Problem Set 3

1. (a) Writing the expectation explicitly, one gets:

$$
\begin{aligned}
E(X_T) = \frac{\sum_{k=2}^{T} k}{T}\frac{0.4}{T} + T\frac{0.4}{T} + (0.1+\frac{1}{T})(1+\frac{1}{T}) \\
= \left(\frac{T^2+T-1}{2T}\right)\frac{0.4}{T} + .4 + .1 + 1.1\frac{1}{T} + \frac{1}{T^2} \\
= 0.2 + \frac{1.3}{T} + .5 + \frac{0.8}{T^2} \xrightarrow[T\to\infty]{} 0.7
\end{aligned}
$$

Also, writing the expression for $E(X_T^2)$ one gets:

$$
E(X_T^2) = \frac{\sum_{k=2}^{T} k^2}{T^2}\frac{0.4}{T} + T^2\frac{0.4}{T} + (0.1+\frac{1}{T})(1+\frac{1}{T})^2
$$

Since all the three terms in the sum are positive and $0.4T^2/T = 0.4T \xrightarrow[T\to\infty]{} \infty$, $E(X_T^2) \to \infty$. So, $\mathrm{Var}(X_T) \to \infty$.

(b) For any bounded, continuous function f of X_T,

$$
\begin{aligned}
E[f(X_T)] = \left(.5-\frac{1}{T}\right) f(0) + \frac{.4}{T}\sum_{k=2}^{T} f\left(\frac{k}{T}\right) + \frac{.4}{T} f(T) + \left(.1+\frac{1}{T}\right) f\left(1+\frac{1}{T}\right) \\
\xrightarrow[T\to\infty]{} .5f(0) + .4\int_0^1 f(t)\,dt + 0 + .1f(1)\,.
\end{aligned}
$$

This latter expression is just $E[f(Z)]$ for a random variable Z that has probability .5 on the point $Z=0$, .4 uniformly spread over (0,1), and .1 on the point $Z=1$. So the distribution of this Z is the limit in distribution (or weak limit) of the distributions of the X_T's.

The cdf of the limiting distribution is

$$
F_Z(a) = \begin{cases} 0 & \text{if } x < 0 \\ 0.4x + 0.5 & \text{if } x \in [0,1) \\ 1 & \text{otherwise.} \end{cases}
$$

Notice that the expectation of the limiting distribution is 0.3 and its variance is finite.

Another route to proving the result is to use the characterization of convergence in distribution as convergence of cdf's at points of continuity, which are here $\mathbb{R}\setminus\{0,1\}$. For each T, the df is given by:

$$F_{X_T}(x) = \begin{cases} 0 & \text{if } x < 0 \\ 0.5 - 1/T + \sum_{k=2}^{T} 1_{\{x \geq k/T\}} \frac{0.4}{T} + 1_{\{x \geq 1+\frac{1}{T}\}}(0.1 + \frac{1}{T}) & \text{if } x \in [0,T) \\ 1 & \text{otherwise.} \end{cases}$$

where $\lfloor xT \rfloor$ is the largest integer that is smaller than xT. If $x < 0$, the sequence is identically null and converges trivially to zero. If $x \in (0,1)$, then

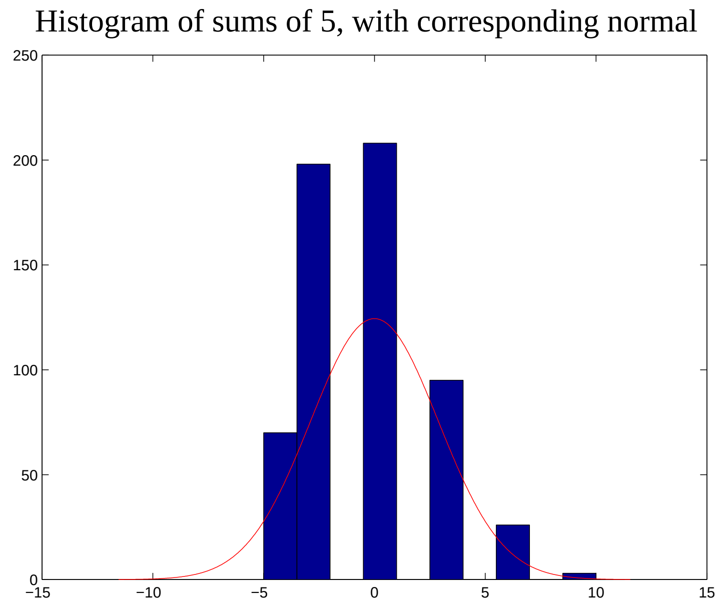
$$F_{X_T}(x) = 0.5 - 1/T + \frac{0.4\lfloor xT \rfloor}{T} \to 0.5 + 0.4x$$

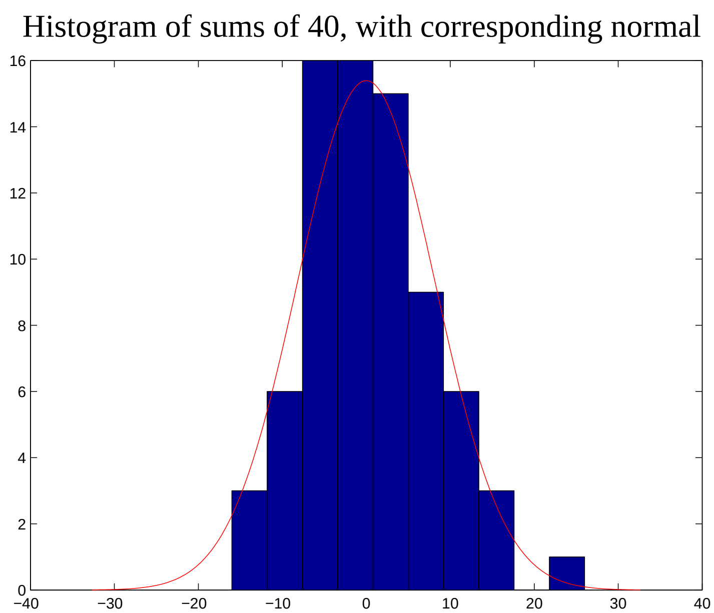
If $x > 1$, for T large enough, $F_{X_T}(x) = 1 - 1/T$ and $1 - 1/T \to 1$. So the convergence in distribution is achieved.

2. This exercise was meant to familiarize you with generating random numbers on a computer, and to illustrate the CLT. Only as I wrote up this answer did I realize that through a typo I started out asking you to generate 100 random draws, and then later asked for a histogram of a sum of 100 of "these" random variables. I meant to make the number of initial draws 1000, and to get reasonable histograms for the sum of 100 should probably have made it 3000.

   An example of a matlab program that produces the histograms for sums of 40 and of 5:

```
z=rand(3000,1);
x=-1+3*(z>.666666666667);
z=-4:.01:4;
npdf=exp(-.5*z.^2)/sqrt(2*pi);
s=sqrt(1.6666666667) % standard deviation of the r.v.
x5=sum(reshape(x,5,600)); %quick way to get sums of 5 successive elements.
                          %x5 is 1x600.
x40=sum(reshape(x,40,75));%x40 is 1x75.
[nh,xh]=hist(x40);
dh=xh(2)-xh(1) % the spacing of the histogram
hold,plot(z*s*sqrt(40),dh*75*npdf/(s*sqrt(40)),'r')
hold off
[nh,xh]=hist(x5);
dh=xh(2)-xh(1) % the spacing of the histogram
hold,plot(z*s*sqrt(5),dh*600*npdf/(s*sqrt(40)),'r')
```

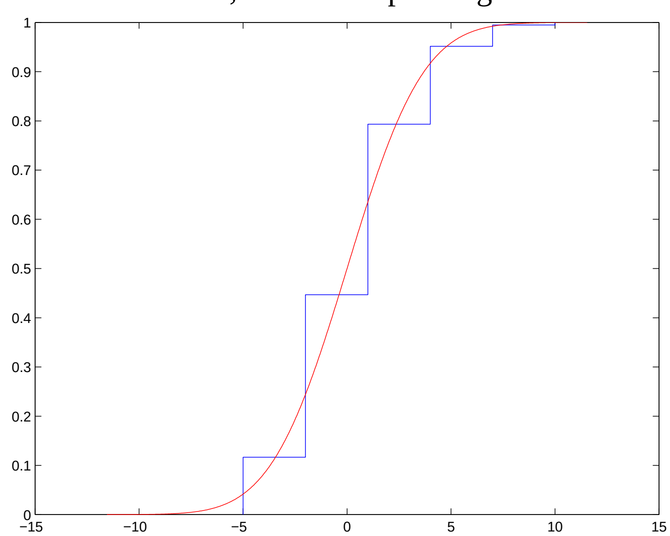




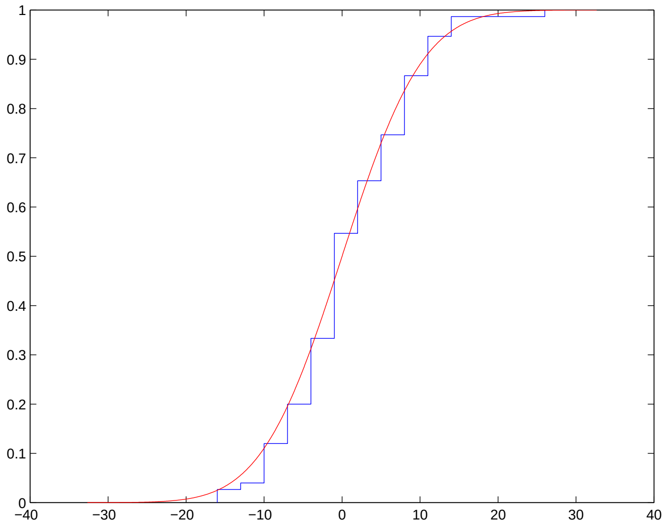
The suggested code for the cdf was not quite right—it produces slight upward slopes between the jumps in the cdf where there should be flats. Here is correct code, with two sample graphs. Note that this code uses a routine `erff.m` that I wrote myself. It uses matlab's `erf.m` to construct the normal cdf. This is just one line of code, which is displayed in matlab 6.5's, and probably also 6.1's, help entry for `erf.m`. If the department computers have the statistics package for matlab, then `normcdf.m` was available, which behaves just like `erff.m`. However I should have mentioned in the problem statement how you were supposed to come up with the normal cdf.

```
x5s=sort(x5);
x40s=sort(x40);
x40double=zeros(1,2*length(x40s));
x40double(1:2:end)=x40s;
x40double(2:2:end)=x40s;
df40=zeros(size(x40double));
df40(1:2:end)=(0:length(x40s)-1)/length(x40s);
df40(2:2:end)=(1:length(x40s))/length(x40s);
plot(x40double,df40)
hold,plot(z*s*sqrt(40),erff(z),'r')
% some save and/or print commands would go in here
x5double=zeros(1,length(x5s)*2);
x5double(1:2:end)=x5s;
x5double(2:2:end)=x5s;
df5=zeros(size(x5double));
df5(1:2:end)=(0:length(x5s)-1)/length(x5s);
df5(2:2:end)=(1:length(x5s))/length(x5s);
plot(x5double,df5)
hold,plot(z*s*sqrt(5),erff(z),'r')
% some save and/or print commands would go in here
```

cdf for sums of 5, with corresponding normal cdf



cdf for sums of 40, with corresponding normal cdf



3. Here is a sequence of matlab commands that generates the eigenvalue decompositions needed to answer this question:

```
[vpre,dpre]=eig(bkcpre);
dpre=diag(dpre)
[vint,dint]=eig(bkcint);
dint=diag(dint)
[vpost,dpost]=eig(bkcpost);
dpost=diag(dpost)\\
[vpre(:,10) vint(:,10) vpost(:,10)]
```

Here are the resulting three vectors of coefficients. Note that the output from the commands above, which for each matrix lists the eigenvalues, will have made clear that the 10'th eigenvector corresponds to the largest eigenvalue for each decomposition.

Australia	0.4049	0.3282	0.3545
Canada	0.5226	0.3684	0.3569
Denmark	0.2981	0.1587	0.2612
Germany	0.1761	0.2704	0.2392
Italy	0.2045	0.3025	0.3200
Japan	0.1046	0.3789	0.2404
Norway	0.3680	0.3298	0.3851
Sweden	0.4683	0.2767	0.2729
UK	-0.1543	0.3678	0.3110
US	0.1086	0.3203	0.3767
eigenvalue	2.3813	3.3004	3.2433

For the interwar and postwar periods the two vectors are quite similar, with positive weights all lying in the interval [.15,.38]. Their correlation (since they are unit length, this is just their dot-product) is .38. But the prewar vector is quite different, with the UK, the US and Japan all showing little connection to the principal component in that period. The fact that the weights are all, or almost all, positive, and similar, suggests that a considerable part of the variation in country GDP's was associated with the worldwide average level of economic activity. The proportion of variance in country i explained by the international component is $q_i^2\lambda$, where q_i is the coefficient for country i in the table above and λ is the largest eigenvalue, listed at the bottom of the table above. The proportion of variance explained averages .24 in the prewar, .33 in the interwar, and .32 in the postwar period. In the prewar period half the countries had less than 10% of variance attributable to the common factor, and the US had the least variance explained by the common factor. In the latter two periods the proportions of variance explained in the US are 34% and 46%