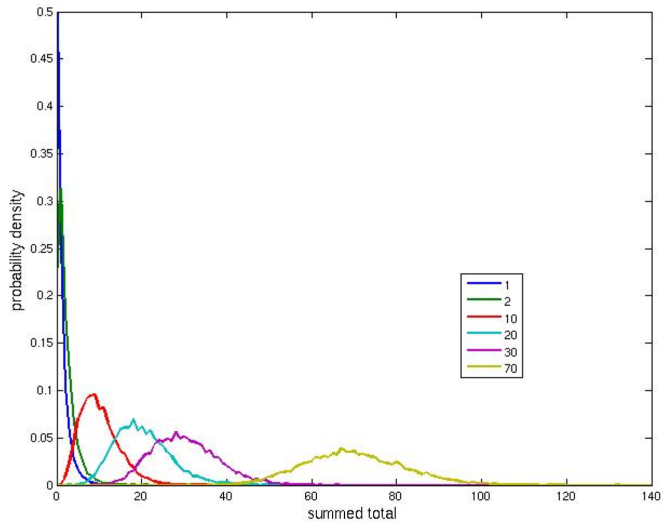

Figure 1: Sample χ^2 distributions for k=1, 2, 10, 20, 30, and 70.

χ^2 and the Central Limit Theorem

In class discussion we briefly pondered the effect of the central limit theorem on the χ^2 distribution. You will recall that a χ^2 variable is defined as:

$$Q = \sum_{i=1}^{k} x_i^2, \qquad (1)$$

where x_i have a Gaussian distribution, k is the number of degrees of freedom, and Q has a χ^2 distribution. Since the x_i all have the same mean and variance, the x_i^2 must also have the same mean and variance, and Q represents a summation of random variables. As k increases, in accordance with the central limit theorem, Q should become progressively more Gaussian. Figure 1 shows sample χ^2 distributions for a few select values of k. Clearly as k increases, the distribution looks more Gaussian.

In class, we asked how Gaussian the χ^2 distribution really is, and in particular, how it compares with a uniform distribution, y_i with values randomly distributed between 0 and 1, so that a summation would be:

$$Q_{uni} = \sum_{i=1}^{k} y_i, \qquad (2)$$

or for that matter with a distribution developed from squares of uniformly distributed data:

$$Q_{uni^2} = \sum_{i=1}^{k} y_i^2. \qquad (3)$$

Figures here help to illustrate the effects. First, the top panel of Figure 2 shows the mean of the distributions as a function of k. Not surprisingly, as you sum more variables, the mean increases linearly. Since the mean increases with k, obviously the fact that all data are positive will become less important, and we will not have to be so concerned about the skewness of the original data. This is clear in Figure 1: for $k = 1$, the distribution is peaked near zero and highly skewed, while for $k = 70$ it has a large non-zero mean and looks nearly Gaussian.

So how quickly do these distributions converge towards a Gaussian? Some basic characteristics of Gaussian distributions are that they have zero skewness (derived from the third moment of the PDF) and a kurtosis of 3 (derived from the fourth moment of the PDF).[1] The middle panel of Figure 2 shows skewness for the 3 distributions as a function of k, and the bottom panel of Figure 2 shows kurtosis as a function of k. In this particular case, the kurtosis of the χ^2 distribution clearly asymptotes to 3, but this occurs more slowly than for the uniform distribution. Interestingly, the uniform and uniform squared data show similar convergence properties. The skewness of the χ^2 distribution never quite reaches zero in these examples, although as k approaches ∞ presumably the skewness would gradually become negligibly small.

There are a few lessons to draw from these figures. First, and most importantly, the χ^2 distribution does not converge terribly quickly to a Gaussian form, so it's probably not a good idea to abandon χ^2 statistics for moderately large k (and you'll have to decide what moderately large means). Second, if the orginal data did not have a Gaussian distribution, then the sum of squared variables may not have a χ^2 distribution, and in the case here of the squared uniform distribution, its properties may converge more rapidly to resemble a Gaussian. For such cases, χ^2 statistics may give poor representations of data uncertainties.

Here's some Matlab code to replicate the tests I did here:

```
a=randn(10000,100); % define a matrix of Gaussian random numbers
a2=rand(10000,100); % define a matrix of uniform random numbers
b=a.^2; % compute the Gaussian noise squared for the chi-squared distribution
b2=a2.^2; % compute an equivalent quantity for the uniform distribution
c=cumsum(b,2); % compute the chi-squared data for k=1,100
c2=cumsum(b2,2); % compute the uniform random number squared and summed values
c3=cumsum(a2,2); % compute the summation of uniform random numbers

% generate histograms for the data
for i=1:100
[n1(i,:),n2(i,:)]=hist(c(:,i),[0:140]);
[m1(i,:),m2(i,:)]=hist(c2(:,i),[0:140]);
[o1(i,:),o2(i,:)]=hist(c3(:,i),[0:140]);
end

% plot kurtosis
figure(1)
plot(1:100,[kurtosis(c); kurtosis(c2); kurtosis(c3)])
title('kurtosis')
legend('\chi^2 distribution','squared uniform distribution','uniform distribution')
xlabel('k')
ylabel('kurtosis')

% etc. etc.

% plot chi-squared distribution
N=sum(n1(1,:));
figure(4)
plot(n2(1,:),n1([1 2 10 20 30 70],:)/N);
axis([0 140 0 .5])
legend('1','2','10','20','30','70')
ylabel('probability density')
xlabel('summed total')
```

[1] Some people like to use "excess kurtosis" rather than kurtosis, which means that they subtract 3 from the kurtosis, and sometimes the "excess kurtosis" is simply called kurtosis.

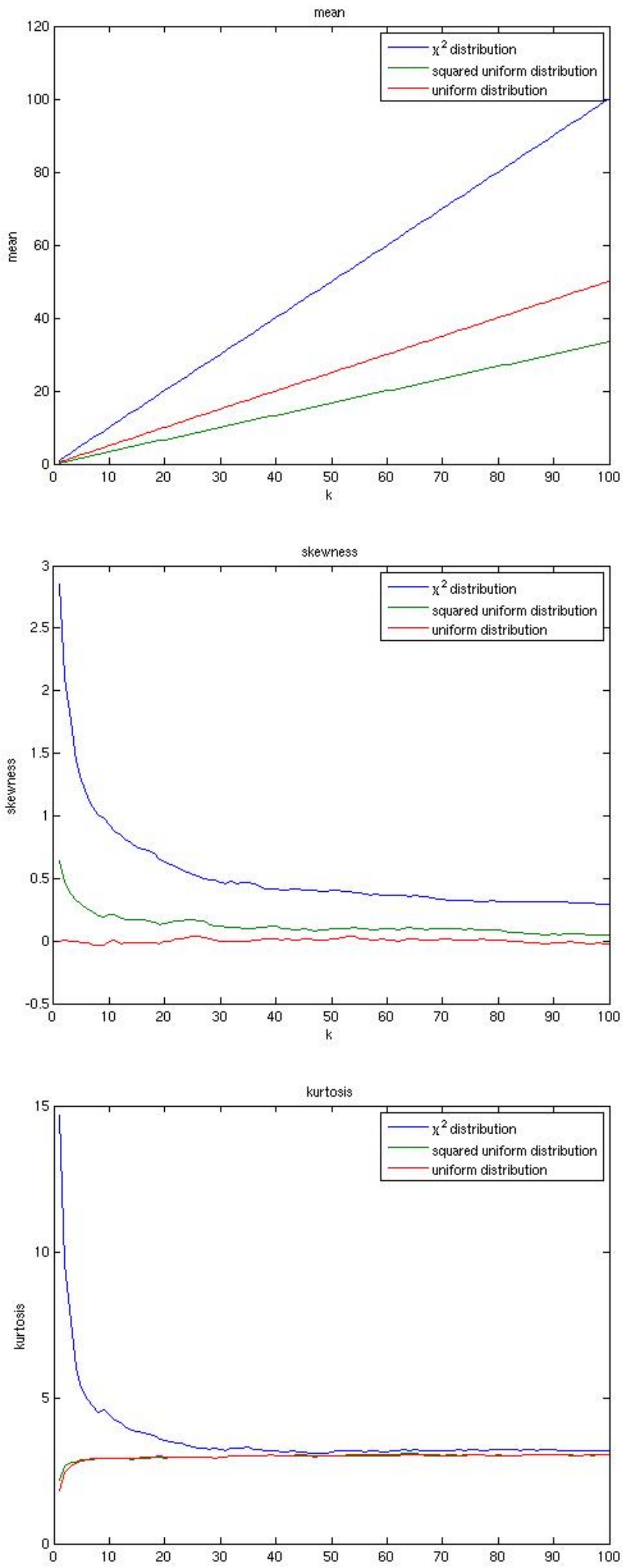

Figure 2: (Top) Mean of random data constructed as a summation of k variables, shown as a function of k. The χ^2 distribution is blue, uniform squared is green, and uniform is red. (Middle) Skewness of same random data. (Bottom) Kurtosis of same data.