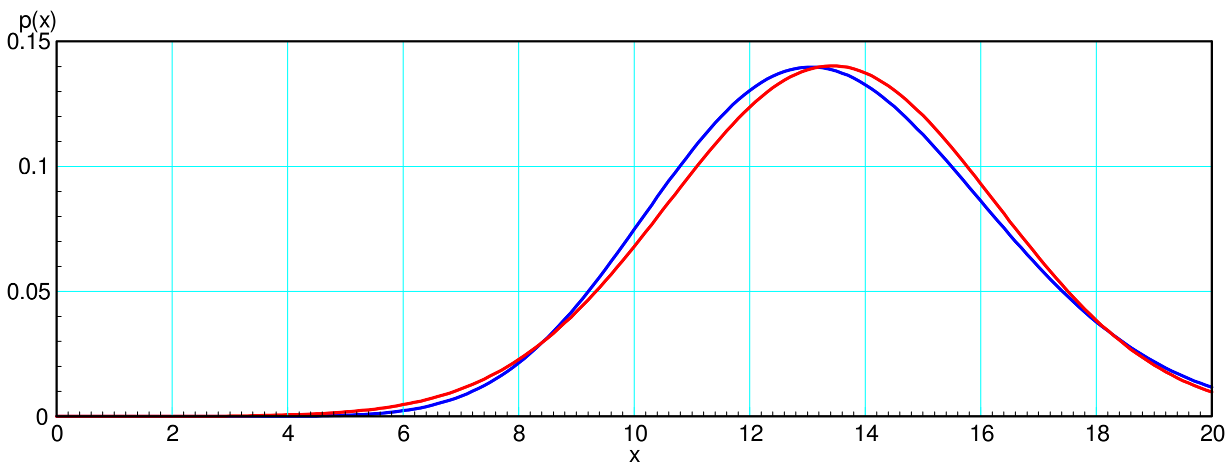
The same holds for other distributions. If you have an exponential distribution with a mean of 1. The pdf vs. a normal distribution with the same mean and variance looks like the following:



pdf of an exponential distribution (blue) and a normal distribution with the same mean and variance



The pdf of the sum of 4 exponential distributions (blue) along with a normal distribution with the same mean and variance (red)



The pdf of the sum of 16 exponential distributions (blue) along with a normal distribution with the same mean and variance (red)

That in essence is the cental limit theorem: the sum of any distribution converges to a normal distribution.

Example 1: Central Limit Theorem with Dice

Let X be the sum of rolling 5 six-sided dice (5d6) and four 10-sided dice (4d10). What is the probability of rolling higher than 54.5?

Exact Solution: Convolve the pdf for 5d6 and 4d10

In Matlab:

```
d6 = [0,ones(1,6)];
d10 = [0, ones(1,10)];
d6x2 = conv(d6, d6);
d6x4 = conv(d6x2,d6x2);
d6x5 = conv(d6x4, d6);
d10x2 = conv(d10, d10);
d10x4 = conv(d10x2, d10x2);
pdf = conv(d6x5, d10x4);
pdf = pdf / sum(pdf);

size(pdf)

ans =     1    71

sum(pdf(46:71))

ans =  0.2382
```

The chance of rolling 45 or higher is 23.82%

```
sum(pdf(51:71))

ans =   0.0748
```

The chance of rolling 50 or higher is 7.48%

Monte-Carlo Simulation

```
N = 0;
for i=1:1e6

d10 = sum( ceil(10*rand(1,4)) );
d6 = sum( ceil(6*rand(1,5)) );

X = d6 + d10;

if(X >= 45)
    N = N + 1;
end

end

N/1e6
```

45 or higher:

```
ans =    0.2384
```

50 or higher

```
ans =    0.07472
```

Method 2: Central Limit Theorem.

The mean and standard deviation of a 6 and 10 sided die are

d6:	mean = 3.5	d10:	mean = 5.5
	var = 2.9167		var = 8.250

So, 5d6 + 4d10 is

mean = 5 * 3.5 + 4 * 5.5

mean = 39.50

var = 5 * 2.9167 + 4 * 8.250

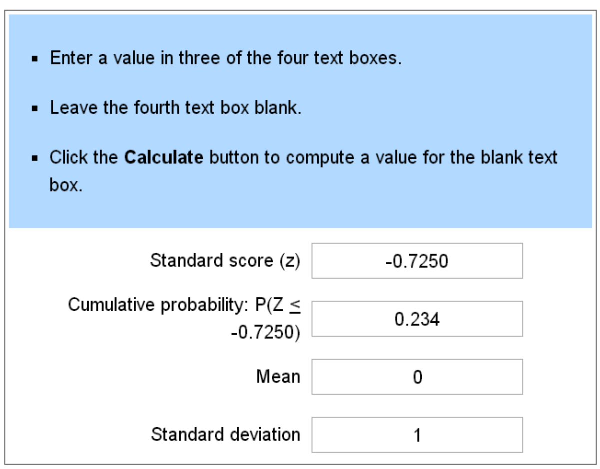
var = 47.584

st dev = 6.8981

The z-score for 44.5 is (roll 45 or higher)

$$z = \left(\frac{44.5-39.5}{6.8981}\right) = 0.7250$$

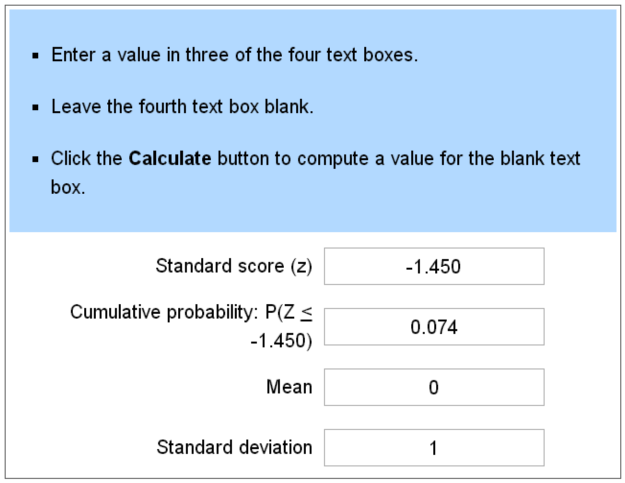
From StatTrek, a z-score of 0.7250 corresponds to a tail with an area of 0.234 (vs. 0.2384 and 0.2382)



The z-score for 49.5 is (roll 50 or higher) is 1.450

$$z = \left(\frac{49.5-39.5}{6.8981}\right) = 1.450$$

The corresponds to a probability of 7.4% (vs. 7.472% and 7.48%)



Example 2: Uniform Distribution.

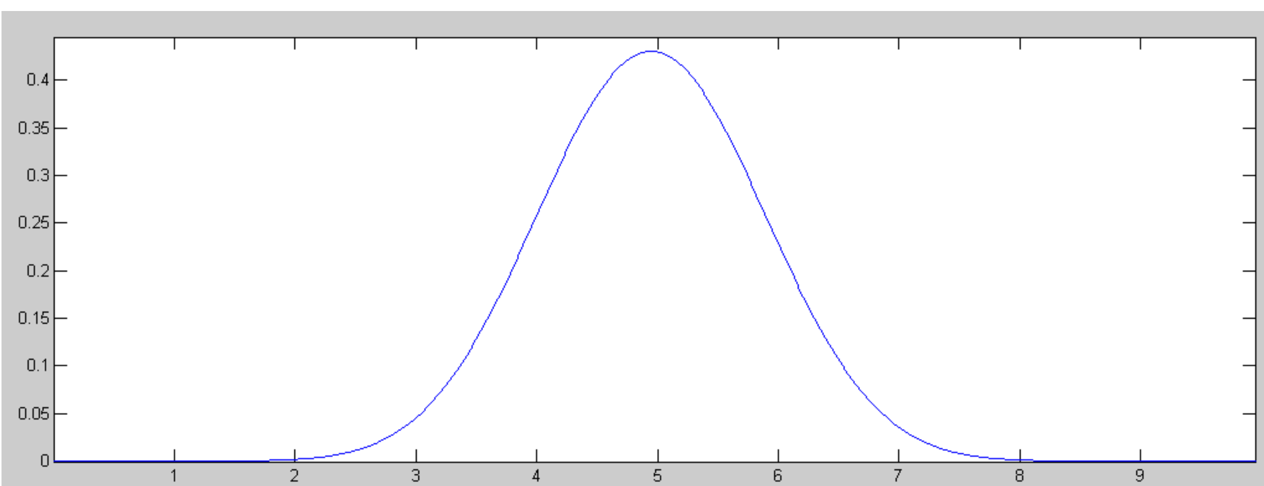
Let A1 .. A10 be uniform distributions over the interval (0, 1).

Let X be the sum of A1 .. A10.

Determine the probability that the sum is more than 9.000

Solution: Convolution with matlab.

```
x = [0:dx:2]';
A = 1*(x < 1);
A2 = conv(A, A) * dx;
A4 = conv(A2, A2) * dx;
A8 = conv(A4, A4) * dx;
A10 = conv(A2, A8) * dx;
sum(A10) * dx

ans =    1.0000

x = [0:2000] * dx;
plot(x,A10)

sum(A10(700:2000)) * dx

ans =    0.0121

sum(A10(600:2000)) * dx

ans =    0.1306

>>
```



Solution: Monte-Carlo Simulation

```
N6 = 0;
N7 = 0;

for i=1:1e5

  X = sum(rand(1,10));
  if(X > 6)
      N6 = N6 + 1;
  end
  if(X > 7)
      N7 = N7 + 1;
  end

end

[N6, N7] / 1e5


    0.1388    0.0137     monte-carlo
    0.1306    0.0121     convolution
```

Solution: Normal Approximation

A uniform distribution over the interval of (0,1) has

mean = 0.5

var = 1/12

The sum of 10 of these has

mean = 5.0

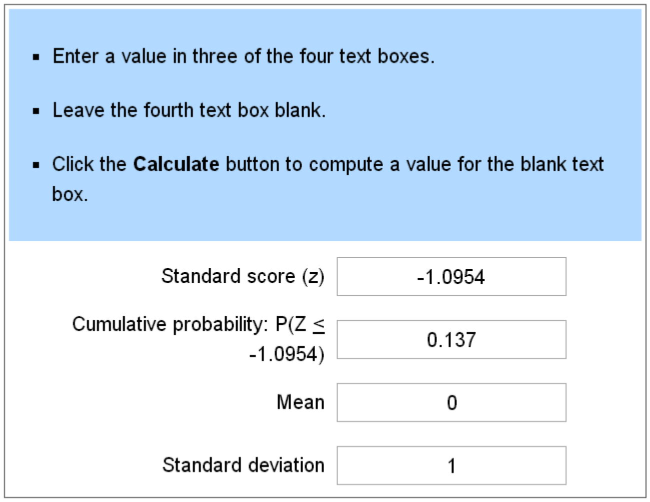
var = 10/12

s = 0.9129

The z-score for 6.00 is

$$z = \left(\frac{6-5}{0.9129}\right) = 1.0954$$

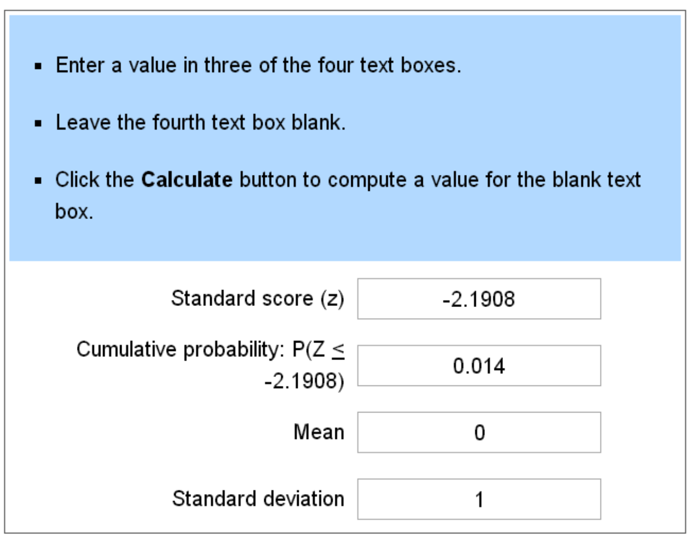
This corresponds to a probability of 0.137



The z-score for rolling 7.00 or higher is

$$z = \left(\frac{7-5}{0.9129}\right) = 2.1908$$

which corresponds to a probability of 0.014



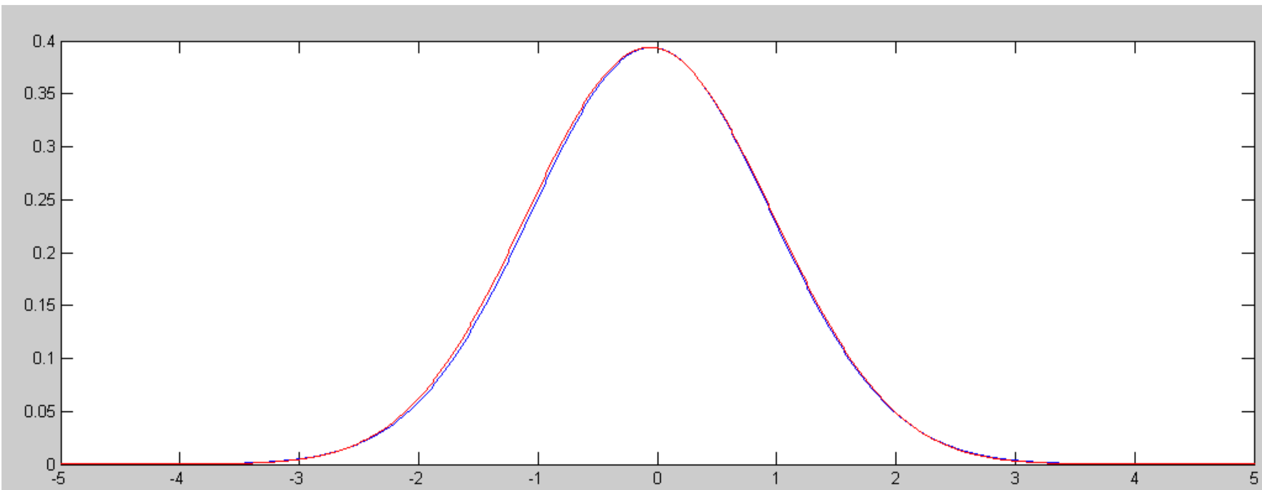
Example 3: Uniform approximation for a Normal Distribution

One trick to creating a normal distribution is to sum twelve uniform distributions and subtracting 6. This gives

mean = 0.0

var = 12 / 12 = 1

The resulting distribution of 12 uniform distributions (found using convolution) vs. a standard normal curve are almost identical



pdf for a Standard Normal Curve (red) and summing 12 uniform distributions and subtracting six