Question 1

Part a

We have a total of 10 topics and need to pick 3, thus the amount of possible combinations is $\binom{10}{3} = 120$.

Part b

To find an expression for the probability that none of the n topics studied come up, I first find an expression for the opposite. That is at least 1 of the topics studied come up in the exam. To find that expression, I use the amount of combinations possible for 3 topics drawn out of 10. I then need to find the amount of combinations such that one or more questions out of n studied come up. This comes down to: $\binom{n}{3}$. Thus, the probability that one or more questions studied come up in the exam is:

$$\frac{\binom{n}{3}}{\binom{10}{3}} \tag{1}$$

We are looking for the opposite thus the probability that none of the n studied topics come up is:

$$1 - \frac{\binom{n}{3}}{\binom{10}{3}} \tag{2}$$

Part c

Knowing that the above expression gives us the probability such that no topic out of the n studied will be in the exam, we can derive it to include two scenarios. The first one being that all of the topics on the exam were studied. To do this we take the number of possible combinations of topics out of the ones studied, this is $\binom{n}{3}$. Please not that in the case that n is less than 3, then $\binom{n}{3} = 0$. To obtain the probability of this scenario happening, we divide it by the number of combinations of exam topics that we calculated in part a: $\binom{10}{3}$.

   The second scenario to consider is when exactly 2 topics on the exam were studied for, to compute this we take the number of combinations of n choose 2. And to get the number of permutations with the 3rd non studied topic, we multiply it by $10 - n$

Thus this gives us the following expression for the probability of failing an exam given that n topics were studied:

$$1 - [\frac{(10-n)*\binom{n}{2} + \binom{n}{3}}{\binom{10}{3}}] \tag{3}$$

We then obtain the following graph using code from appendix a.
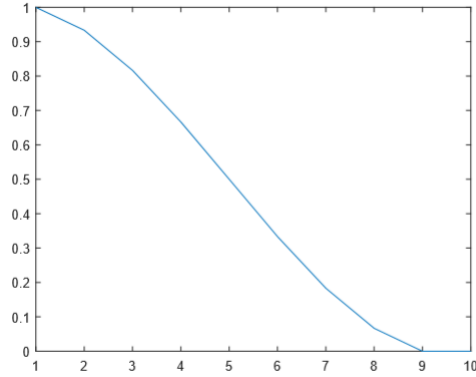
Part d

We can use the expression found in part d to come up with an expression for the same situation where 4 questions are selected instead of 3.

We first notice that our scenarios change, we now have 3 different scenarios to consider in order to find the probability of passing the exam, that is knowing 2, 3 and 4 topics from the exam.

For knowing 4 out of 4 topics, we use the same expression, changing 3s into 4s: $\binom{n}{4}$. When 3 out of 4 topics were studied, we take the number of permutations for $\binom{n}{3}$ and multiply that by $10 - n$ the same way we did before such as to get the total permutations including the 4th topic that was not studied for. The final case, when 2 topics out of 4 were studied, we use the same principle as the latter, we take the amount of permutations possible within these 2 topics and multiply it by the number of permutations possible with the remaning $10-n$ topics. This leaves us with: $\binom{10-n}{2}\binom{n}{2}$

We sum these 3 scenarios and divide by the total amount of permutations

of 10 topics in a 4 questions exam:

$$\frac{\binom{10-n}{2}\binom{n}{2} + (10-n)\binom{n}{3} + \binom{n}{4}}{\binom{10}{4}} \tag{4}$$

We then obtain the following graph using code from appendix b.
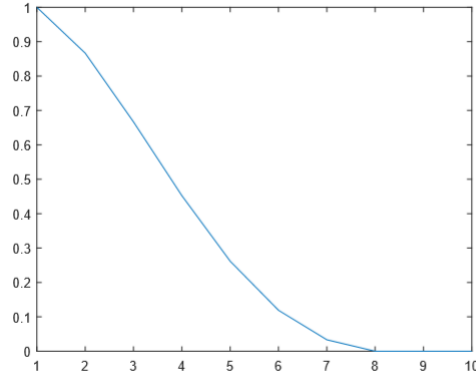
Comparing this plot with the one from question c, we notice that for any equal amount of topics studied, we have an overall higher change of passing with 4 topics in an exam instead of 3. We also notice that the probability of passing becomes 1 at n = 8 instead of n = 9 for a an exam model containing 3 questions.

Part e

The code for the stochastic simulation of the exam setup with three questions is appended to this document in appendix c. Please note that it is the function called **simulation3Questions** and that it **may be used in the following questions**.

Part f

The code for the extended simulation from part e, such that it runs the simulation N times and returns the empirical mean is appended in appendix d. Please note that it is the function called **simulationEmpiricalMean** and that it **may be used in the following questions**. Let's now compute the confidence intervals using the central limit theorem when $n = 7$ with $N = 1000$ and $N = 10000$

We first need the mean and the variance thus we'll start by computing $E[X_i]$

and $var(X_i)$. We've seen in part c that an expression for

$$1 - E[X_i] = 1 - [\frac{(10-n)*\binom{n}{2}+\binom{n}{3}}{\binom{10}{3}}] \leftrightarrow E[X_i] = [\frac{(10-n)*\binom{n}{2}+\binom{n}{3}}{\binom{10}{3}}]$$

Thus, with $n = 7$, $E[X_i] = 0.8167$. To compute the variance, since the Y_i are independent,

$$var(\frac{1}{n}\sum_{i=1}^{N}Y_i) = \frac{1}{n^2}\sum_{i=1}^{N}var(Y_i) = \frac{var(Y_i)}{n} = \sigma^2 \tag{5}$$

Thus, using the empirical mean $E[X_i] = 0.8167$ and the equation for the variance in a bernouilli distributed random variable: $var(X) = \sqrt{p*(1-p)}$, can get the standard deviation σ.

$$\sigma = \sqrt{\frac{E[X_i]*(1-E[X_i])}{N}} = \sqrt{\frac{0.8167*0.1833}{N}} = \sqrt{\frac{0.3869}{N}} \tag{6}$$

Here we have $N = 1000$ and $N = 10000$, thus: $\sigma_{N=1000} = 0.01967$
$\sigma_{N=10000} = \frac{0.3869}{10000} = 0.00622$
    The 95% confidence interval using the CLT is then

$$[\mu 2\sigma, \mu + 2\sigma]_{N=1000} = [0.77736, 0.85604]_{N=1000} \tag{7}$$

$$[\mu 2\sigma, \mu + 2\sigma]_{N=10000} = [0.80426, 0.82914]_{N=10000} \tag{8}$$

Part g

Appendix

Section 0

The code in the following sections may make use of the following function: **customNChooseK**. It is a function which uses nchoosek when $n >= k$. If not then it will return 0. Matlab's default implementation of this situation is to throw an error which is why I decided to write this helper function.

```
function [m] = customNChooseK(n,k)
    if(n < k)
        m = 0;
    else
        m = nchoosek(n,k);
    end
end
```

Section b - Question 1d

```
y1 = zeros([1 10]);
x1 = 1:10;
% Compute the probability for all values of n
for n = x1
    y1(n) = 1 - (((10-n)*
    (customNChooseK(n,3))/customNChooseK(10,4))+
    (customNChooseK(n,4)/customNChooseK(10,4)) +
    (customNChooseK(10-n,2)*(customNChooseK(n,2)/
    customNChooseK(10,4))));
end
%plot(x1, y1);
```

Section a - Question 1c

```
y1 = zeros([1 10]);
x1 = 1:10;
% Compute the probability for all values of n
for n = x1
    y1(n) = 1 -
        (((10-n)*
        (customNChooseK(n,2))/customNChooseK(10,3)) +
        (customNChooseK(n,3)/customNChooseK(10,3)));
end
plot(x1, y1);
```

Section b - Question 1d

```
y1 = zeros([1 10]);
x1 = 1:10;
% Compute the probability for all values of n
for n = x1
    y1(n) = 1 - (((10-n)*
    (customNChooseK(n,3))/customNChooseK(10,4))+
    (customNChooseK(n,4)/customNChooseK(10,4)) +
    (customNChooseK(10-n,2)*(customNChooseK(n,2)/
    customNChooseK(10,4))));
end
%plot(x1, y1);
```

Section b - Question 1e

```matlab
function [X] = simulation3Questions(n)
    % number of topics
    a=1:10;
    % questions drawn
    questions = randperm(numel(a),3);
    % draw for studied subjects
    studiedSubjects = randperm(numel(a),n);

    % get common topics
    pos=intersect(questions, studiedSubjects);
    if(length(pos) >= 2)
        X = 1;
    else
        X =0;
    end
end
```

Section b - Question 1f

```matlab
function [X] = simulationEmpiricalMean(n, N)
    passCount = 0;
    for i = 1:N
        % number of topics
        a=1:10;
        % questions drawn
        questions = randperm(numel(a),3);
        % draw for studied subjects
        studiedSubjects = randperm(numel(a),n);

        % get common topics
        pos=intersect(questions, studiedSubjects);
        if(length(pos) >= 2)
            passCount = passCount + 1;
        end
    end

    X = passCount / N;
end
```