EXERCISE ON PRINCIPAL COMPONENTS, CLT

(1) Suppose $P[X_T = 0] = .5 - 1/T$, $P[X_T = k/T] = .4/T$, $k = 2, \ldots, T$, $P[X_T = T] = .4/T$, and $P[X_T = 1 + 1/T] = .1 + 1/T$.
   (a) Show that $EX_T \xrightarrow[T\to\infty]{} K > 0$ and $\text{Var}(X_T) \xrightarrow[T\to\infty]{} \infty$ as $T \to \infty$. Find K.
   (b) Does the sequence $\{X_T\}$ converge in distribution? If so, prove that it converges and describe the limiting distribution. Does the limiting distribution have expectation K and infinite variance? If there is no limiting distribution, prove that there is not.

(2) Use a computer to generate 100 independent draws from the distribution of a random variable that is -1 with probability 2/3 and 2 with probability 1/3.
   (a) Plot histograms for the sum of 5, of 10, of 40, and of 100 of these random variables. On each histogram plot, also plot a normal pdf that has the same variance as the theoretical (not the sample) variance of the random variable whose histogram you're plotting.
   (b) Plot sample cdf's for the sums of 5, 10, 40 and 100. On each of these plots show also the normal cdf for the same theoretical variance.

   Matlab has a `hist` *command that automatically generates histograms. The only effort required from you is figuring out how to scale the normal pdf to match the histogram. To plot a new curve on an existing plot, use the matlab* `hold` *command. Recall that the sample cdf jumps up by 1/N, where N is sample size, at each data point. In matlab, use* `sort` *first, then* `plot(x,(1:length(x))/N)`*. Of course you can use other software than matlab if you prefer.*

(3) The course web site has links to data for three correlation matrices, representing correlations of deviations from "trend" for GNP in 10 countries, for the three periods pre-war, inter-war, and post-war. They come from a paper by Backus and Kehoe in the AER, September 1992. For each of these matrices, find the first principal component and show a vector that contains the proportion of variance in each country's GNP that is accounted for by the first principle component. Do your results suggest support for the claim by Backus and Kehoe that cross country links were strongest in the inter-war, next strongest in the post-war, and weakest pre-war? Do they suggest anything about changes in which countries are most strongly linked to the rest of the world?

   The matrices and the list of country names are all in the file `bkcor.mat`*, which can be read in to matlab with the* `load bkcor` *command. To see the list of variable names after loading the file, use* `whos`*. The same information is available on the web site in four separate text files, in case you want to work with them outside matlab.*

The matrices we are using here are not exactly those in the original article. None of the original correlation matrices are positive definite. This is possible because each correlation was calculated from data over a different sample period, because of data availability problems. However, the estimation of these correlations could certainly have been made more accurate if the authors had used the fact that the true correlation matrix has to be positive definite. For this exercise I have just damped down the sizes of the off-diagonal elements until the matrix became positive definite.