Central Limit Theorem

Part 1.

a) Suppose $F_{\mathbf{X}}(x)$ and $f_{\mathbf{X}}(x)$ are respectively the CDF and PDF of a continuous random variable $\mathbf{X}$. Let $\mathbf{Y} = F_{\mathbf{X}}(x)$. Determine $f_{\mathbf{Y}}(y)$, i.e. the PDF of $\mathbf{Y}$.

b) Download one of the images from "www.site.uottawa.ca/∼wabed" and display it in Matlab. (Hint: Use subroutines "`imread`" and "`imshow`").

c) Plot the relative frequency of the pixel values in the image. Matlab stores the pixel values in the image in a "uint8 array" and in order to be able to calculate the relative frequency of pixel values you have to convert them to "double array"'. To do that, use the subroutine "`double`".

d) Treat the relative frequency above as the PMF of a random variable $\mathbf{X}$. Plot the CDF $F_{\mathbf{X}}(x)$.

e) Generate a new image by changing the value x of each pixel in the above image to $[F_{\mathbf{X}}(x) \times 255]$. Display the new modified image, and plot the histogram of it. Explain what you see!

Part 2.

According to the central limit theorem, if $\mathbf{X}_1$, $\mathbf{X}_2$,...., $\mathbf{X}_n$ are identically distributed independent random variables with mean μ and variance σ^2, their average $\sum_i \mathbf{X}_i/n$ will approach normal distribution with mean μ and variance σ^2/n. We want to test this theorem empirically in Matlab.

a) Generate a 1000 random numbers uniformly distributed on $[0, 1]$. Compute their average.

b) Repeat the previous step 1000 times and save the resulting averages into a vector $\mathbf{M}$. Draw a histogram of $\mathbf{M}$. Does it resemble a normal distribution?

c) Normalize the histogram of $\mathbf{M}$ to make it into the probability mass function (PMF). To make an approximation to the density function, you also need to divide the histogram by the width of the bin. On the same plot draw the graph of the normal with the mean and variance as dictated by the central limit theorem (what are the mean and the variance you should use?). The normalized histogram should approximate the normal distribution.

Part 3.

When an analyst attempts to fit a statistical model to observed data, he or she may wonder how well the model actually reflects the data. How "close" are the observed values to those which would be expected under the fitted model? One statistical test that addresses this issue is the chi-square goodness of fit test. This test is commonly used to test association of variables in two-way tables, where the assumed model of independence is evaluated against the observed data. In general, the chi-square test statistic is of the form

$$\chi^2 = \sum \frac{(observed - expected)^2}{expected}$$

Suppose we flip a coin 100 times and record the number of tails and heads. If it is hypothesized that the coin is unbiased (i.e. $P(T) = 0.5$), would you reject this hypothesis at 1% and 5% significance level?.

To answer this question use Matlab to perform this experiment and compute the χ^2 statistic. (Use Table 3.5 p.142 in the book).