Tutorial problems:

Ghahramani: Section 11.1 - Problems 2, 4, 13; Section 11.2 - Problems 5, 15. Also the following questions:

1. Let \( X_i \overset{d}{=} N(\mu_i, \sigma_i^2) \) for \( i = 1, 2 \cdots \) be mutually independent. Use moment generating functions to derive the distribution of the linear combination
\[
L = \sum_{i=1}^{n} a_i X_i
\]
where \( \{a_i\} \) are constants.

2. In Lab 4, Exercise A you considered the distribution for the number of heads \( X \overset{d}{=} Bi(n, H) \) in \( n \) tosses of a coin chosen at random from a mixture of coins with different probabilities of a head \( H \) (ie \( H \) is a random variable). In question 8 of Exercise A (then an extension task) you were asked to speculate on the distribution for \( X \) if \( H \overset{d}{=} R(0, 1) \). You can now solve this question theoretically. Write down the pmf for \( X \) when \( H \overset{d}{=} R(0, 1) \) by first calculating the pgf \( P_X(z) \). [Hint: To obtain the pgf first calculate the conditional pgf \( P_{X|H}(z) \)].

3. Let \( X, Y \) and \( A \) be independent random variables, with known moment generating functions \( M_X(t), M_Y(t) \) and with \( P(A = 1) = 1 - P(A = 0) = p \in (0, 1) \). Compute the mgf of the random variable:
\[
T = AX + (1 - A)Y.
\]
The distribution of \( T \) is called a ‘mixture’ of the distributions of \( X \) and \( Y \). [Hint: Make use of the conditional mgf \( M_{T|A}(z) \)].

4. Let \( A_n = \frac{S_n}{n} \) denote the sample average of \( n \) observations on the discrete distribution with pmf
\[
p_X(x) = \frac{1}{6} \quad x = 1, 2, 3, 4, 5, 6.
\]
Compute an approximate value for \( P(3 < A_n < 4) \) for \( n = 12 \) and for \( n = 24 \).

5. Let \( A_n = \frac{S_n}{n} \) denote the sample average of \( n \) observations on a gamma distribution \( \gamma(2, \frac{1}{4}) \). Compute an approximate value for \( P(7 < A_n < 9) \) for \( n = 128 \) and for \( n = 64 \).

Additional problems: Ghahramani: Section 11.1 - Problems 8, 12, 14; Section 11.2 - Problems 2, 3, 4.

Final problem sheet: Note that a problem sheet will be handed out on Wednesday in the last week of semester. This sheet will provide practice problems for the topics covered in the last few lectures.

Note: There are no homework questions in this handout. The final homework to be submitted is Homework set 10 due in the last tutorial for the semester.
In this lab you

- illustrate the Central Limit Theorem and convergence in distribution by simulating sequences of distribution functions converging to the distribution function for a standard normal distribution.

- use StatPlay to investigate the Central Limit Theorem further (your tutors will provide a brief overview of how to use StatPlay for this purpose).

- perhaps have time to review previous lab exercises if required with the assistance of your tutor.

Exercise A - Central Limit Theorem - Graphs of distribution functions

Suitably modified the Matlab m-file **Lab11ExA.m** will generate estimates of the distribution functions \( F_n \) for the standardised random variables \( Z_n \) as defined on lecture slide 433. Z_n standardises the sample sum \( S_n = X_1 + \cdots + X_n \) for a sample of size \( n \). The \( n \) independent observations \( X_1, \ldots, X_n \) in the sample all have the same underlying distribution \( X \) with mean \( \mu \) and standard deviation \( \sigma \). The Central Limit Theorem tells us that the dfs \( F_n \) should tend to the df for a standard normal distribution \( \Phi(x) \) irrespective of the form of the underlying distribution \( X \).

Lab11ExA plots a sequence dfs on any one run - you can specify the selected sequence of \( n \) values in the Initialisation section of the program. On each plot the selected dfs are compared with the df for a standard normal (plotted as a dotted black line). The program starts with the underlying distribution \( X \) set to be a discrete Bernoulli distribution with \( p = 0.5 \). Don’t worry too much about understanding the main plot command in the program.

1. Run the program for \( n = [3, 5, 10, 50] \). For \( X \) Bernoulli what is the distribution of \( S_n \)? Use your answer to check the distribution functions plotted for \( n = 3 \) and 5 by using an appropriate pmf. Check both the height of some jumps and the position at which they occur.

2. Run the program for some other sequences with larger \( n \) values and observe the rate of convergence (for \( n \) too large the run time for the program might be quite long).

3. Change the value of \( p \) to 0.25 by changing the line in the function Finverse. This changes the mean \( \mu \) and standard deviation \( \sigma \) so you will need to calculate the new values and type them into the ‘Initialisation section’. Repeat the above steps and review the rate of convergence as \( n \) increases.

4. To change the underlying distribution you need to change the function ‘Finverse’. Comment out the first line and uncomment the line to generate observations on \( R(0, 1) \). Again you will need to change the values of \( \mu \) and \( \sigma \) specified in the ‘Initialisation section’. The CLT works remarkably quickly for the symmetric uniform distribution so you need only consider very small values of \( n \) to see convergence to the normal.
5. The program can generate observations on any distribution using the inverse transformation method by appropriate changes to the function ‘Finverse’. A line of code for generating $X \sim \exp(\lambda)$ is already included - note that $\lambda$ is set in the Initialisation section. Uncomment this line and change $\mu$ and $\sigma$ so that they are calculated from the given value for $\lambda$. Then investigate the convergence to normality for a variety of different $n$ and $\lambda$ values. In particular note the asymmetry in the df s for small values of $n$ corresponding to the skewness in the exponential distribution.

6. **Optional:** Add code to generate observations on some other distributions using the inverse transformation method and investigate the convergence to normality.

**Exercise B - Central Limit Theorem - StatPlay**

Your tutors will give the whole group a brief overview of the functionality for the StatPlay Sampling playground which can be used to explore the Central Limit Theorem. Rather than graphing the distribution function StatPlay graphs estimates of the pdf s. You should replicate some of the examples illustrated in Exercise A. Also try the Central Limit Theorem for a ‘Freehand’ distribution which you can draw yourself in any shape you wish.

**Exercise C - Revision**

If you have questions on any previous labs you may wish to take advantage of any remaining time to ask your tutors for assistance.