CS 215 ASSIGNMENT

Tathagat Verma - 180050111
Neel Aryan Gupta - 180050067

Question 1

$X := (X_1, X_2)$ $\quad X_1, X_2 \to$ i.i.d from uniform distribution over $(-1, 1)$.

a.) $P($ X takes values within circle of radius 1 $)$

$= P(X_1^2 + X_2^2 \le 1)$

$$= \int_{-1}^{1} \left(\int_{-\sqrt{1-x_1^2}}^{\sqrt{1-x_1^2}} \frac{dx_2}{2} \right) \frac{dx_1}{2} \qquad \underbrace{\{P(x_2^2 \le 1 - x_1^2 \mid x_1)\}}_{\text{sum over all } X_1}$$

$$= \frac{1}{4} \int_{-1}^{1} \left(2 \times \sqrt{1 - x_1^2} \right) dx_1$$

$$= \frac{1}{2} \int_{-1}^{1} \sqrt{1 - x_1^2}\, dx_1$$

Put $x_1 = \sin\theta \Rightarrow dx_1 = \cos\theta\, d\theta$

$$= \frac{1}{2} \int_{-\pi/2}^{\pi/2} \cos^2\theta\, d\theta = \cancel{\frac{1}{2} \left| \sin\theta \right|_{-\pi/2}}$$

$$= \frac{1}{4} \int_{-\pi/2}^{\pi/2} (\cos 2\theta + 1)\, d\theta = \frac{1}{4} \left(\frac{\sin 2\theta}{2} + \theta \right)_{-\pi/2}^{\pi/2}$$

$$= \frac{1}{4}\left(\cancel{\frac{\sin \pi}{2}}^{\,0} + \frac{\pi}{2} - \cancel{\frac{\sin(-\pi)}{2}}^{\,0} - \left(-\frac{\pi}{2}\right)\right)$$

$$= \frac{\pi}{4}$$

$$\therefore \boxed{P(\|X\|_2 \le 1) = \pi/4} \quad \text{(Ans.)}$$

(b.) To obtain value of π;

we will generate a large number (say N) of bivariate random variables X, with the given specifications on X_1 & X_2.

Then ~~&~~ we'll calculate the fraction of points which lies in the unit circle. For large N this value will converge to that predicted theoretically i.e. $P(\|X\|_2 \le 1) = \pi/4$.

Hence π can be estimated as

$$4 \times (\text{fraction of points satisfying } X_1^2 + X_2^2 \le 1)$$

A justification can simply be the Weak Law of Large Numbers.

~~X_1, X_2~~

we can define a random variable Y
as 1 when x is inside unit circle
as 0 when X is outside unit circle.
So Y is a Bernoulli random variable with parameter $= \pi/4$.

Now after generating N values of X, we have $Y_1, Y_2, Y_3, \ldots, Y_N$ as iid
& $E(Y) = \pi/4 \rightarrow$ finite.

$$S_N = \frac{Y_1 + Y_2 + \cdots + Y_N}{N}.$$

$S_N \rightarrow E[Y]$ as $N \rightarrow \infty$ from Weak Law of Large Numbers.

i.e. $\frac{\pi}{4}$ = fraction of points lying inside unit circle.

Part (c)

Estimate of π with

$N = 10$	is	2.00000
$N = 10^2$	is	3.24000
$N = 10^3$	is	3.24400
$N = 10^4$	is	3.12280
$N = 10^5$	is	3.13972
$N = 10^6$	is	3.14213
$N = 10^7$	is	3.14178
$N = 10^8$	is	3.14165

With the normal code,
for ~~10~~ values of N as large as 10^9,
the problem that arises ~~is~~ is a lot of
memory is taken by the array ; exceeding
"array size limit", in generating 10^9 numbers.
To avoid this, we can run the code of
10^7, ~~&~~ 100 times, an at each time out of
the 100, we'll just require arrays of
size 10^7 which is within the array size
limit. So we can compute the ~~avex~~. fraction at
each step out of the 100 to get final value.

The normal code doesnot handle this case when $N = 10^9$, but running 10^7 a 100 times does work.

{At each step we are effectively creating & destroying arrays of length 10^7, never exceeding it & using the value of that ~~fx~~ fraction to compute overall fraction.}

(d) For this, we will use central Limit Theorem,

Again we define a random variable Y which is 1 when X is inside unit circle
0 when X is outside unit circle.

Y is a Bernoulli Random variable with parameter p. ($p = \pi/4$).
Varience of Y is $p(1-p)$.

$$S_N = \frac{\Sigma Y_i}{N}.$$

for large values of N, S_N will belong to a Gaussian.

$$S_N \sim \mathcal{N}\left(E(Y), \frac{Var(Y)}{N}\right)$$

$$\Rightarrow 4S_N \sim \mathcal{N}\left(4 \times p, \frac{16 \times p(1-p)}{N}\right)$$

To have $4S_N$ lie within

$[\pi - 0.01, \pi + 0.01]$ with 0.95 probability,

we note that for a Gaussian, 95% of sample datapoints lie with the range
$(\mu - 1.96\sigma, \mu + 1.96\sigma)$

$\therefore 0.01 \geq 1.96\sigma$

$\therefore 1 \geq 196 \times \sqrt{\frac{16\, p(1-p)}{N}}$

$\therefore N \geq 16 \times p \times (1-p) \times 196^2$

we use $p = \frac{3.1416}{4}$.

$\therefore N \geq \cancel{16} \times \frac{3.1416 \times 0.8584 \times 196^2}{\cancel{16}}$

$\Rightarrow \boxed{N \geq 103{,}598}$ (Ans)

Estimate of sample size N is 103,598.

(Ans)