ECE 340
Problems on statistics
Solutions

Problem 1

1) The Matlab commands to find the sample mean $\hat{\hat{X}}_n$ and the sample variance $\widehat{\sigma^2_{X_n}}$ are the following

```
>>x = [2.9 3.05 3.1 2.89 3.21 2.99 2.78 2.91 3.03 2.84 2.93 3.12]
>>sample_x_mean = sum(x)/size(x,2); % or we can use 'mean(x)'
>>sample_var = (sum((x - sample_x_mean).^2))/(size(x,2)-1)
% or use var(x)
```

Results are the following

sample_x_mean =

2.9792

sample_var =

0.0160

The sample mean $\hat{\hat{X}}_n$ is simply the average of the numerical values that make up the sample

$$\hat{\hat{X}}_n = \frac{1}{\mathrm{n}}\sum_{i=1}^{n} X_i\,, \quad n = 12$$

The sample variance $\widehat{\sigma^2_{X_n}}$ is the unbiased estimate of the population variance, which describes the magnitude of the fluctuation or the variation of the data around the mean value. The sample variance is defined as the following

$$\widehat{\sigma^2_{X_n}} = \frac{1}{\mathrm{n}-1}\sum_{i=1}^{n}(X_i - \hat{\hat{X}}_n)^2\,, \quad n = 12$$

2) When t =1, we have $\varepsilon = 0.03$. By the Chebyshev's inequality, we know that

$$P\left\{\left|\hat{\hat{X}}_n - \mu\right| \geq \varepsilon\right\} = P\left\{\left|\hat{\hat{X}}_n - 3\right| \geq 0.03\right\} \leq \frac{\sigma^2}{n\varepsilon^2} = \frac{0.16}{100(0.03)^2} = 1.7778 > 1$$

So the upper bound obtained from Chebychev's inequality is 1.
When t =5, we have $\varepsilon = 0.15$. By the Chebyshev's inequality,

$$P\left\{\left|\hat{\hat{X}}_n - \mu\right| \geq \varepsilon\right\} = P\left\{\left|\hat{\hat{X}}_n - 3\right| \geq 0.15\right\} \leq \frac{\sigma^2}{n\varepsilon^2} = \frac{0.16}{100(0.15)^2} = 0.0711$$

So the estimated $P\left\{\left|\hat{\hat{X}}_n - 3\right| \geq 0.15\right\}$ is 0.0711.
When t =10, we have $\varepsilon = 0.3$, and

$$P\left\{\left|\hat{\hat{X}}_n - \mu\right| \geq \varepsilon\right\} = P\left\{\left|\hat{\hat{X}}_n - 3\right| \geq 0.3\right\} \leq \frac{\sigma^2}{n\varepsilon^2} = \frac{0.16}{100(0.3)^2} = 0.0178$$

So $P\left\{\left|\hat{\hat{X}}_n - 3\right| \geq 0.15\right\}$ is upper bounded by 0.0178.

3) To make the sample mean be within 1% of the true mean with a probability 0.9, our problem is to find a sufficiently large *n*. By the Chebyshev's inequality we know that

$$P\left\{\left|\hat{\hat{X}}_n-\mu\right|\le\varepsilon\right\}=1-P\left\{\left|\hat{\hat{X}}_n-\mu\right|\ge\varepsilon\right\}\ge 1-\frac{\sigma^2}{n\varepsilon^2}=0.9,\quad \sigma^2=0.16,\qquad \varepsilon=0.03$$

By solving $1-\frac{\sigma^2}{n\varepsilon^2}=0.9$ we have the solution for *n* as the following

$$n=1777.78$$

Here we round up the value to be 1778, i.e. the sample size should be at least 1778 in order to make the sample mean be within 1% of the true mean with probability no less than 0.9.

4) To find $P\{\hat{\hat{X}}_n\notin[3-3t,3+3t]\}$, we do the following:

$$\begin{aligned}&P\left\{\hat{\hat{X}}_n\notin[3-3t,3+3t]\right\}\\&=1-P\left\{\hat{\hat{X}}_n\in[3-3t,3+3t]\right\}\\&=1-P\left\{3-3t\le\hat{\hat{X}}_n\le 3+3t\right\}\\&=1-P\left\{-3t\le\hat{\hat{X}}_n-3\le 3t\right\}\\&=1-P\left\{-\frac{3t}{\sqrt{\sigma^2/n}}\le\frac{\hat{\hat{X}}_n-3}{\sqrt{\sigma^2 n}}\le\frac{3t}{\sqrt{\sigma^2/n}}\right\}\\&=1-\left[\Phi\left(\frac{3t}{\sqrt{\sigma^2/n}}\right)-\Phi\left(-\frac{3t}{\sqrt{\sigma^2/n}}\right)\right]\end{aligned}$$

$=2-2\Phi\left(\frac{3t}{\sqrt{\sigma^2/n}}\right)=1-\operatorname{erf}\left(\frac{3t}{\sqrt{2}\sqrt{\sigma^2/n}}\right)$, since $\Phi(x)=0.5+0.5\operatorname{erf}(x/\sqrt{2})$ [***prove this***]

For the case *t=1%:*

$$P\{\hat{\hat{X}}_n\notin[3-3t,3+3t]\}=1-\operatorname{erf}\left(\frac{3t}{\sqrt{2}\sqrt{\sigma^2/n}}\right)=0.4533$$

For the case *t=5%:*

$$P\{\hat{\hat{X}}_n\notin[3-3t,3+3t]\}=1-\operatorname{erf}\left(\frac{3t}{\sqrt{2}\sqrt{\sigma^2/n}}\right)=1.7683e-004$$

For the case *t=10%:*

$$P\{\hat{\hat{X}}_n\notin[3-3t,3+3t]\}=1-\operatorname{erf}\left(\frac{3t}{\sqrt{2}\sqrt{\sigma^2/n}}\right)=6.3838e-014$$

5) Put

$$\begin{aligned}q&=P\left\{\hat{\hat{X}}_n\in[3-a,3+a]\right\}\\&=P\left\{3-a\le\hat{\hat{X}}_n\le 3+a\right\}\\&=P\left\{-a\le\hat{\hat{X}}_n-3\le a\right\}\\&=P\left\{-\frac{a}{\sqrt{\sigma^2/n}}\le\frac{\hat{\hat{X}}_n-3}{\sqrt{\sigma^2 n}}\le\frac{a}{\sqrt{\sigma^2/n}}\right\}\end{aligned}$$

$=\left[\Phi\left(\frac{a}{\sqrt{\sigma^2/n}}\right)-\Phi\left(-\frac{a}{\sqrt{\sigma^2/n}}\right)\right]$, since $\frac{\hat{\hat{X}}_n-3}{\sqrt{\sigma^2 n}}$ is a zero-mean and unit-variance Gaussian rv and $\Phi$is the cumulative distribution function of the zero-mean and unit-variance Gaussian rv. So,

$q = 2\Phi\left(\frac{a}{\sqrt{\sigma^2/n}}\right) - 1 = \text{erf}\left(\frac{a}{\sqrt{2}\sqrt{\sigma^2/n}}\right)$ since Φ(x)= 0.5+0.5 erf(x/$\sqrt{2}$)

Hence, $a = \sqrt{2\sigma^2/n}\,\text{erf}^{-1}(q)$.

We can use Matlab's "erfinv" function to calculate the inverse erf.
For *n*=100, we find *a*=0.0784 and *I*=[2.9216, 3.0784]; for *n*=1000, *a*= 0.0248 and *I*=[2.9752, 3.0248]. Note that the interval becomes narrower as *n* increases while keeping the confidence level *q* fixed. **Interpret this.**

Problem 2

By the CLT (central limit theorem), we can estimate $P\{X \geq 110{,}000\}$ with the following

$$P\{X \geq 110{,}000\} = \int_{110000}^{\infty} f_X(x)dx$$

where $f_X(x) = \frac{1}{12524\sqrt{2\pi}} \exp\left(-\frac{(x-98711)^2}{2(12524)^2}\right)$

So, we have

$$P\{X \geq 110{,}000\} = \int_{110000}^{\infty} \frac{1}{12524\sqrt{2\pi}} \exp\left(-\frac{(x-98711)^2}{2(12524)^2}\right) dx$$

$$= 1 - \Phi\left(\frac{110000 - 98711}{12524}\right) = 1 - (0.5 + 0.5\text{erf}\left(\frac{110000 - 98711}{\sqrt{2}\,12524}\right)) = 0.1837$$

Problem 3

The Matlab code used to find out the sample means and the sample variances are the following:

```
==========================================================================

clc
clear all
close all

n = [1000 5000 10000 100000];
for i = 1:1:size(n,2)
    X = random('Poisson',10,n(i),1);
    xmean(i) = mean(X);
    xvar(i) = var(X);
end
xmean
xvar
==========================================================================
```

The results in the command window are shown below:

xmean =

9.8450   10.0198   10.0138   9.9966

xvar =

9.9609   9.8674   9.9280   9.9182

The Poisson random variable we generated has a mean value of 10, we also know that the variance of a Poisson random variable has the same value as the mean value, thus the variance of the Poisson random variable we generated should have a variance 10 as well. We see that as we have a larger sample size, the sample mean and sample variance are closer to the true mean and the true variance, respectively.