SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN
Established in collaboration with MIT

Assessed Problem Set 3 Solutions

Due: Wed, 31st Oct, 23:59PM
Delivery method: E-dimension
Format: Portable Document Format (PDF)

For all the python based questions in this problem set, please copy and past the code that you used for each function/operation and the result of the calculation.

Problem 1: Propagation of Error
Snell's law relates the angle of refraction θ_2 of a light ray in a medium of refractive index n_2 to the angle of incidence θ_1 of a ray travelling in a medium of index n_1 through the equation $n_2 \sin\theta_2 = n_1 \sin\theta_1$.

i Find n_2 and its uncertainty from the following measurements: $\theta_1 = 22.02 \pm 0.02°$, $\theta_2 = 14.45 \pm 0.02°$, and $n_1 = 1.0000$

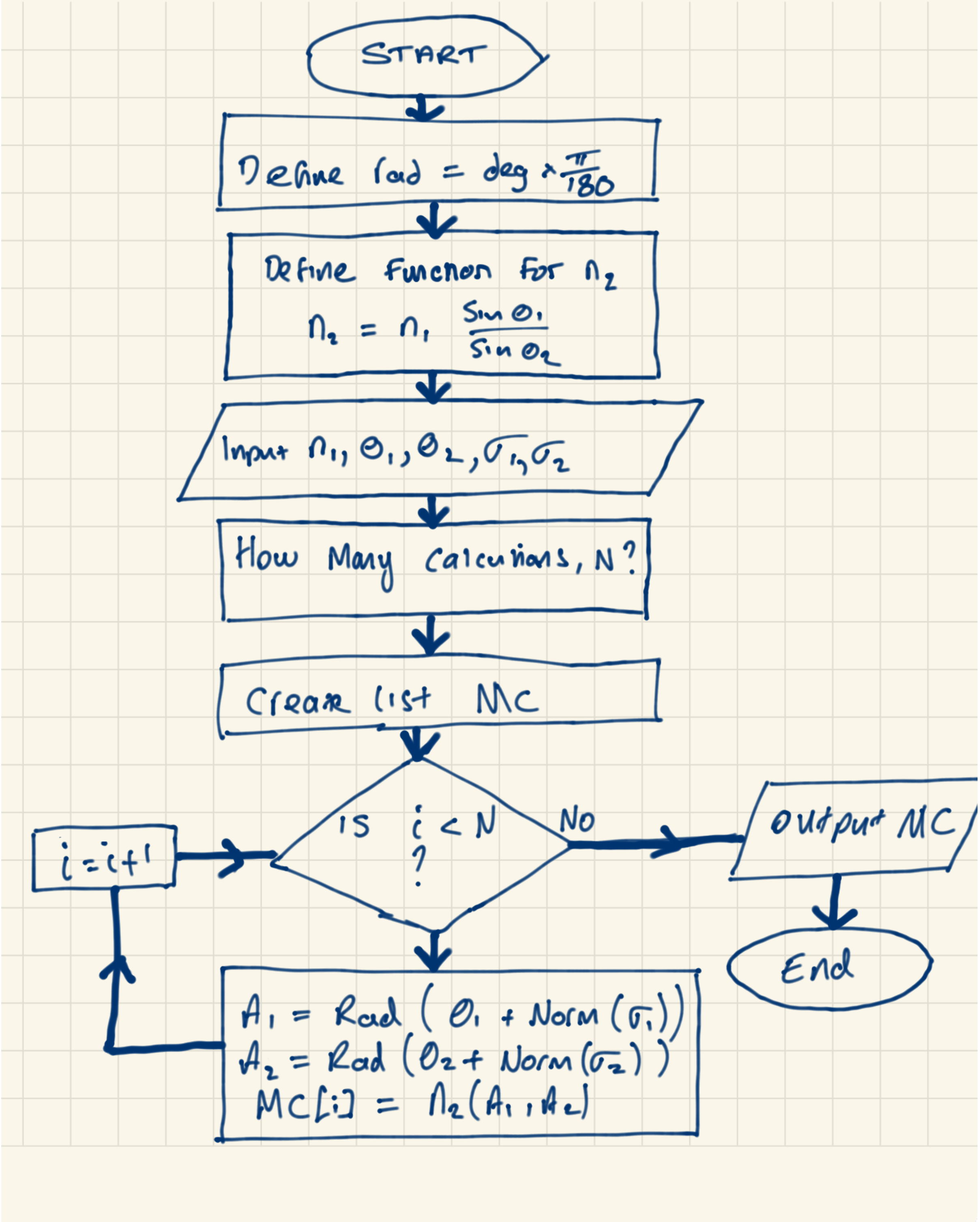
ii Draw a block flow diagram to show how you would compute the mean and uncertainty in n_2 using a Monte Carlo algroithm

iii Compute the mean and standard deviation using your Monte Carlo algorithm. Plot a histogram to show the distribution of n_2 for N=10, 100, and 10^4 computations. Comment on the result.

Problem 1 Solutions
Prolem 1(i)

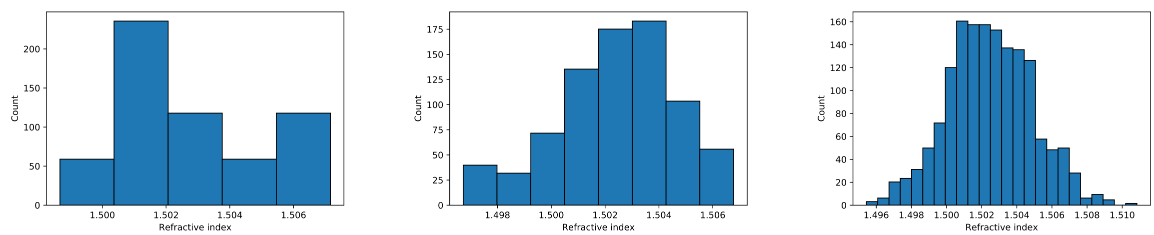
$$
\begin{aligned}
n_2 &= \frac{n_1 \sin\theta_1}{\sin\theta_2} \\
\frac{\partial n_2}{\partial \theta_1} &= n_1 \frac{\cos\theta_1}{\sin\theta_2} = \frac{\cos(22.02 \times \pi/180)}{\sin(14.45 \times \pi/180)} = 3.715 \\
\frac{\partial n_2}{\partial \theta_2} &= n_1 \frac{\sin\theta_1}{-\sin^2\theta_2} \cos\theta_2 = \frac{\sin(22.02 \times \pi/180)}{-\sin^2(14.45 \times \pi/180)} \cos(14.45 \times \pi/180) = 5.86 \\
\sigma &= \left[\left(\frac{\partial n_2}{\partial \theta_1} \sigma_1 \right)^2 + \left(\frac{\partial n_2}{\partial \theta_2} \sigma_2 \right)^2 \right]^{1/2} \\
&= \left[(3.715(0.02 \times \pi/180))^2 + (5.86(0.02 \times \pi/180))^2 \right]^{1/2} \\
&= 2.42 \times 10^{-3} \\
n_2 &= 1.5025 \pm 0.0024
\end{aligned}
$$

Problem 1(ii)


START
Define rad = deg × π/180
Define function for n2
n2 = n1 sin θ1 / sin θ2
Input n1, θ1, θ2, σ1, σ2
How many calculations, N?
Create list MC
IS i < N ?
No
Output MC
End
i = i+1
A1 = Rad (θ1 + Norm (σ1))
A2 = Rad (θ2 + Norm (σ2))
MC[i] = n2(A1, A2)

Problem 1(iii)

```
import scipy.stats as ss
import numpy as np
from scipy.stats import f
import matplotlib.pyplot as plt

#convert degrees to radians
def d2r(deg):
    return deg*np.pi/180

#define the Snells Law Equation
def snell(n1, theta1, theta2):
    rad1=d2r(theta1)
    rad2=d2r(theta2)
    return n1*(np.sin(rad1))/np.sin(rad2)


#Enter the constants
theta1mean=22.02
theta2mean=14.45
thetaerror=0.02

#create a list called n2 to store the MC refractive indices
n2=[]
for i in range(1000):   #Perform a MC simulation
    theta1=ss.norm.rvs(theta1mean,thetaerror)
    theta2=ss.norm.rvs(theta2mean,thetaerror)
    n2.append(snell(1, theta1,theta2))

#plot the historgram of the results
plt.figure()
plt.hist(n2, bins='auto', edgecolor='black', normed=1)
plt.xlabel("Refractive␣index")
plt.ylabel("Count")
plt.savefig('Q1_1000.pdf')
```



(a) N=10 (b) N=100 (c) N=1000

Figure 1: Snells Law Histograms

Comment The standard deviation of n_2 does not depend on the number of samples. However, the standard error of the sampling distribution does depend on the number samples.

Problem 2: Goodness of fit

Table 1 gives the age distribution of part-time college students as determined five years ago.

Age	18-24	25-34	35-44	45-54	55 and over
Percent	25%	35%	25%	10%	5%

Table 1: Age distribution of part time college students 5 years ago

1500 part-time students were recently surveyed across the Singapore. 352 were in the age group 18-24, 501 were in the age group 25-34, 371 were in the age group 35-44, 126 were in the age group 45-54, and the remainder were in the age group 55 or over.
Here we aim to understand whether the proportions from 5 years ago are representative of today's college student proportions?

i Write the null hypothesis, H_0, and alternative, H_a, hypothesise

ii Compute the χ^2 test statistic (if you use python, please include your code, if you do it manually by hand, please copy out a table showing your calculations)

iii Is there a statistically signifficant difference in the proportions at the $\alpha = 0.01$ signifficance level?

Problem 2i

$H_o: \quad p_1 = .25, \quad p_2 = .35, \quad p_3 = .25, \quad p_4 = .10, \quad p_5 = 0.05$ H_a: The current proportions have changed in the last 5 years.

Problem 2ii

Age Distribution

Age	Obsvered (o)	Expected (e)	(o-e)^2/e
18—24	352.00	375.00	1.41
25–34	501.00	525.00	1.10
35—44	371.00	375.00	0.04
45-54	126.00	150.00	3.84
>55	150.00	75.00	75.00
Total	1500.00	1500.00	
Chi^2			**81.39**

Table 2: Age distribution based on numbers (observed) now and proportions from 5 years ago (expected)

Problem 2iii

There 5 sets of numbers (rows) that we are comparing. Therefore, dof=5-1=4. $\chi^2_{\alpha=0.01} = 13.277$. Since $81.39 > 13.277$, we must reject H_0 and conclude that the age distributions are different.

Problem 3: Confidence

In a random sample of 85 bearings, 10 have a surface roughness that is larger than specifications allow.

i Estimate of the proportion, $\hat{p}$, of bearings that exceed the roughness specification.

ii Calculate the 95% two-sided confidence interval for the proportion of bearings that are rougher than specification.

iii How large should the sample be if we want to be *at least* 95% confident that our estimate $\hat{p}$ is within 0.05 of the true proportion, p?

Problem 3i

$$\hat{p} = 10/85 = 0.1176$$

Problem 3ii

$$\hat{p} - z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{1}$$

$$0.1176 - 1.96\sqrt{\frac{0.1176(1-0.1176)}{85}} \leq p \leq 0.1176 + 1.96\sqrt{\frac{0.1176(1-0.1176)}{85}} \tag{2}$$

$$0.0491 \leq p \leq 0.1861 \tag{3}$$

Problem 3iii

Let $\hat{P}$ be the true proportion. The error in estimating $\hat{P}$ with the sample of bearings is: $E = \hat{P} - \hat{p}$.

$$E = \hat{P} - \hat{p} = z_{\alpha/2}\sqrt{\frac{p-(1-p)}{n}}$$

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 p(1-p)$$

For 95% confidence, we have a signifficance level of $\alpha = 0.05$, and since it is a two tail test, we need to look up z value for 0.025, which is $z_{\alpha/2=0.025} = 1.96$. So:

$$n = \left(\frac{1.96}{0.05}\right)^2 0.1176(1-0.1176) = 160$$

So based on the proportion in the sample the minimum number of samples would be 160.

<u>Note:</u> If the proportion is unknown, we can still estimate a minimum number of samples by using the largest possible value of $p(1-p)$, which occurs when $p = 0.5$. When $p = 0.5$, n would be 385.