TAMS24 — Computer Lab 1

When you work in Matlab, you need to write your commands/codes in m-file, so that you can easily go back and make changes. After you finish each problem, write down your solutions on the last page.

Problem 1.
Simulation of observations from a discrete distribution

We have often used the roll of the dice to illustrate various calculations of probabilities. You should now use Matlab to simulate such throws. Start Matlab and open an m-file. Name the file to uppg1.m and save it in the appropriate place. Type the following commands in the m-file.

```
n=600; % n tossings

% Simulate one tossing
kast = randi(6,n,1);
kast(101:115)
```

In order to compile/run, in the command window input the name of the file: uppg1.m.

Write `help randi` in the command window to see how `randi` works. See the special examples in the help section.

What is the meaning of the code `kast(101:115)`?

We have 6 different outcomes: 1, 2, 3, 4, 5, 6. Now we want to know the frequency f_i of each outcome $i = 1, \ldots, 6$. To do this, we write `sum(kast==i)` for $i = 1, \ldots, 6$ in the command window. What frequencies do you get?

We can even use Histogram to plot the frequencies. To this end, we should first "comment out"using % in front, that is, `% kast = randi(6,n,1)` (if we don't comment out, then Matlab will simulate new tossings).

Write in the m-file the following

```
% Histogram
figure
hist(kast,[1:6])
```

Now we want to find the sample mean and sample standard deviation of these 600 tossings. Write in the m-file:

```
% Sample mean
medel = mean(kast)
% Sample standard deviation
stdev = std(kast)
```

Compare these two (simulated) values with the theoretical values.

Problem 2.
Confidence interval for the difference of population means

In the command window write `clear` and `clc` in order to clear all previous memories and graphs stored in Matlab. Now open a new m-file and name it as *uppg2.m*.

We now generate 10 observations from $N(22, 6)$ and 31 observations from $N(16, 2)$, using `normrnd`. Write in the command window `help normrnd` to see how this works. Write in the m-file:

```
n = 10;
mu1 = 22;
sigma1 = 6;
m = 31;
mu2 = 16;
sigma2 = 2;

% Generate observations
x = normrnd(mu1,sigma1,n,1);
y = normrnd(mu2,sigma2,m,1);
```

Run this in the command window. When you have done this, comment out `x = normrnd(...);` and `y = normrnd(...);` that is, `% x = normrnd(mu1,sigma1,n,1);` and `% y = normrnd(mu2,sigma2,m,1);`. By doing this we will not generate new observations every time when we run the m-file.

a) We may think that these observations are from two independent populations $N(\mu_1, \sigma_1)$ and $N(\mu_2, \sigma_2)$ respectively. It is clear that two population variances σ_1 and σ_2 are different. We now construct a 95% (two-sided) confidence interval for $\mu_1 - \mu_2$. Write in the m-file:

```
[H1,P1,CI1,STATS1] = ttest2(x,y,0.05,'both','unequal')
```

'unequal' means $\sigma_1 \neq \sigma_2$. In this case, we know that the help variable is approximately t-distributed. Can you find the corresponding degrees of freedom by hand? (you can use `std(x)` and `std(y)` to get the sample standard deviations).

b) Now we want to see if $\sigma_1^2 = \sigma_2^2$. A 95% confidence interval for σ_1^2/σ_2^2 can be obtained as follows:

```
[H2,P2,CI2,STATS2] = vartest2(x,y, 0.05,'both')
```

Is $\sigma_1^2 = \sigma_2^2$?

c) If we **wrongly** assume that these two samples are from $N(\mu_i, \sigma)$, $i = 1, 2$, with a same variance, then what is a (wrong) 95% confidence interval for $\mu_1 - \mu_2$? Try to write/guess the commands yourself below:

In this case, what is the degrees of freedom in the t-distribution ?

Problem 3.
Confidence interval using normal approximations

Open a new m-file and name it as *uppg3.m*.

In this problem we will construct confidence intervals when the population is a Binomial random variable $Bin(n, p)$. In order to use Normal approximation, in Lecture we required that n is large enough such that $np > 10$ and $n(1 - p) > 10$. It is interesting to see what happens if n is not that large.

We will construct a 95% confidence interval for p for large n, and not so large n.

a) We generate 1000 observations from $Bin(16, 0.3)$.

```
n = 16;
p = 0.3;
x = binornd(n,p,1000,1);
```

With these 1000 observations, we can estimate p (1000 times)

```
phat = x/n;
```

and the corresponding 95% confidence interval is

```
lower_lim = phat - 1.96*sqrt(phat.*(1-phat)/n);
upper_lim = phat + 1.96*sqrt(phat.*(1-phat)/n);
```

Now we have 1000 such 95% confidence intervals. Since it is 95%, there should be around 950 such intervals containing the real value $p = 0.3$, and 50 not containing the real value. Now we count how many such intervals which do not contain the real value $p = 0.3$ as follows

```
missar = sum(lower_lim > p) + sum(upper_lim < p)
```

'`missar`' gives you the number of intervals not containing the real value $p = 0.3$. Compare this number with the expected number.

b) Repeat these procedures for 1000 observations from $Bin(80, 0.3)$.

Compare again.

Problem 4.
Confidence interval and hypotheses testing

Let $x_1, \ldots, x_n$ be observations from $N(\mu, \sigma)$. We want to test if $\mu = 5$ or not:

$$H_0 : \mu = 5 \quad \text{mot} \quad H_1 : \mu \neq 5,$$

We will create 1000 samples (every sample size is $n = 20$) from $N(5, 1.2)$.

```
n=20;
N=1000;
mu = 5;
sigma = 1.2;

X = normrnd(mu, sigma, N, n); % Simulations for mu=5
```

We can get the (1000) sample means and (1000) sample standard deviations as follows

```
muhat = mean(X,2);
s = std(X,[],2);
```

a) Assume/pretend that σ is unknown, write down the formula of a 95% confidence interval for μ :

$$I_\mu = ???$$

Use the same procedures in Problem 3 to count how many such 95% confidence intervals not containing the real value $\mu = 5$. The command `tinv(0.975,n-1)` gives you $t_{0.05/2}(n-1)$.

b) In hypotheses testing with an unknown σ, we have in Lecture

$$\text{TS} = \frac{\bar{x} - 5}{s/\sqrt{n}}, \qquad C = (-\infty, -t_{0.05/2}(n-1)) \cup (t_{0.05/2}(n-1), +\infty)$$

We know that if $\text{TS} \in C$ then we reject H_0. Here we have 1000 such TS and 1000 such C (the same), and we want to count how many times H_0 is rejected.

```
TS = (muhat - 5)./(s/sqrt(n));

t = tinv(0.975,n-1);

reject = sum(TS < -t) + sum(TS > t)
```

'`reject`' is the number of times that H_0 is rejected for these 1000 samples.

Question: will the number in a) be always the same as the number in b) ? why ?

SOLUTION PAGE. Write down your name and personal number.

1) ....

2) ....

You should be able to explain your solutions and show your commands in Matlab.

Problem 1 **OK** ......

$f_1 = \ldots\ldots$ $f_2 = \ldots\ldots$ $f_3 = \ldots\ldots$ $f_4 = \ldots\ldots$ $f_5 = \ldots\ldots$ $f_6 = \ldots\ldots$

Sample mean $\bar{x} = \ldots\ldots$; theoretical mean $E(X_i) = \ldots\ldots$; sample standard deviation $s = \ldots\ldots$

Problem 2 **OK** ......

a) $I_{\mu_1-\mu_2} = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$

Degrees of freedom for the t-distribution: ......

b) Confidence interval: ....

Is $\sigma_1^2 = \sigma_2^2$? ....

c) $I_{\mu_1-\mu_2} = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$

Degrees of freedom for the t-distribution:......

Problem 3 **OK** ......

Let x be an observation of the population $X \sim Bin(n, p)$. Write down the formula of a 95% confidence interval for p.

$$I_p = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

With 1000 confidence intervals, how many intervals do you expect not containing the real value p: ......

a) $np =$: .... $n(1-p) =$: ....

The number of intervals not containing the real value $p = 0.3$ when $n = 16$: ......

b) $np =$: .... $n(1-p) =$: ....

The number of intervals not containing the real value $p = 0.3$ when $n = 80$: ......

Problem 4 **OK** ......

a) $I_\mu =$ ....

The number of intervals not containing the real value $\mu = 5$: ......

b) The number of times H_0 is rejected: ......

Are the two numbers in a) and b) always the same ?