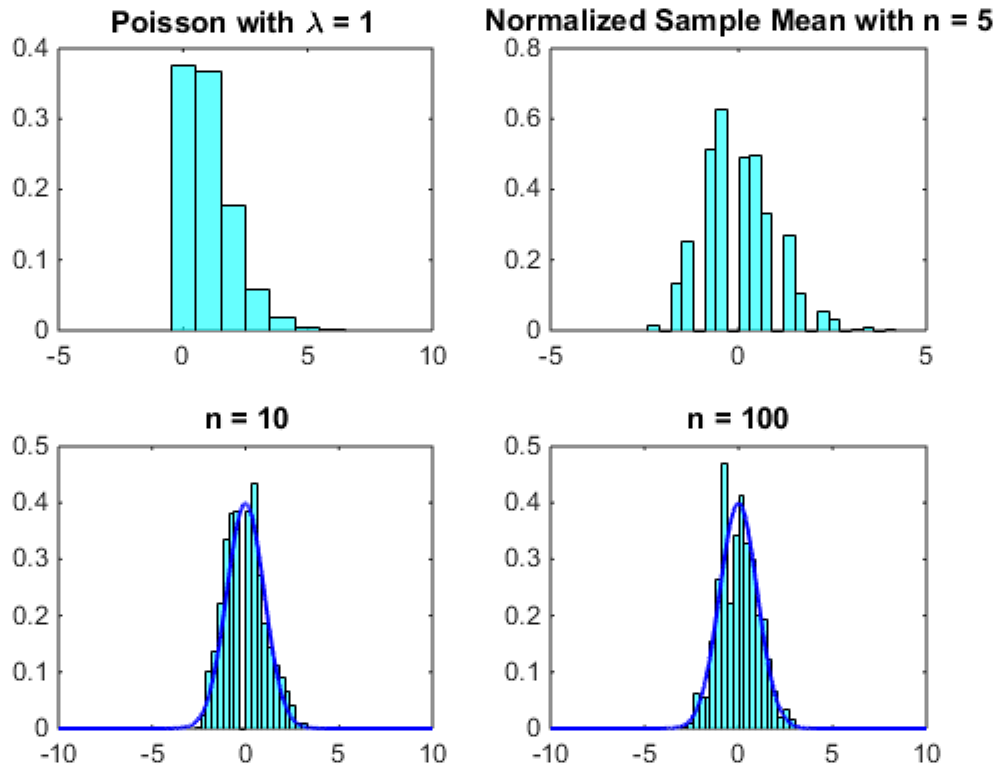
Table of Contents

An Introduction

The central limit theorem is an extremely powerful result in probability theory. It states that the sum of independent random variables from identical distributions tends towards a normal distribution as the number of summands increases. In this script, several different forms of this theorem will be analyzed for their validity of approximation, however, it is first important to get a sense for the power of this theorem. A function SampleMean has been written to create a sample mean of random samples from a specific distribution, and a second function nzeroone has been written to scale these sample means to have means of zero and variances of one. The first visual example will be the discrete Poisson distribution with a lambda of 1. A random sample from it will be compared to the standard normal distribution after application of SampleMean and nzeroone with n = 1, 5, 10 and 100. Note both the power of this approximation, and the swiftness of diminishing returns; n = 100 does not seem to obviously be a much better approximation than n = 10.

```
y = linspace(-10,10,10000);
pn = pdf('Normal',y,0,1);
poiss = SampleMean('Poisson',1,1000);
sample5 = SampleMean('Poisson',5,1000);
sample10 = SampleMean('Poisson',10,1000);
sample100 = SampleMean('Poisson',100,1000);
figure;
subplot(2,2,1);
histogram(poiss,'Normalization','pdf','FaceColor','cyan');
title('Poisson with \lambda = 1');
subplot(2,2,2);
histogram(nzeroone(sample5,5,1,1),'Normalization','pdf','FaceColor','cyan');
title('Normalized Sample Mean with n = 5');
subplot(2,2,3);
histogram(nzeroone(sample10,10,1,1),'Normalization','pdf','FaceColor','cyan');
hold on;
plot(y,pn,'blue','LineWidth',1.5);
title('n = 10');
subplot(2,2,4);
histogram(nzeroone(sample100,100,1,1),'Normalization','pdf','FaceColor','cyan');
hold on;
plot(y,pn,'blue','LineWidth',1.5);
title('n = 100');
```

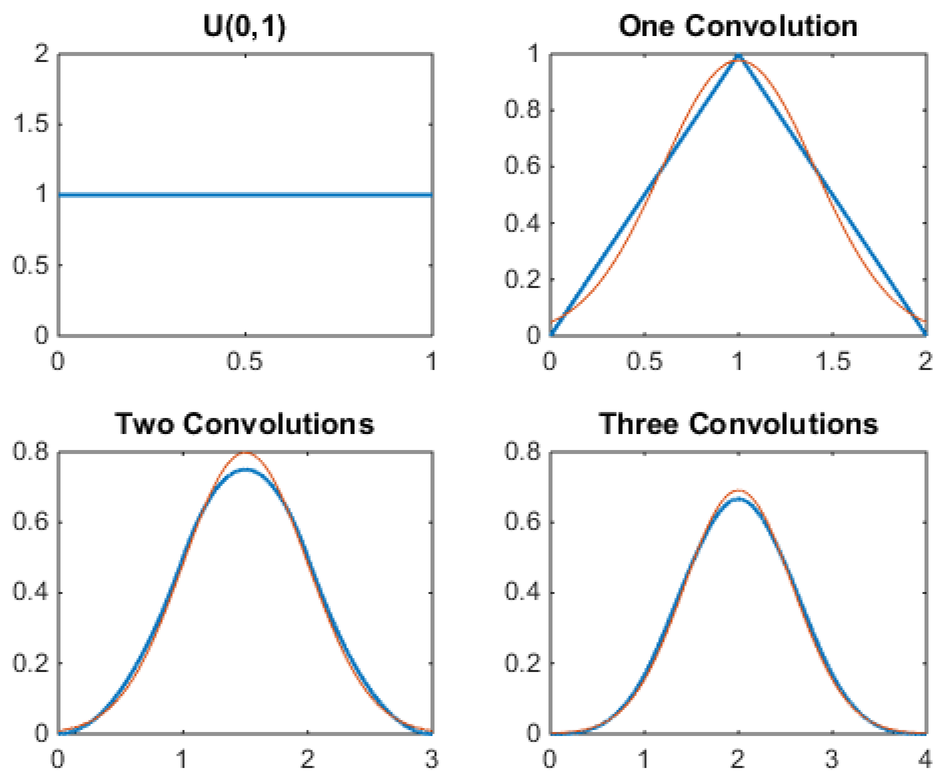


The Convolution Method

Although the above example may help for understanding the central limit theorem in a qualitative sense, it would be more helpful to work deterministically with functions rather than actual random events when we are looking to analyze the error associated with the theorem. For this, we will present a result regarding the convolution of probability density functions. In this section, we will state the result and observe its consequences, however, we will not prove it. The result is as follows: suppose f(x) is a probability density function, then the probability density function of the sample mean of two independent random variables corresponding to this distribution is given by g = f * f. In this example, we take a very simple distribution, U(0,1), and we take three concurrent convolutions. Scaling of mean and variance aside, it begins to look more like the normal distribution. Note that in convolving two distributions we must scale by the sampling value in x, as MATLAB performs discrete convolution, not integral convolution. On the plots, blue lines represent the convolved PDFs, while orange lines represent the estimated normal PDF.

```
unif = pdf('Uniform',linspace(0,1,1000),0,1);
figure;
subplot(2,2,1);
plot(linspace(0,1,1000),unif,'LineWidth',1.5);
title('U(0,1)');
sum2 = (1/1000)*conv(unif,unif);
subplot(2,2,2);
plot(linspace(0,2,1999),sum2,'LineWidth',1.5);
hold on;
plot(linspace(0,2,1999),.5*normpdf(.5*linspace(0,2,1999),.5,sqrt(1/24)));
title('One Convolution');
sum3 = (1/1000)*conv(sum2,unif);
```

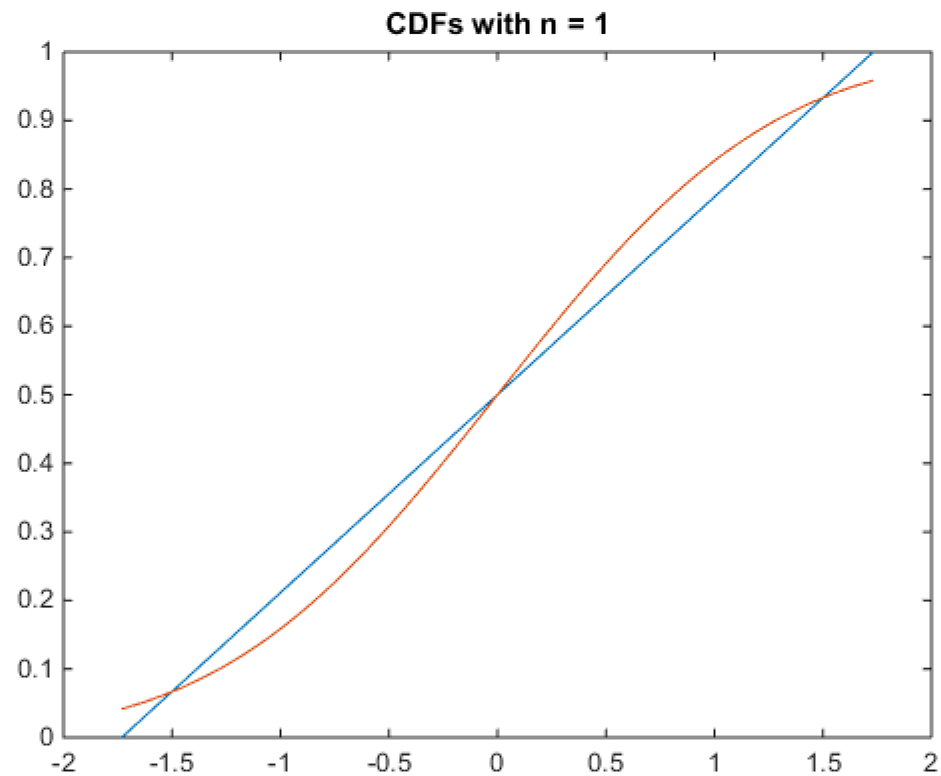
```
subplot(2,2,3);
plot(linspace(0,3,2998),sum3,'LineWidth',1.5);
hold on;
plot(linspace(0,3,2998),(1/3)*normpdf((1/3)*linspace(0,3,2998),(1.5/3),sqrt(1/36))
title('Two Convolutions');
sum4 = (1/1000)*conv(sum3,unif);
subplot(2,2,4);
plot(linspace(0,4,3997),sum4,'LineWidth',1.5);
hold on;
plot(linspace(0,4,3997),(1/4)*normpdf((1/4)*linspace(0,4,3997),.5,sqrt(1/48)));
title('Three Convolutions');
```

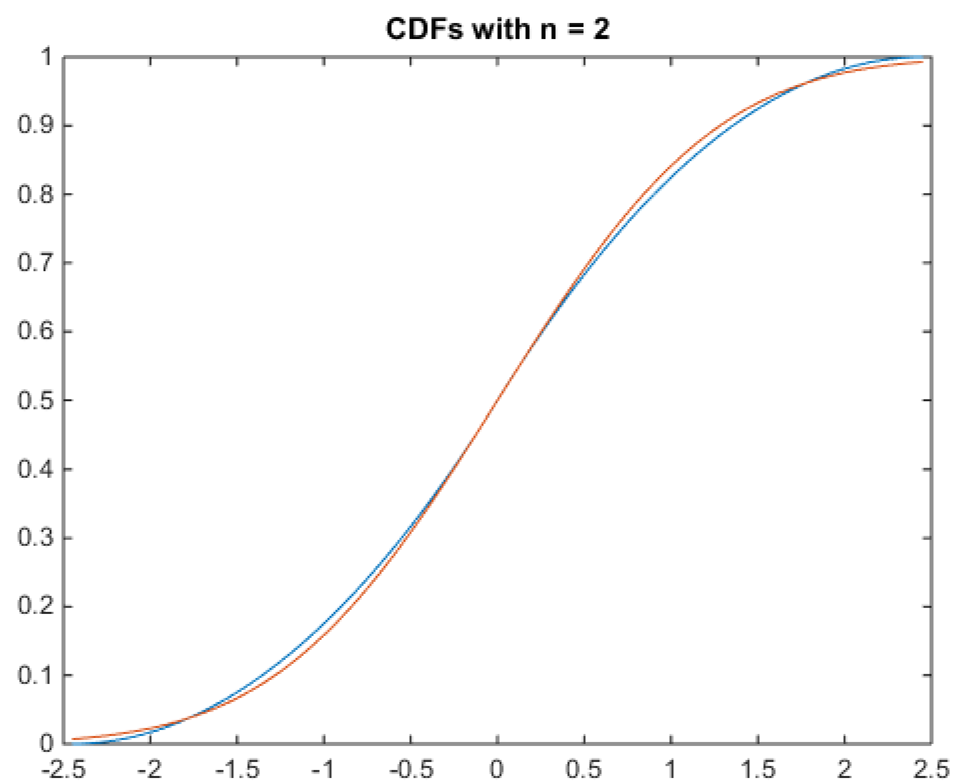


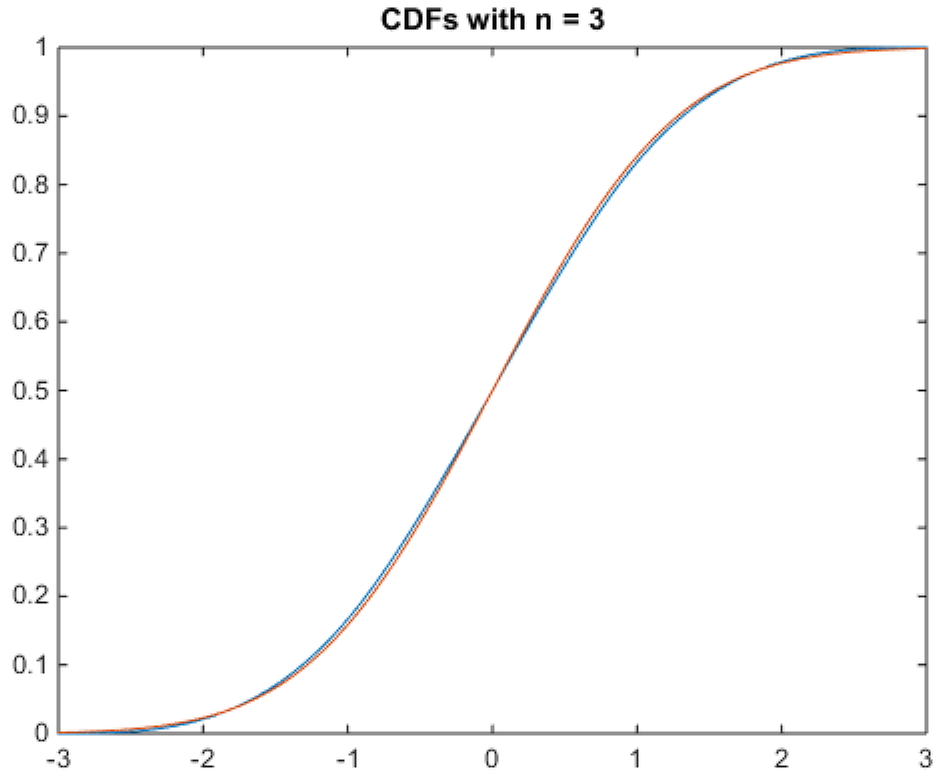
Error Analysis: The Berry-Esseen Theorem Part 1

The Berry-Esseen Theorem gives an upper bound for the difference between the scaled sample mean's cumulative distribution function and the cumulative distribution function of the standard normal distribution. Of course, this bound depends on the number of samples, as well as the standard deviation of the sampled distribution. The special version of this theorem that will be demonstrated here applies only to distributions with mean of zero, so we will discuss the uniform distribution from -1 to 1. The bound is given to be the product of C, a distribution-dependent constant, with the third absolute moment of expectation, divided by the square root of the number of samples, and the cube of the standard deviation. In the following section, the sample mean is presented in blue, while the normal cumulative distribution function is in orange. In this section we will find and show the differences between the cumulative distribution functions. In the next, we will use this data to find a potential C value. Note that conveniently, for a random variable X with the U(-1,1) distribution, a scalar multiple cX has a U(-c,c) distribution.

```
sig = sqrt(1/3);
updf = pdf('Uniform',linspace(-1/sig,1/sig,1000),-1/sig,1/sig);
ucdf = 1/max(cumtrapz(updf))*cumtrapz(updf);
figure;
plot(linspace(-1/sig,1/sig,1000),ucdf);
hold on;
plot(linspace(-1/sig,1/sig,1000),cdf('Normal',linspace(-1/sig,1/sig,1000),0,1));
title('CDFs with n = 1');
maxdif1 = max(abs(ucdf - cdf('Normal',linspace(-1/sig,1/sig,1000),0,1)));
updf2 = pdf('Uniform',linspace(-.5*sqrt(2)/sig,.5*sqrt(2)/sig,1000),-.5*sqrt(2)/si
usum2 = (6/1000)*cumtrapz((1/1000)*conv(updf2,updf2));
figure;
plot(linspace(-sqrt(2)/sig,sqrt(2)/sig,1999),usum2);
hold on;
plot(linspace(-sqrt(2)/sig,sqrt(2)/sig,1999),cdf('Normal',linspace(-sqrt(2)/sig,sq
title('CDFs with n = 2');
maxdif2 = max(abs(usum2 - cdf('Normal',linspace(-sqrt(2)/sig,sqrt(2)/sig,1999),0,1
updf3 = pdf('Uniform',linspace((-1/3)*sqrt(3)/sig,(1/3)*sqrt(3)/sig,1000),(-1/3)*s
uconv3 = (1/1000)*conv(updf3,(1/1000)*conv(updf3,updf3));
usum3 = (8/1000)*cumtrapz(uconv3);
figure;
plot(linspace(-sqrt(3)/sig,sqrt(3)/sig,2998),usum3);
hold on;
plot(linspace(-sqrt(3)/sig,sqrt(3)/sig,2998),cdf('Normal',linspace(-sqrt(3)/sig,sq
title('CDFs with n = 3');
maxdif3 = max(abs(cdf('Normal',linspace(-sqrt(3)/sig,sqrt(3)/sig,2998),0,1) - usum
```


CDFs with n = 1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2


CDFs with n = 2
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5

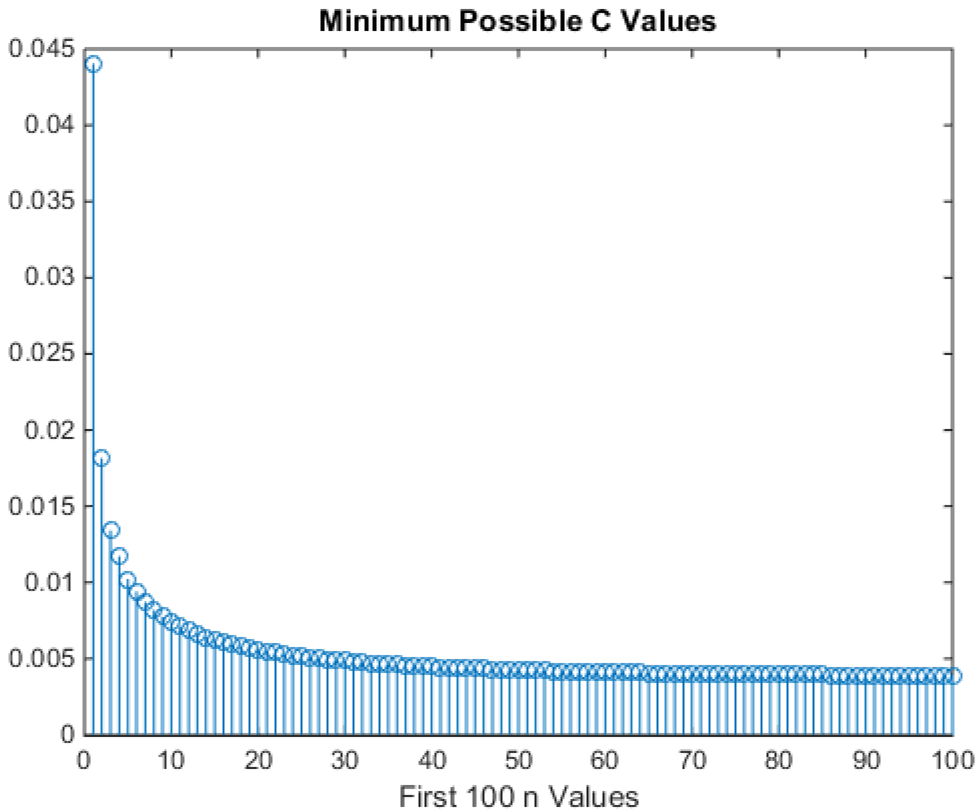


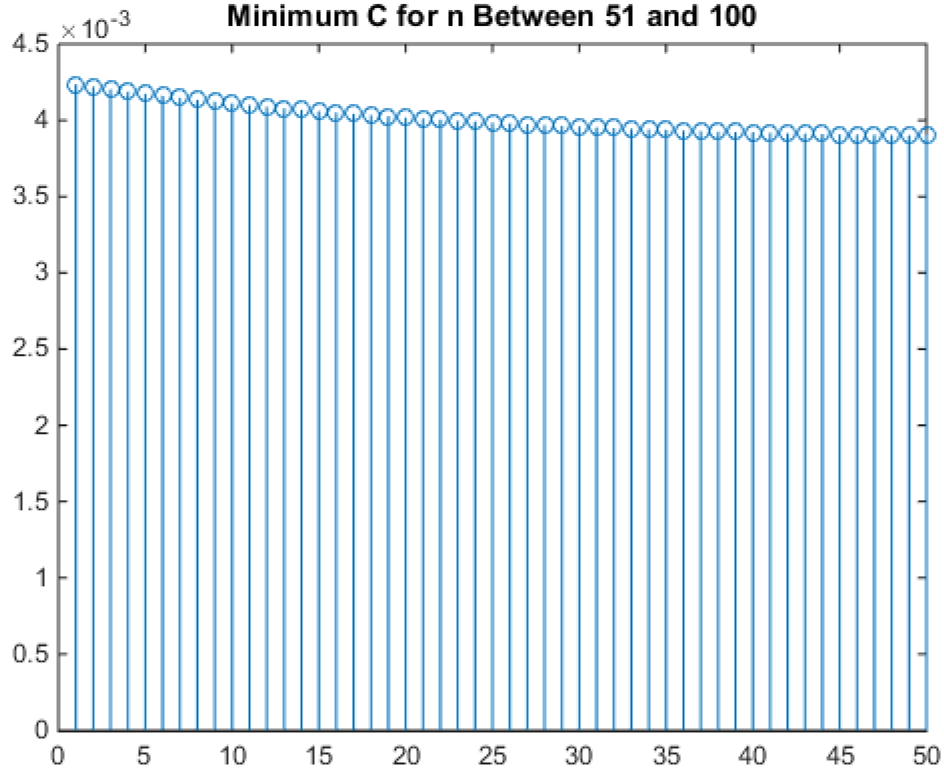
Error Analysis: The Berry-Esseen Theorem Part 2

We now have the tools necessary to begin quantifying the closeness of a central limit theorem approximation. In this section, we will attempt to further quantify this for the distribution U(-1,1). Let us first recall that the Berry-Esseen theorem requires knowledge of the absolute third central moment, which is .25 for the distribution in question. We also must know the cube of the standard deviation, which is the cube of the square root of one third, or about .2. With this information, it is easy to compute a minimum value for the constant C with our information for n=1, n=2, and n=3. We will walk through this process here, and then appreciably generalize our findings in the section. For n=1, we have saved *maxdif1* and found it to be about .0572, so C can be no less than .044. Recall that C is a constant that satisfies the inequality for all values of n. So if we look at n=3, we find that C has a minimum value of .0134; that is a lot smaller, but recall that this is not particularly meaningful without other values. What is the trend for this distribution then? We write and call a function *maxdif*, which employs the convolution and integration technique from the last section, to learn more about C.

```
sig3 = (1/sqrt(3))^3;
vec = zeros(1,100);
vec(1) = maxdif1*4*sig3;
for i = 2:100
    vec(i) = maxdif(i)*4*sig3*sqrt(i);
end
figure;
stem(vec);
title('Minimum Possible C Values');
```

```
xlabel('First 100 n Values');
figure;
stem(vec(51:100));
title('Minimum C for n Between 51 and 100');
```





A Quick Recap of Our Work with the Berry-Esseen Theorem

In the last two sections, we briefly worked with the Berry-Esseen Theorem, and saw that the minimum possible value of C for this distribution decreased with n, meaning *the* C value was given by the minimum C at n = 1, or .044. We see, however, that there is a quick trend towards stagnation, with C values near 1/4000 being nearly constant between n values of 51 and 100. With this, we can quantify our estimations. For example, we are now equipped to answer the following realistic question:

If I would like to estimate the Normal cumulative distribution function to the nearest one one-hundredth at all points, how many samples should my sample mean of U(-1,1) contain?

We could even use the pseudo-C-value, 1/4000, if we know that a sufficiently large sample will be taken. Of course for any distribution, with mean 0, we could use this exact same process again. For now, let us answer this posed problem.

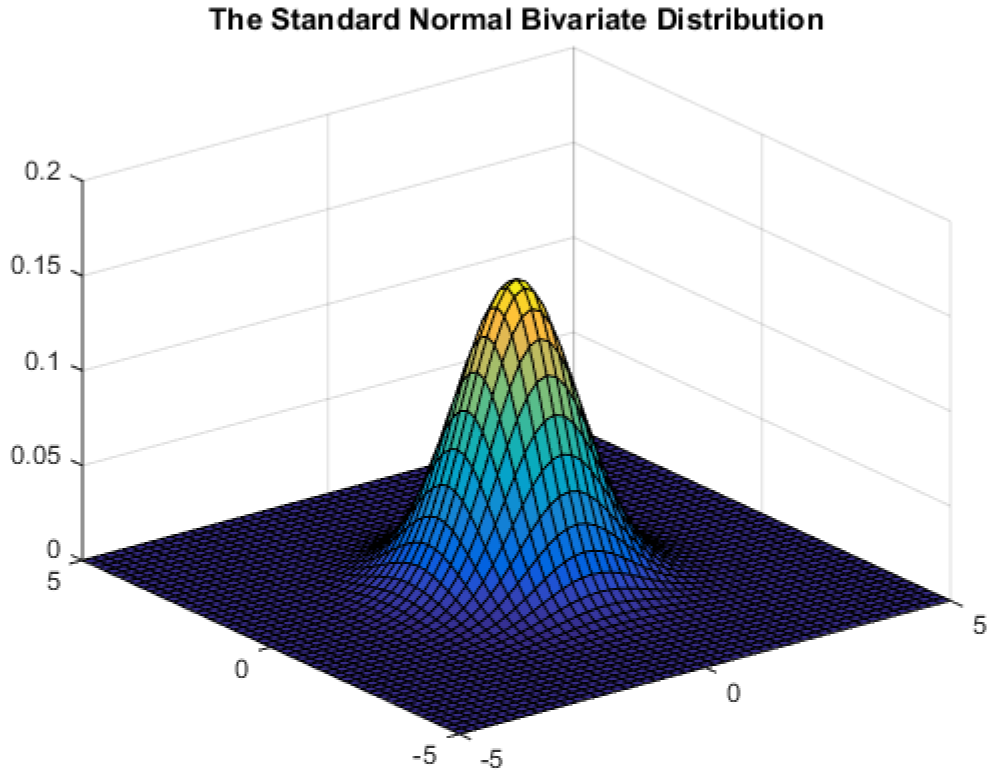
```
maxd = 1/100;
C = .044;
n = ((C*.25)/(sig3*maxd))^2;

% From this, we see that the lowest integer value of n that satisfies this
% is 33. What if the question asked for an n that satisfies the normal
% distribution to the nearest five-thousandth? Well surely we would need many
% samples for that, so use pseudo-C of 1/4000.
```

```
newmaxd = 1/5000;
pC = 4/1000;
newn = ((pC*.25)/(sig3*newmaxd))^2;

% We see that 675 samples will be enough for this estimation. Let's quickly
% check these two estimates.

wasitcorrect1 = maxdif(33);

% We see a maximum difference of one thousandth, surely less than
% one hundredth.

wasitcorrect2 = maxdif(675);

% This gives a remarkably small number, 1E-231. This shows that our
% pseudo-C value was more than enough to give us a good estimate.
```

The Multi-Dimensional Central Limit Theorem 1: One More Dimension!

The central limit theorem is actually more general than the one stated several sections ago; it actually applies to arbitrary-dimensional probability distributions. Mathematically, the meaning of a multi-dimensional distribution is fairly simple, as instead of being determined by a single random variable, it is determined by a vector of random variables. It is mathematically painful, but conceptually simple, to understand a ten-dimensional probability distribution in this way. Of course, for the limitations of visualization, we will focus here on describing the case in three dimensions. For such distributions, we are concerned with 2x1 vectors of random variables, with 2x1 mean vectors respectively associated with them. In place of variance, in this case we care about covariance, which is determined by a 2x2 matrix with entries corresponding to the variances of each random variable and the covariances of the variables with respect to one another. Abstractly this is rather confusing, so we should look at the most important such distribution: the standard multivariate normal distribution. Unsurprisingly, it is given by the 2x1 vector containing two independent standard normal random variables. This means we have a mean vector and covariance matrix as is seen in the following section, with a three-dimensional surface to illustrate the corresponding density function.

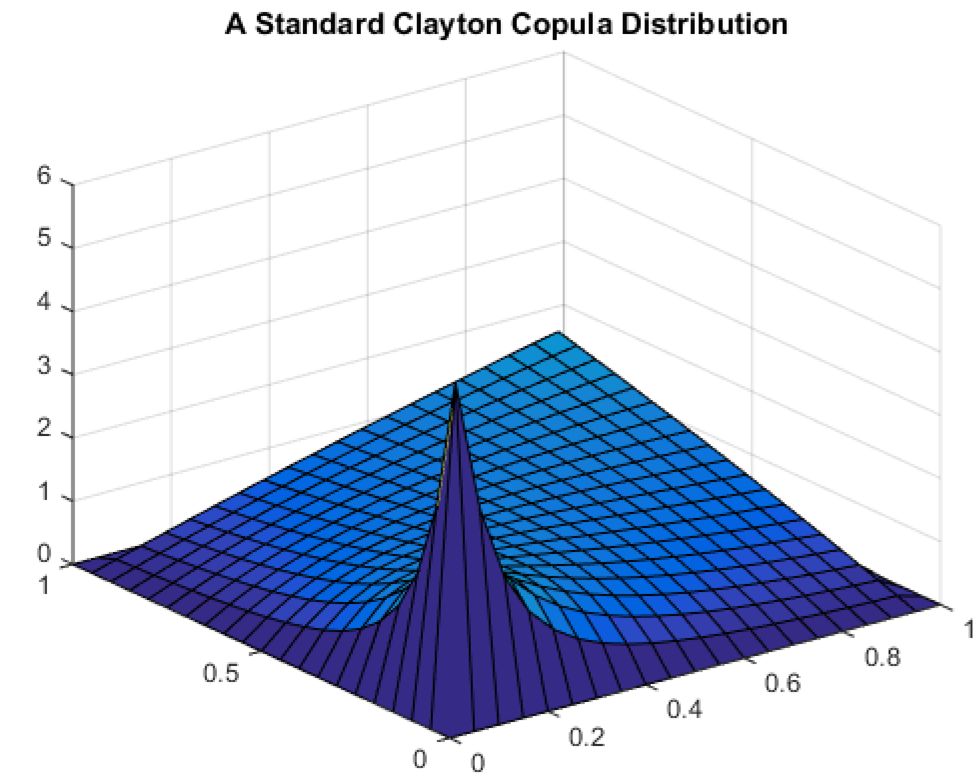
```
snorm3mean = [0 0];
snorm3cov = [1 0; 0 1];
[X,Y] = meshgrid(-5:.2:5,-5:.2:5);
Z = mvnpdf([X(:) Y(:)],snorm3mean,snorm3cov);
Z = reshape(Z,51,51);
surf(-5:.2:5,-5:.2:5,Z);
title('The Standard Normal Bivariate Distribution');
```

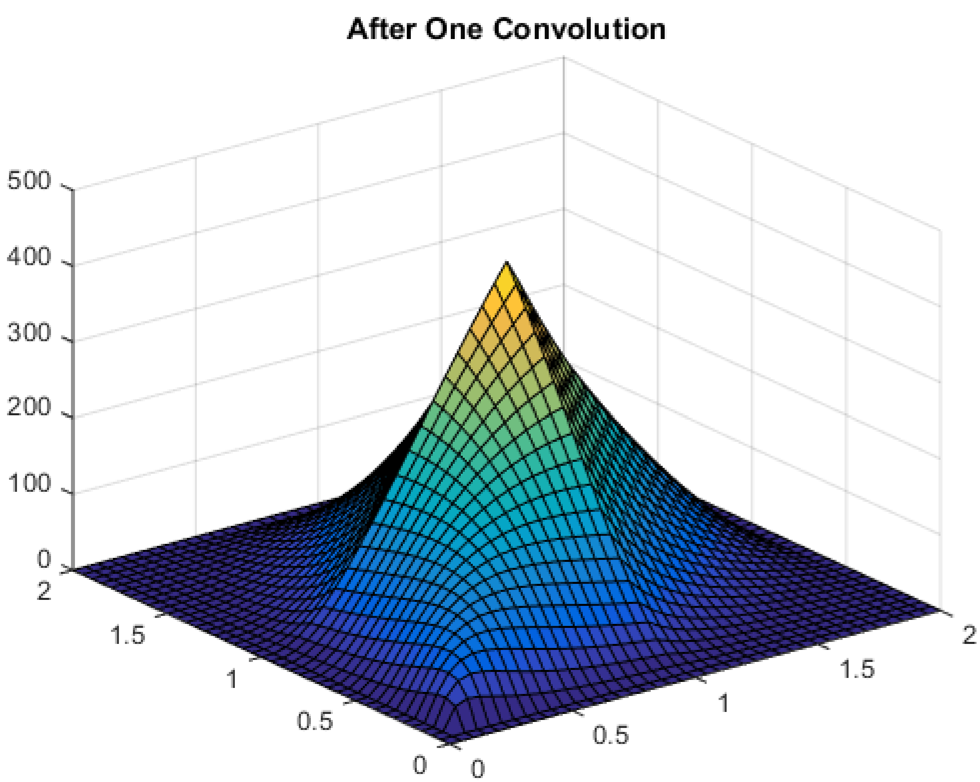


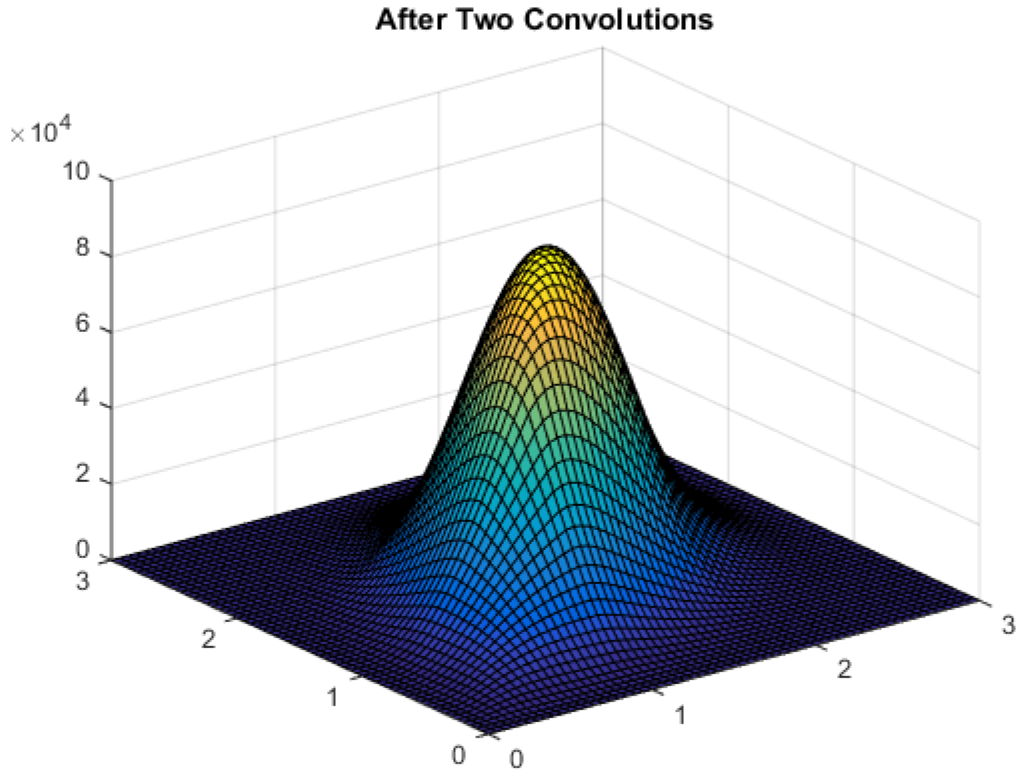
The Multi-Dimensional Central Limit Theorem 2: The Theorem

As you would expect, our generalized theorem states that sample means of nx1 independent and identically distributed random vectors tend towards normal distributions as more samples are taken, where "normal" looks a bit like the surface shown above. For the sake of mathematical cleanliness and brevity, we will not go into the error bounding of such a limit, but we will illustrate its effect in this section. We will look at a common distribution; the Clayton copula distribution; and see how its samples transform to look more like the normal distribution shown above. This method is simply the convolution method from earlier, however we use 2-dimensional convolution instead.

```
[X1,Y1] = meshgrid(0:.05:1,0:.05:1);
Z1 = copulapdf('Clayton',[X1(:), Y1(:)],1);
Z1 = reshape(Z1,21,21);
figure;
surf(0:.05:1,0:.05:1,Z1);
title('A Standard Clayton Copula Distribution');
conved = conv2(Z1,Z1);
figure;
surf(linspace(0,2,41),linspace(0,2,41),conved);
title('After One Convolution');
conved2 = conv2(Z1,conved);
figure;
surf(linspace(0,3,61),linspace(0,3,61),conved2);
title('After Two Convolutions');
```


A Standard Clayton Copula Distribution
6
5
4
3
2
1
0
1
0.5
0
0
0.2
0.4
0.6
0.8
1


After One Convolution
500
400
300
200
100
0
2
1.5
1
0.5
0
0
0.5
1
1.5
2



A Conclusion

With this, we conclude our survey of the central limit theorem. I hope that this has been an informative and interesting read. This document was created in MATLAB R2014b by Brian Frost-LaPlante, Cooper Union Electrical Engineering class of 2019, as part of the final project for ECE-210, a MATLAB seminar. Thank you for reading!

Published with MATLAB® R2014b